

Samuel G. Armato III, PhD  
 Geoffrey McLennan, MD,  
 PhD  
 Michael F. McNitt-Gray,  
 PhD  
 Charles R. Meyer, PhD  
 David Yankelevitz, MD  
 Denise R. Aberle, MD  
 Claudia I. Henschke, MD,  
 PhD  
 Eric A. Hoffman, PhD  
 Ella A. Kazerooni, MD, MS  
 Heber MacMahon, MD  
 Anthony P. Reeves, PhD  
 Barbara Y. Croft, PhD  
 Laurence P. Clarke, PhD  
**For the Lung Image  
 Database Consortium  
 Research Group**

**Index terms:**

Computers, diagnostic aid  
 Lung, CT, 60.1211  
 Lung, nodule, 60.31

**Published online**

10.1148/radiol.2323032035  
**Radiology 2004; 232:739–748**

**Abbreviations:**

CAD = computer-aided diagnosis  
 LIDC = Lung Image Database  
 Consortium  
 NCI = National Cancer Institute

<sup>1</sup> From the Department of Radiology, MC 2026, University of Chicago, 5841 S Maryland Ave, Chicago, IL 60637 (S.G.A.). Affiliations for all other authors and the members of the Lung Image Database Consortium Research Group are listed at the end of this article. Received December 17, 2003; revision requested February 16, 2004; revision received March 11; accepted March 16. Supported in part by USPHS Grants U01CA091085, U01CA091090, U01CA091099, U01CA091100, and U01CA091103. **Address correspondence to S.G.A.** (e-mail: [s-armato@uchicago.edu](mailto:s-armato@uchicago.edu)).

S.G.A., H.M., and C.E.M. are shareholders in R2 Technology. E.A.K. receives research support from and serves on the medical advisory board of GE Medical Systems.

© RSNA, 2004

# Lung Image Database Consortium: Developing a Resource for the Medical Imaging Research Community<sup>1</sup>

To stimulate the advancement of computer-aided diagnostic (CAD) research for lung nodules in thoracic computed tomography (CT), the National Cancer Institute launched a cooperative effort known as the Lung Image Database Consortium (LIDC). The LIDC is composed of five academic institutions from across the United States that are working together to develop an image database that will serve as an international research resource for the development, training, and evaluation of CAD methods in the detection of lung nodules on CT scans. Prior to the collection of CT images and associated patient data, the LIDC has been engaged in a consensus process to identify, address, and resolve a host of challenging technical and clinical issues to provide a solid foundation for a scientifically robust database. These issues include the establishment of (a) a governing mission statement, (b) criteria to determine whether a CT scan is eligible for inclusion in the database, (c) an appropriate definition of the term *qualifying nodule*, (d) an appropriate definition of “truth” requirements, (e) a process model through which the database will be populated, and (f) a statistical framework to guide the application of assessment methods by users of the database. Through a consensus process in which careful planning and proper consideration of fundamental issues have been emphasized, the LIDC database is expected to provide a powerful resource for the medical imaging research community. This article is intended to share with the community the breadth and depth of these key issues.

© RSNA, 2004

Image processing and computer vision techniques are being developed to facilitate radiologic tasks through a paradigm known as computer-aided diagnosis (CAD). CAD has been defined as a diagnosis made by a radiologist with the benefit of information generated by computerized image analysis (1). Although some investigators distinguish between the concepts of computer-aided detection and CAD, others have interpreted CAD broadly as encompassing both the detection task and the classification task. For the present article, we will adopt the view that CAD involves the application of computerized techniques to a broad range of clinical challenges in radiology.

The investigation of CAD techniques for chest radiography has a long history (2). Throughout this time, special emphasis has been placed on automated analyses of lung nodules within the complex background of superimposed anatomic structures that results when the three-dimensional human body is projected onto a two-dimensional image. Computed tomography (CT) generates images in a manner that eliminates the superposition of anatomic structures; the trade-off, however, is that much more image data per patient are acquired for radiologist interpretation.

## MOTIVATION TO FORM A DATABASE

### Computer-aided Diagnosis

Two developments promise to greatly affect the practice of CT. The first is the advent of multi-detector row CT scanners combined with gantry rotation times of less than 500

msec per rotation. These scanners generate an extensive amount of image data per examination (3), but they have been shown to increase patient throughput (4). A single thoracic CT examination routinely generates more than 300 thin-section images from a single breath hold with state-of-the-art 16-section scanners. Thus, it is now possible to acquire volumetric image data of the thorax with images composed of isotropic voxels with millimeter and submillimeter resolution. The second development is the growing awareness of lung cancer screening by using a low-dose helical CT protocol (5–7). These developments, combined with the fact that CT examinations are being used more often for a wide range of diagnostic tasks, are dramatically increasing the workload of radiologists, which, in turn, may result in more errors of omission (8–11). Consequently, CAD techniques may become a practical necessity in the interpretation of CT scans; for example, CAD may be used in the evaluation of lung nodules.

The development of CAD for use in the evaluation of lung nodules on CT images has accelerated in recent years. These CAD methods may generally be divided into the categories of nodule detection and classification. A number of researchers around the world have been developing computerized nodule detection techniques (12–26). These groups have incorporated a variety of approaches, including gray-level threshold methods (12,16,20,23–26), fuzzy clustering algorithms (13), spatial filtering (15), template matching (22), object-based deformation procedures (17), morphologic analysis (18), and model-based techniques (14,21). The computerized classification of lung nodules as cancerous or noncancerous has also received attention from a number of groups that used various combinations of shape and gray-level distribution characteristics (27–37). Some of these groups use nodule features depicted on a single CT scan for classification, while others capture information regarding change in nodule features over time, as demonstrated with multiple CT examinations in the same patient.

Although the development of CAD methods for lung nodules on CT scans has accelerated, all of these diverse approaches have one common constraint: Access to well-characterized image data is limited. This deficiency represents a fundamental limitation to CAD research in this area. Consequently, it is difficult to develop and test CAD methods in a robust and reliable fashion.

## CAD Requirements

Two fundamental requirements common to all CAD research are patient image data and a definition of “truth” for the specific task. For investigators not affiliated with medical centers, access to image data requires collaborative research agreements that may be hindered by federal regulations that govern the transmission of patient data, including images, to outside institutions (38). These investigators may choose to apply their talents to nonmedical problems, or they may confine their investigations to simulated image data with limited real-world application. Even for investigators associated with medical centers, the seemingly ready supply of clinical image data that might be reaped for research purposes is not without its barriers: The task of identifying and collecting appropriate images for any specific research activity can be a laborious process. In addition, the important need to secure patients’ consent for use of their images and clinical data in research can consume additional time and resources.

Once a set of appropriate images has been assembled, a medical opinion concerning the task-specific truth must be rendered on the content of those images. For example, investigators developing automated lung nodule detection methods will require the opinion of an experienced radiologist regarding the presence and precise location of nodules on the CT scans. More appropriately, a panel of experienced thoracic radiologists would be used to establish the truth for the nodule detection task, since the variability among radiologists in the detection of lung nodules is substantial (39). Truth for other CAD tasks will require other data, such as follow-up CT scans that enable radiologists to evaluate nodule growth, pathology reports, or radiologist-drawn nodule outlines. The collection of this information can be time consuming. Furthermore, the notion of a single truth in any particular instance is a fallacy; differences of opinion—even among experienced subspecialty radiologists—are a reality, and the resulting variability in the truth must be understood and appreciated by CAD investigators.

The task of gathering image data sets along with associated truth is expensive, and when different research groups use separate databases, a reliable comparison of CAD methods reported in the literature is impossible. The increasing need for CAD in the clinical practice of radiology lends urgency to the creation of com-

mon image databases with established truth to facilitate a direct comparison of CAD methods.

## Origins of the Lung Image Database Consortium

The role of imaging in science is expanding, and consequently, the need for appropriate image-archiving and image-sharing mechanisms is evolving (40,41). Recognizing that the development of CAD methods by the imaging research community would be facilitated and stimulated through access to a well-characterized repository of CT image data, the National Cancer Institute (NCI) released a request for applications in April 2000, entitled *Lung Image Database Resource for Imaging Research*, as a U01 funding mechanism (also known as a cooperative agreement). The intent of this initiative was “to support a consortium of institutions to develop consensus guidelines for a spiral CT lung image resource, and to construct a database of spiral CT lung images” (42). Through the request for applications, the NCI recognized that “the generation of standardized databases requires the development of consensus on many issues related to database design, accessibility, metrics and statistical methods for evaluating image-processing algorithms” (43). From among the applicants who submitted grant applications in response to the request for applications, the following five institutions were selected through a peer-review process to form the Lung Image Database Consortium (LIDC): Cornell University; the University of California, Los Angeles; the University of Chicago; the University of Iowa; and the University of Michigan. The LIDC web site may be found at [cip.cancer.gov](http://cip.cancer.gov).

The five-member consortium represents access to an expanded pool of CT data for CAD research. The LIDC’s charge was not just to collect CT scans with lung nodules; this consortium has spent a great deal of effort thus far to harness the distinct experiences and divergent opinions of the member institutions and other participating individuals to provide a solid foundation for a robust database that will meet the anticipated needs of CAD investigators. Toward this end, the LIDC has identified a number of critical technical and clinical issues that must be addressed to ensure the effort of creating the database is properly focused. This foundation-laying process has been evolutionary in nature, as every issue raised has spawned multiple other issues for

consideration. This article is intended to share with the community the breadth and depth of several of these key issues.

## MISSION STATEMENT

The scientific staff of the NCI, through the request for applications, established the need for the Lung Image Database by articulating constraints that hindered rapid advancements by the medical imaging research community. Once the member institutions of the LIDC were convened, the role of the NCI intentionally shifted from that of a motivator to that of a facilitator. The LIDC Steering Committee (with two voting investigators from each of the five member institutions and two voting representatives from the NCI) was effectively granted authority to set its own course. In this context, the first order of business addressed was the formulation of a mission statement that expressed the consensus view of the purpose of the LIDC. Two excerpts capture the essence of the LIDC mission statement:

To develop an image database as a web-accessible international research resource for the development, training, and evaluation of CAD methods for lung cancer detection and diagnosis using helical computed tomography (CT).

The database should enable the correlation of performance of CAD methods for detection and classification of lung nodules with spatial, temporal and pathologic ground truth.

These two statements convey the intended public accessibility of CT image data and truth data contained within the database to facilitate all aspects of lung nodule CAD research.

To provide a true research resource, the database must contain more than images. Consequently, the database will consist of an image repository and an associated relational database in which nodule features (eg, radiologist outlines, subjective subtlety ratings, and lobar location); technical parameters of the scan available in the digital imaging and communications in medicine, or DICOM, header (eg, exposure rate, reconstruction algorithm, and scanner model); and patient information (eg, age, sex, smoking history, and any available diagnostic information, such as the results of follow-up studies or pathologic examinations) are recorded. The relational database component will give users of the database the ability to extract customized image subsets based on search results.

## INCLUSION CRITERIA

The LIDC will not prescribe CT scanning protocols. Moreover, no CT examination will be performed specifically for the LIDC effort; rather, scans will be culled from the archives and clinical caseloads of the member institutions. Accordingly, the database will contain scans acquired with technical parameters in accordance with the protocols established at each site and with the CT scanners available at each site; thus, a range of scanner manufacturers and models will be represented. Cases considered for inclusion in the database will undergo a specific selection process. The selected cases must capture the complexities of CT scans encountered during routine clinical practice; this interest, however, must be balanced by practical considerations and the desire to exclude cases that will be of limited use to CAD investigators. We have established inclusion criteria on two levels: First, the CT scan itself must meet several defined criteria. Second, nodules that appear on the scan must satisfy another set of criteria.

A distribution of nodule size, radiologic pattern, subtlety, abnormalities, and anatomic location will be sought by means of monitoring case accrual. Similarly, we will attempt to maintain a distribution of patient demographics. Clearly, strict adherence to target distributions established a priori will limit the overall number of scans that may be accrued in the database; therefore, scan quantity will have priority over any target distributions of patient or nodule characteristics.

### Scan Inclusion Criteria

CT scans from both diagnostic and screening studies will be included in the database. Ideally, we would prefer to include the complete radiologic history of a patient, beginning with the screening study at which any nodule was first demonstrated and encompassing all subsequent diagnostic follow-up studies. Scans acquired during and after any prescribed treatment would complete the history. While the screening study will certainly demonstrate the complete thoracic anatomy, subsequent studies may be limited to the anatomic location of a suspected nodule. These limited-anatomy scans will be included in the database; although such scans will be irrelevant for computerized detection methods, they will certainly benefit computerized classification methods. We recognize that including the complete history of a patient

will often be difficult in practice; the database, therefore, may contain any combination of screening scans, full-thorax diagnostic scans, and limited-anatomy diagnostic scans for any specific patient.

In an effort to increase the yield of scans from the five institutions, the database will contain both prospective and retrospective cases. To maintain the technical relevance of the database, however, scans with reconstruction interval or section thickness greater than 5 mm will be excluded. No requirements with regard to scanner pitch, exposure, tube voltage, or reconstruction algorithm will be imposed. To achieve a robust database, the constituent scans will represent a variety of technical parameters from a variety of scanner models.

Image quality is an issue that has been difficult to define. Since image artifacts that are caused by patient factors, such as respiratory motion, and scanner factors, such as beam hardening, are a reality in medical imaging, CAD developers will need to contend with their presence. Consequently, while we will not deliberately fill the database with scans that contain severe artifacts, scans with high levels of noise or with streak, motion, or metal artifacts will be included. When such scans are included, however, they will be denoted by a "marginal" or "unacceptable" rating in the image quality field. The overall image quality of a scan, including the presence of artifacts and other factors, will be assessed by an LIDC radiologist and recorded in the relational database so that investigators may explicitly exclude or include images with marginal or unacceptable image quality from an image dataset.

Since the database is intended to reflect the realities of clinical practice, it will not comprise only pristine cases that contain only nodules. Nodule cases will be included despite the presence of other abnormalities on the scan, unless the other abnormality is spatially contiguous with the nodules and substantially interferes with visual interpretation. The database will be augmented by a collection of "normal" cases (eg, scans that do not contain nodules but may or may not contain other abnormalities), the number of which has yet to be determined.

### Nodule Inclusion Criteria

The diameters of nodules (both calcified and noncalcified) included in the database will not exceed 30 mm, which is consistent with upper limits of nodule

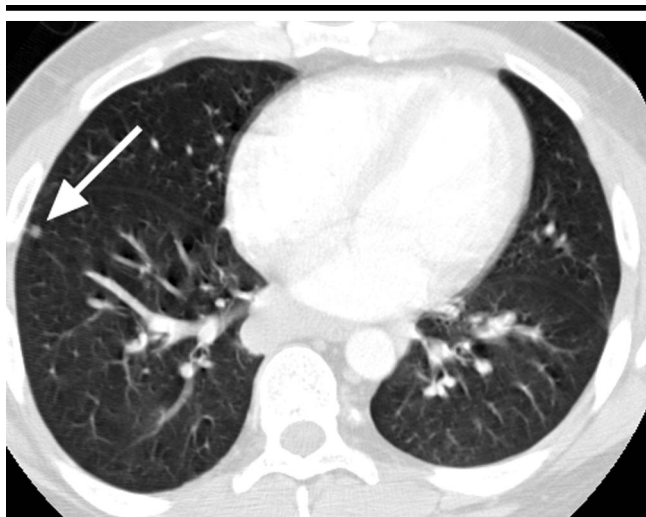


Figure 1. Transverse thoracic CT scan shows a lesion (arrow) that clearly satisfies the Fleischner Society definition of a nodule.

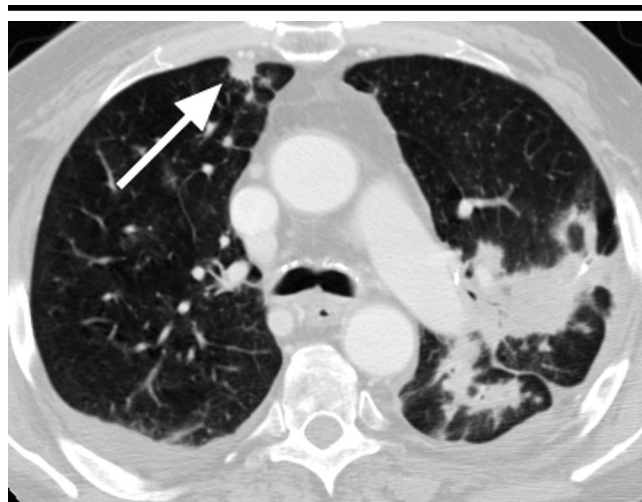


Figure 2. Transverse CT scan shows a lesion (arrow) that appears to satisfy the Fleischner Society definition of a nodule.

size found in the literature (44). Furthermore, the minimum effective diameter for included nodules has been set at 3 mm. Some investigators maintain that nodules smaller than 5 mm may be of limited clinical importance, while others contend that since the benefits of detecting small lung cancers are presently unknown, to exclude from consideration nodules in the 1–2-mm size range may limit the relevance of the database. The 3-mm lower limit for the database strikes a compromise between these views and also takes into consideration the practical issue that all lesions identified as nodules in the database will require effort to define spatial location and extent and to follow through subsequent examinations. As a further compromise, the presence of nodules smaller than 3 mm that are suspicious for cancer will be indicated, but the spatial extents of these small nodules will not be defined.

The nodules in the database may be primary lung cancers, metastatic disease, or a noncancerous process. An upper limit on the number of nodules per scan, however, has been set at six as a general guideline. Although this number is arbitrary, scans with more than six noncalcified nodules are much less likely to represent primary lung cancer (5). This upper limit on the number of nodules again takes into consideration the practical issue that all lesions identified as nodules in the database will require effort to define spatial location and extent and to follow through subsequent examinations.

Perhaps the most important nodule inclusion criterion is that the nodules on the scans must be nodules. Exactly what

this means has been the subject of extensive discussions and will be described in the next section.

#### NODULE DEFINITION

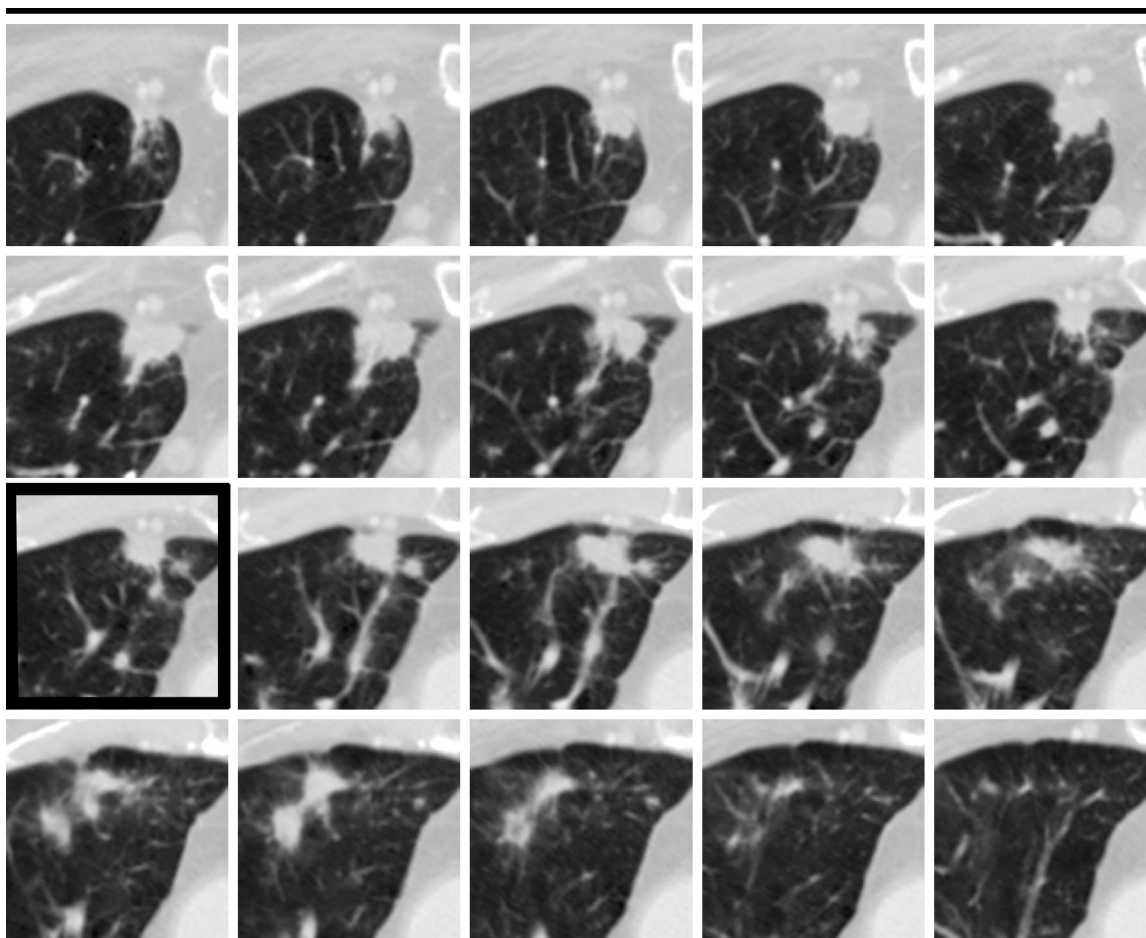
The fact that nodules must be “nodular” has led the LIDC to the seemingly simple yet challenging question, “what is a nodule?” In 1984, the Fleischner Society published a glossary of terms for thoracic radiology (45), in which a lung nodule was defined as “any pulmonary or pleural lesion represented in a radiograph by a sharply defined, discrete, nearly circular opacity 2–30 mm in diameter.” Twelve years later, the Fleischner Society published a glossary of terms specifically for thoracic CT (44) in which a lung nodule was defined as a “round opacity, at least moderately well margined and no greater than 3 cm in maximum diameter.” The Fleischner Society’s pathologic definition of a nodule as a “small, approximately spherical, circumscribed focus of abnormal tissue” (44) reflects the three-dimensional nature of the physical lesion manifested radiologically as a nodule on CT scans.

While these words provide an idealized definition of a lung nodule demonstrated on CT scans, the natural complexities of biologic systems render the practical application of such a definition more difficult. For example, most would agree that the lesion shown in Figure 1 satisfies the previous definition of a nodule. Moreover, the lesion shown in Figure 2 appears to satisfy the previous definition of a nodule when that single section is con-

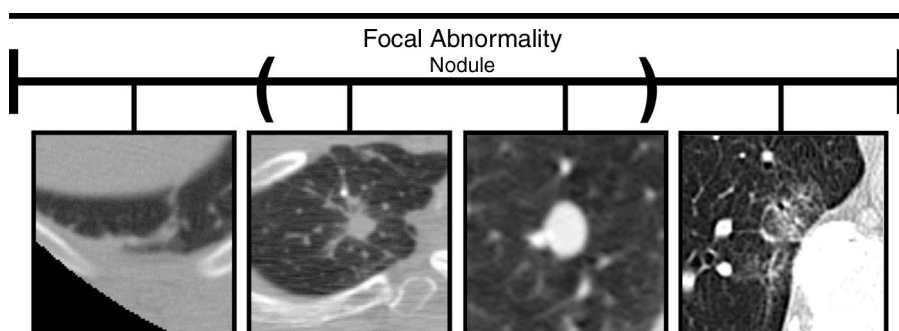
sidered in isolation. A lesion with an axial dimension greater than approximately the reconstruction interval used in the CT examination will, of course, appear in more than a single section. When the complete extent of the lesion demonstrated in the single section of Figure 2 is considered (Fig 3), the more complicated nature of this lesion becomes apparent. So the question remains, is this lesion a nodule?

The previous discussion makes clear that a utilitarian definition of a nodule may not be straightforward, since the notion of a nodule may not represent a single entity capable of verbal definition. The term *nodule* is more appropriately applied to a spectrum of abnormalities (irrespective of presumed histologic findings), which is itself a subset of a broader spectrum of abnormalities that we will term *focal abnormalities*, as illustrated in Figure 4. On the basis of this conceptualization, all nodules are focal abnormalities, but not all focal abnormalities are nodules.

For example, the linear scar shown on the far left in Figure 4 does not satisfy the shape criterion of a nodule, while the nonspecific opacity on the far right of Figure 4 does not satisfy the margination requirement. Two example nodules, one spiculated and irregularly shaped (Fig 4, left-center) and one compact (Fig 4, right-center), are shown for comparison. The axis label has intentionally been omitted from the schematic in Figure 4; the two spectra, in reality, span a multi-dimensional space that comprises lesion characteristics such as shape, texture, and margin sharpness. The LIDC is currently creating a visual nodule library



**Figure 3.** The lesion shown in Figure 2 is shown in all transverse sections in which it appears on the CT scan. The irregular shape of the three-dimensional lesion is apparent. The region shown with a black border was extracted from the section shown in Figure 2.



**Figure 4.** The term “nodule” refers to a spectrum of abnormalities that is itself part of a larger spectrum of abnormalities termed “focal abnormalities.” This scar (far left) is a focal abnormality that would not be considered a nodule, as is the nonspecific opacity (far right), which is likely due to previous infection. The spiculated lesion (left-center) and the compact lesion (right-center) are two representative examples of lesions within the nodule spectrum. All images were obtained from transverse CT examinations.

that seeks to capture image-based examples of lesions that span the spectrum of *nodule* and the superset spectrum of *focal abnormality*. With the benefit of this visual aid, we will recast a verbal definition of the term *nodule*. All nodules in the

database will be characterized by a panel of experienced thoracic radiologists according to nodule characteristics captured in the visual library so that panel concordance on these attributes may be captured for each nodule.

## TRUTH ASSESSMENT

The database, as previously discussed, will comprise more than a collection of CT images. In accordance with our mission, the LIDC will provide an assessment of truth along with the images in the database. We anticipate that investigators who use the database will require information regarding nodule spatial location and extent, as estimated from the CT images, and information regarding verified nodule diagnosis, which may be derived from pathologic specimens or appropriate radiologic follow-up. Ideally, both categories of information would be obtained for all nodules in the database; practical considerations, however, will limit our ability to achieve this goal.

### Spatial Location and Extent Estimates

A *qualifying nodule* is defined, for the purposes of the LIDC database, as a nodule that satisfies the nodule inclusion cri-

teria discussed previously; namely, it has an effective diameter of 3–30 mm and is considered to lie within the nodule spectrum. Each scan selected for the database (with the exception of the cohort of “normal” scans described in the Scan Inclusion Criteria section) will have at least one but, as a general rule, not more than six qualifying nodules. The outlines for each qualifying nodule will be obtained to record not only the position (eg, a centroid calculated from the radiologist outline) but also the spatial extent of the lesions that are the focus of the LIDC database. In addition, the approximate centers of nodules that are smaller than 3 mm and suspicious for cancer will be indicated by the LIDC radiologists; outlines for these small nodules will not be recorded.

A panel of thoracic radiologists will independently construct outlines encompassing each qualifying nodule. The necessity of this task for the completeness of the database and for the benefit of investigators who will use it was debated by the LIDC Steering Committee on several occasions. The decision to supply this information generated two important issues for consideration: First, the manner in which nodule outlines would be obtained must be established. Second, the extent of interradiologist variability should be anticipated. Recent research has convinced the LIDC to expect substantial interreader variability in both the detection and the outlining of nodules (39).

The task of outlining each nodule in every section in which it appears for every CT scan selected for the database is certainly daunting. Placement of rectangular bounding boxes around each nodule was initially considered as a means by which spatial extent information could be obtained; however, while such an approach would be useful for investigators developing nodule detection techniques, the LIDC recognized that nodule classification and segmentation (and perhaps other research interests) would not be served by such coarse information. With the need for nodule outlines established by means of consensus, the next decision was whether these outlines must be constructed manually or whether semiautomated methods could be used. The trade-off is between time and bias: Manual outlines would require a great deal of radiologist time, but they would reflect the unbiased expertise of radiologists. Semiautomated outlines are expected to improve the reproducibility of nodule outlines and require a smaller investment of



**Figure 5.** Lung nodule shown with outlines constructed by two radiologists independently. One radiologist constructed an outline that captured the core solid component of the nodule (black outline), while another radiologist constructed a less conservative outline that encompassed the nonsolid component, as well (white outline).

radiologist time, at least in some cases, but may exhibit bias introduced by the specific computer method used. Moreover, some may object to the development of computerized methods that are based on a truth established with the aid of another computerized method. The LIDC, through a series of pilot studies, is currently investigating the appropriateness of manual and semiautomated outlining techniques for the database.

The anticipated degree of interradiologist variability in the outlining task was a concern for the LIDC. Specifically, we were concerned that inconsistencies among the outlines constructed by different radiologists might render the resulting spatial extent estimates impractical for use by CAD researchers. Consider the nodule outlines in Figure 5. One radiologist constructed an outline that captured the core solid component of the nodule, while another radiologist independently constructed an outline that also encompassed the nonsolid component of the nodule. Both outlines, however, may be considered correct depending on the specified task, and both represent a valid interpretation of the nodule boundary.

The outlines in Figure 5 were constructed by two radiologists who were simply asked to outline the nodule. Differences in the interpretation of that general task are apparent. If, for example, the stated task was to outline the solid com-

ponent of the nodule, variability of task interpretation would likely be reduced, and both radiologists would likely have constructed outlines similar to the black outline of Figure 5. It is important to note that such outlines would still maintain a degree of variability because of radiologist interpretation of the edges of the nodule's solid component, but at least interpretation of the task would be more consistent. Accordingly, consensus on the specific outlining task is required to allow for a meaningful collection of nodule outlines. The variation in nodule outlines among the panel of radiologists will be included with the database to provide a statistical map describing the combined opinion of experienced readers. Nodule size will be derived from these nodule outlines.

An issue considered by the LIDC is the extent to which manual or semiautomated nodule outlines that are based on the CT scan actually represent the boundary of the physical nodule. This concern is greatest for outlines obtained from the upper- and lowermost sections that depict a nodule, since these sections exhibit the most pronounced partial volume effect because of the finite and anisotropic CT point spread function. Computerized nodule segmentation methods that might be developed by users of the database may be designed to compensate for these partial volume effects;

the evaluation of such methods would then be affected adversely by spatial extent estimates that are based on radiologic appearance. Nevertheless, the consensus of the LIDC was that the most appropriate nodule outlines for the database would encompass all "abnormal pixels" that either belong to or are derived from a nodule. The opinion and experience of each radiologist will dictate how he or she accommodates adjacent vessels, airways, or pleura.

In addition to nodules, all abnormal lesions or foci of abnormality that are larger than 3 mm will be identified by recording the approximate spatial location of the center of the lesion. As discussed previously, in addition to nodules, scans selected for the database may contain other focal abnormalities that do not lie within the nodule spectrum or disease that may not even be considered a focal abnormality, such as emphysema. The coarse identification of such regions that do not correspond to normal anatomic structures will provide a more complete inventory of pulmonary abnormalities on each scan and allow for a more thorough evaluation of CAD methods. To illustrate the latter point, a nodule detection scheme may identify as a nodule a lesion that is actually a scar. While such a finding would be counted as false-positive with regard to the specific nodule detection task, it is distinctly different from false-positive findings caused by pulmonary vessels, since the scar is indeed an anomaly. In fact, investigators may wish to use the complete truth assessment that will be offered with the database to explicitly develop CAD algorithms that can be used to identify scars or classify abnormal findings as a scar or a nodule.

### Verified Diagnosis

Every effort will be made to obtain a verified diagnosis for each qualifying nodule as either cancerous or noncancerous when it becomes available through chart review. This information will be essential for investigators developing automated lung nodule classification techniques. In addition, the pathologic subtype of nodules (both cancerous and noncancerous), which is based on the basis of the World Health Organization classification (46), will be extracted from patient charts and included with the database. To confirm the patient records, the pathology reports for all patients and any available histopathologic slides or surgical specimens for a subset of the pa-

tients (5%) will be independently reviewed by the LIDC panel of experienced pathologists. We recognize that many scans may not have available pathologic information; furthermore, we recognize that not all nodules on any one scan will have such information. The database will document the scans and nodules for which a verified diagnosis has been obtained with fine-needle aspiration biopsy, surgical resection, or extended radiologic observation in which no growth is demonstrated over a 2-year period (47).

### PROCESS MODEL

The LIDC has developed a process model, according to which various aspects of the database will be gathered. This process model will govern the mechanics of identifying appropriate CT scans, assessing truth at the various levels previously described, cataloging the scans and nodules, recording pertinent information about each nodule in the database, and orchestrating these activities across the five geographically disparate institutions. We are steadily resolving the technical challenges posed by a task of this magnitude.

The process model details the steps required to translate a CT scan acquired for the clinical evaluation of a patient into a viable element of the database. With appropriate local institutional review board approval, radiologists at each of the five institutions will identify thoracic CT scans from their clinical caseload or image archive, which may include images acquired as part of a research study such as the Early Lung Cancer Action Program (5) or the National Lung Screening Trial (48) that meet the previously discussed inclusion criteria established by the LIDC. Radiologists will then evaluate image quality and any artifacts that may be present. Informed consent procedures will also be followed, as required by each institutional review board.

Identification of a candidate scan will set in motion a sequence of events that begins at the local institution and then extends to the other four institutions. The scan will be transferred to the local research computer, where it will be catalogued either manually by a local data manager or automatically through a local DICOM receiver. This step is critical to ensure that the local institution has the ability to monitor any subsequent imaging or pathologic data acquired for that patient. Software will be applied so that all protected health information con-

tained in the DICOM receiver header of the image will be removed in accordance with Health Insurance Portability and Accountability Act, or HIPPA, guidelines (38).

This scan then will be made available to the other four institutions for assessment of spatial location and extent. Rather than a forced consensus panel approach, the LIDC has opted for a panel that uses a combination of blinded and unblinded reviews by multiple radiologists (eg, one radiologist at each of the other four institutions) to establish estimates of nodule spatial location and extent. The blinded and unblinded reviews are both part of the same process in which radiologists attempt to identify, as completely as possible, all nodules on a scan. In this approach, the designated radiologist at each site first performs a blinded review of the scan by identifying the spatial location and radiologic characteristics of all abnormalities on the scan that are larger than 3 mm, as measured with electronic calipers. Nodules smaller than 3 mm that the radiologist deems suspicious for cancer are also identified.

Other information collected for each lesion will include lesion type (eg, scar or nodule), the radiologist's subjective level of confidence that the lesion represents a focal abnormality in general or a nodule more specifically, radiologic texture (eg, solid, part solid, or nonsolid) if considered a nodule, a five-point lesion subtlety score (ranging from "obvious" to "extremely subtle"), presence of calcifications, and lobar location. Furthermore, outlines will be constructed for all qualifying nodules on calibrated monitors with magnification capabilities and an initial window and level setting of 1500 HU and -500 HU, respectively, which may then be adjusted by the radiologist on the basis of individual preference. Additional subjective assessments of characteristics such as shape and margin will be recorded. If the patient has one or more scans already in the database, a reconciliation of lesions will be performed to identify the same lesion across multiple scans. These data will be recorded in the fields of the relational database. During this initial review, each radiologist will interpret the scan independently from the radiologists at the other four institutions; hence, we use the term *blinded review*.

Once the four radiologists (each from a different institution) have performed the blinded review, the results of the blinded review of each radiologist will be made

available to all of the other radiologists who reviewed the scan. Each radiologist will then perform an unblinded review of the scan with the additional information provided by the other radiologists. During this unblinded review, the radiologists will review all marked structures (eg, their own markings, as well as the markings of the other radiologists who reviewed the scan) and decide whether to include each marked structure as a nodule. It is important to note that a forced consensus will not be imposed; rather, all of the nodules indicated by the reviewing radiologists will be tallied and recorded in the database.

Information obtained from all radiologists during both blinded and unblinded reviews will be included in the database to provide a rich source of data for investigators. For example, nodules recorded by only two of the radiologists during the blinded review will constitute a different detection target than nodules initially identified by all radiologists during the blinded review. Even more interesting might be nodules recorded by only two of the radiologists during the blinded review and then recorded by only the same two radiologists during the unblinded review, which implies that other radiologists observed this structure and declined to consider it a nodule. As another example, the spatial extent of a nodule may be described in probabilistic terms on the basis of the number of radiologists' outlines that encompass each pixel.

## ASSESSMENT METHODS

Development of the LIDC database must necessarily anticipate the ultimate use of CT scans and associated data by CAD investigators. Statistical power is an important concern. To benefit CAD research, the database must provide a sufficient number and mix of scans to enable meaningful conclusions to be drawn from the results of studies in which these scans are used. This presents the LIDC with a dilemma regarding the issue of database sizing (49); an intriguing aspect of the database is that we cannot fully anticipate its potential uses, and any discussion of assessment methods is inherently task dependent. For practical simplicity, we may consider the two general tasks of nodule detection and nodule classification and construct a statistical framework around these two lines of research. Even this coarse stratification is not without its challenges, since nodule detection and nodule classification have

different requirements in terms of truth information and performance evaluation. Ultimately, the data collection process will limit the population of the database: A limit exists on the number of scans that may be practically collected and subjected to the various levels of evaluation at the five institutions. A pilot study is planned that will allow us to better understand the case number issue.

An unbiased estimate of algorithm performance and a meaningful comparison of CAD methods necessitates the consideration of a number of important factors. The LIDC research group will inherently establish case distribution, lesion subtlety, and truth assessment. Other factors, such as the effect of scoring methods on reported CAD performance (50), the appropriateness of different training and testing paradigms (51), and the proper use of various task-dependent evaluation metrics, will remain at the discretion of the investigators who use the database. An initial survey of some of these essential issues as they pertain to lung nodule detection in CT has been reported recently by Dodd et al (49); the LIDC intends to further explore the intricacies of these issues and document the consensus on the most appropriate approaches for a variety of CAD tasks. Moreover, the LIDC will provide suggestions, references, and pointers to publicly available software, when appropriate.

## DISCUSSION POINTS

One factor that presents a constant challenge to any effort such as this is timeliness. As the LIDC endeavors to create a resource for the medical imaging research community, CT scanner technology continues to evolve; thus, as we take the time required to formulate, consider, and resolve fundamental issues, the technologic landscape continues to change. For example, the decision to accept CT scans acquired with section thickness as thick as 5 mm was made during LIDC Steering Committee meetings in 2001 at a time when 2.5-mm section thickness was generally regarded as state-of-the-art. Three years later, however, 5-mm section thickness apparently lags two steps behind the 1.25-mm section thickness possible with state-of-the-art scanners. Although the potential of CAD is expected to improve for CT scans with increased transverse resolution, we recognize that many medical centers worldwide still routinely acquire thoracic CT scans with 5-mm or larger section thickness. A bal-

ance must be maintained to ensure that the database is not obsolete before it becomes available and that the task of creating the database is not paralyzed by the pace of evolving technology.

Not all scans in the database will be useful for all CAD research activities, given the reality that not all desired data will be available for all nodules or all scans. For investigators pursuing nodule classification techniques based on volumetric analysis of sequential scans, only qualifying nodules that appear in at least two scans from the same patient within the database and that include associated pathologic information or demonstrate stability over a defined period of time will prove useful. For investigators pursuing nodule classification techniques based on nodule features within a single scan, the requirement for at least two scans may be removed, but pathologic information is still required. Nodule segmentation techniques may make use of all qualifying nodules, regardless of whether pathologic information is available; however, nodules smaller than 3 mm in diameter will not have appropriate spatial extent estimates for such analyses. Nodule detection techniques will benefit mainly from all qualifying nodules on full-thorax scans, since qualifying nodules on limited-anatomy scans will not be appropriate for the detection task. The database will be organized in such a manner that the relevance of each scan or each nodule with regard to various CAD tasks will be specified.

The nodule outlining process truly has no reference standard. Adjudication among the radiologists is unnecessary, since we intentionally want to provide a range of outlines based on the reality of differences of opinion among experienced chest radiologists. The variation in nodule outlines will provide a statistical map of nodule spatial extent estimates. This variability will allow researchers to perform a variety of interesting analyses. For example, if a CAD algorithm generates an estimate of the nodule border that provides a measure of edge texture or, with follow-up scans, a measure of nodule growth that allows prediction of the pathologic status of a lung nodule better than the outlines produced by the panel of radiologists, then the performance of that algorithm may be considered to exceed, in some sense, the accuracy of the radiologists' border estimates. If the CAD algorithm generates an estimate of the nodule border that lies within the range of the radiologists' outlines, then the algorithm is at least con-



sistent with the interobserver variability of the radiologist panel.

A database such as this presents an opportunity for truly blinded evaluation of CAD techniques. As an example of such an evaluation paradigm for the registration accuracy of brain images, the Retrospective Image Registration Evaluation project at Vanderbilt University (52) allows research groups to download CT, magnetic resonance, and positron emission tomography image data sets, to which the groups may apply their own retrospective image registration techniques; the results are then submitted to the central project site for comparison with a prospective marker-based registration standard. Since results based on this standard are not revealed to the research groups, direct comparison of different registration techniques may be captured in a controlled manner. The analogy for the LIDC database would be the segregation of dedicated image training sets and test sets. In this scenario, investigators would only have access to designated training images for the development of their CAD techniques; the final technique would be applied to test images, which were not previously available to the investigators. The LIDC Steering Committee, however, decided against such a segregation of cases. The main reasons for this decision involved the limitations that would be imposed on investigators' use of the database and the inability of the LIDC to anticipate the full range of applications for which investigators will use the database.

The public release of a database such as this is an important issue of timing. Ideally, the complete database containing all CT scans and verified diagnoses for all patients whose scans are included would be made available at the end of this project. We are aware that this approach, in practice, would severely limit the CAD research community; as the LIDC strives to compile the complete database, many researchers are eager to accept an initial collection of CT scans even without the follow-up scans and verified diagnoses that will eventually become available. To accommodate the larger goal of facilitating CAD research, we anticipate an initial release of approximately 100 CT scans by the end of 2004. These cases will be processed through the infrastructure we are creating so that a scan from the clinical workflow at a member institution can be placed into the central LIDC database. As previously described, this process includes identification of a qualifying scan, deletion of information that could be

used to identify the patient, multiple reader and site estimates of the spatial location and extent of all nodules with blinded and unblinded reviews, acquisition of information for all database fields, and final cataloging.

The complete contents of the database will become openly available without restriction to the medical imaging research community and will be stored at and managed by the NCI. Image data sets will probably be provided on digital video disks or DVDs at the request of investigators. Although search requests and database descriptors will be available through the Web, Web-based retrieval of image data may be impractical because of the size of the image files. Details of the distribution process, including potential fees, remain to be evaluated by the LIDC. The burden of monitoring the use of the database will then shift to individual investigators and the scientific community through the peer-review process both for publications and for grants. Investigators and journal editors will be responsible for disclosing and grant review study sections will be responsible for demanding details on the precise manner in which image and clinical data were used. For example, the training and/or testing paradigm should be described so that other investigators may repeat the study on the same scans with their own methods. The LIDC intends to provide guidance but does not intend to impose restrictions on what should be reported in terms of important training and test set characteristics, such as size and subtlety. We fully expect that the success of the database and its effect on the community will become evident through the literature.

## SUMMARY

To stimulate the advancement of CAD research for lung nodules in thoracic CT scans, the NCI launched a cooperative effort known as the LIDC. The LIDC consists of five academic institutions from across the United States that are working together to develop an image database that will serve as an international research resource for the development, training, and evaluation of CAD methods for lung nodules on CT scans. Prior to the collection of CT images and associated patient data, the LIDC has been engaged in a consensus process to identify, address, and resolve a host of challenging technical and clinical issues to provide a solid foundation for a scientifically robust database. These issues include the

establishment of (a) a governing mission statement, (b) criteria to determine whether a CT scan is eligible for inclusion in the database, (c) an appropriate definition of the term *qualifying nodule*, (d) an appropriate definition of truth requirements, (e) a process model through which the database will be populated, and (f) a statistical framework to guide the application of assessment methods by users of the database. Through a consensus process in which careful planning and the proper consideration of fundamental issues has been emphasized, the LIDC database is expected to provide a powerful resource for the medical imaging research community.

**Author affiliations:** University of Chicago, Ill (H.M.); University of Iowa, Iowa City (G.M., E.A.H.); University of California, Los Angeles (M.F.M.G., D.R.A.); University of Michigan, Ann Arbor (C.R.M., E.A.K.); Cornell University, New York, NY (D.Y., C.I.H., A.P.R.); National Cancer Institute, Bethesda, Md (B.Y.C., L.P.C., L.E.D.); University of Pittsburgh, Pa (D.G.).

**Members of the LIDC Research Group:** Lori E. Dodd, PhD, David Gur, ScD, Nicholas A. Petrick, PhD, Edward Staab, MD, Daniel C. Sullivan, MD, Robert F. Wagner, PhD, Peyton H. Bland, PhD, Keith Brautigam, BA, Matthew S. Brown, PhD, Barry DeYoung, MD, Roger M. Engelmann, MS, Andinet A. Enquobahrie, MS, Carey E. Floyd, Jr, PhD, Junfeng Guo, PhD, Aliya N. Husain, MD, Gary E. Laderach, BS, Charles E. Metz, PhD, Brian Mullan, MD, Richard C. Pais, BS, Christopher W. Piker, BS, James W. Sayre, DrPH, Adam Starkey.

## References

- Giger ML. Current issues in CAD for mammography. In: Doi K, Giger ML, Nishikawa RM, Schmidt RA, eds. Digital mammography '96: proceedings of the 3rd international workshop on digital mammography. Amsterdam, the Netherlands: Elsevier, 1996; 53-59.
- van Ginneken B, ter Haar Romeny BM, Viergever MA. Computer-aided diagnosis in chest radiography: a survey. IEEE Trans Med Imaging 2001; 20:1228-1241.
- Rubin GD. Data explosion: the challenge of multidetector-row CT. Eur J Radiol 2000; 36:74-80.
- Jhaveri KS, Saini S, Levine LA, et al. Effect of multislice CT technology on scanner productivity. AJR Am J Roentgenol 2001; 177:769-772.
- Henschke CI, McCauley DI, Yankelevitz DF, et al. Early lung cancer action project: overall design and findings from baseline screening. Lancet 1999; 354:99-105.
- Kaneko M, Eguchi K, Ohmatsu H, et al. Peripheral lung cancer: screening and detection with low-dose spiral CT versus radiography. Radiology 1996; 201:798-802.
- Sone S, Takashima S, Li F, et al. Mass screening for lung cancer with mobile spiral computed tomography scanner. Lancet 1998; 351:1242-1245.
- Li F, Sone S, Abe H, MacMahon H, Armato SG 3rd, Doi K. Lung cancers missed at low-dose helical CT screening in a general

- population: comparison of clinical, histopathologic, and imaging findings. *Radiology* 2002; 225:673-683.
9. Gurney JW. Missed lung cancer at CT: imaging findings in nine patients. *Radiology* 1996; 199:117-122.
  10. White CS, Romney BM, Mason AC, Austin JH, Miller BH, Protopapas Z. Primary carcinoma of the lung overlooked at CT: analysis of findings in 14 patients. *Radiology* 1996; 199:109-115.
  11. Kakinuma R, Ohmatsu H, Kaneko M, et al. Detection failures in spiral CT screening for lung cancer: analysis of CT findings. *Radiology* 1999; 212:61-66.
  12. Giger ML, Bae KT, MacMahon H. Computerized detection of pulmonary nodules in computed tomography images. *Invest Radiol* 1994; 29:459-465.
  13. Kanazawa K, Kubo M, Niki N, et al. Computer assisted lung cancer diagnosis based on helical images. In: Chin RT, Ip HHS, Naiman AC, Pong TC, eds. *Image analysis applications and computer graphics: proceedings of the third international computer science conference*. Berlin, Germany: Springer-Verlag, 1995; 323-330.
  14. Ryan WJ, Reed JE, Swensen SJ, Sheedy PF Jr. Automatic detection of pulmonary nodules in CT. In: Lemke HU, Vannier MW, Inamura K, Farman AG, eds. *Computer assisted radiology (CARS '96)*. Amsterdam, the Netherlands: Elsevier Science, 1996; 385-389.
  15. Okumura T, Miwa T, Kako J, et al. Image processing for computer-aided diagnosis of lung cancer screening system by CT (LSCT). *Proc SPIE* 1998; 3338:1314-1322.
  16. Fiebich M, Wietholt C, Renger BC, et al. Automatic detection of pulmonary nodules in low-dose screening thoracic CT examinations. *Proc SPIE* 1999; 3661:1434-1439.
  17. Lou SL, Chang CL, Lin KP, Chen TS. Object-based deformation technique for 3-D CT lung nodule detection. *Proc SPIE* 1999; 3661:1544-1552.
  18. Taguchi H, Kawata Y, Niki N, et al. Lung cancer detection based on helical CT images using curved surface morphology analysis. *Proc SPIE* 1999; 3661:1307-1314.
  19. Satoh H, Ukai Y, Niki N, et al. Computer aided diagnosis system for lung cancer based on retrospective helical CT image. *Proc SPIE* 1999; 3661:1324-1335.
  20. Armato SG 3rd, Giger ML, MacMahon H. Automated detection of lung nodules in CT scans: preliminary results. *Med Phys* 2001; 28:1552-1561.
  21. Brown MS, McNitt-Gray MF, Goldin JG, Suh RD, Sayre JW, Aberle DR. Patient-specific models for lung nodule detection and surveillance in CT images. *IEEE Trans Med Imaging* 2001; 20:1242-1250.
  22. Lee Y, Hara T, Fujita H, Itoh S, Ishigaki T. Automated detection of pulmonary nodules in helical CT images based on an improved template-matching technique. *IEEE Trans Med Imaging* 2001; 20:595-604.
  23. Gurcan MN, Sahiner B, Petrick N, et al. Lung nodule detection on thoracic computed tomography images: preliminary evaluation of a computer-aided diagnosis system. *Med Phys* 2002; 29:2552-2558.
  24. Fan L, Novak CL, Qian J, Kohl G, Naidich DP. Automatic detection of lung nodules from multi-slice low-dose CT images. *Proc SPIE* 2001; 4322:1828-1835.
  25. Ko JP, Betke M. Chest CT: automated nodule detection and assessment of change over time—preliminary experience. *Radiology* 2001; 218:267-273.
  26. Fiebich M, Wormanns D, Heindel W. Improvement of method for computer-assisted detection of pulmonary nodules in CT of the chest. *Proc SPIE* 2001; 4322:702-709.
  27. Cavouras D, Prassopoulos P, Pantelidis N. Image analysis methods for solitary pulmonary nodule characterization by computed tomography. *Eur J Radiol* 1992; 14:169-172.
  28. Henschke CI, Yankelevitz DF, Mateescu I, Brettle DW, Rainey TG, Weingard FS. Neural networks for the analysis of small pulmonary nodules. *Clin Imaging* 1997; 21:390-399.
  29. Kawata Y, Niki N, Ohmatsu H, et al. Quantitative surface characterization of pulmonary nodules based on thin-section CT images. *IEEE Trans Nucl Sci* 1998; 45:2132-2138.
  30. Tozaki T, Kawata Y, Niki N, et al. Pulmonary organs analysis for differential diagnosis based on thoracic thin-section CT images. *IEEE Trans Nucl Sci* 1998; 45:3075-3082.
  31. McNitt-Gray MF, Hart EM, Wyckoff N, Sayre JW, Goldin JG, Aberle DR. A pattern classification approach to characterizing solitary pulmonary nodules imaged on high resolution CT: preliminary results. *Med Phys* 1999; 26:880-888.
  32. McNitt-Gray MF, Wyckoff N, Sayre JW, Goldin JG, Aberle DR. The effects of co-occurrence matrix based texture parameters on the classification of solitary pulmonary nodules imaged on computed tomography. *Comput Med Imaging Graph* 1999; 23:339-348.
  33. Reeves AP, Kostis WJ, Yankelevitz DF, Henschke CI. Three-dimensional shape characterization of solitary pulmonary nodules from helical CT scans. In: Lemke HU, Vannier MW, Inamura K, Farman AG, eds. *Computer assisted radiology and surgery (CARS '99)*. Amsterdam, the Netherlands: Elsevier Science, 1999; 83-87.
  34. Kawata Y, Niki N, Ohmatsu H, et al. Quantitative analysis of internal texture for classification of pulmonary nodules in three-dimensional thoracic images. *Proc SPIE* 2000; 3979:863-871.
  35. Kawata Y, Niki N, Ohmatsu H, et al. Computer-aided differential diagnosis of pulmonary nodules based on a hybrid classification approach. *Proc SPIE* 2001; 4322:1796-1806.
  36. Minami K, Kawata Y, Niki N, et al. Computerized characterization of contrast enhancement patterns for classifying pulmonary nodules. *Proc SPIE* 2001; 4322:1936-1943.
  37. Armato SG 3rd, Altman MB, Wilkie J, et al. Automated lung nodule classification following automated nodule detection on CT: a serial approach. *Med Phys* 2003; 30:1188-1197.
  38. Department of Health and Human Services. Standards for privacy of individually identifiable health information: final rules. *Federal Register* 2002; 67:53182-53272.
  39. Rogers SR, Brown MS, Goldin JG, Suh RD, Kim HJ, Aberle DR. Automated lung nodule detection in CT: nodule inclusion criteria for determining ground truth (abstr). *Radiology* 2002; 225(P):407.
  40. Vannier MW, Summers RM. Sharing images. *Radiology* 2003; 228:23-25.
  41. Vannier MW, Staab EV, Clarke LP. Cancer imaging informatics workshop report from the biomedical imaging program of the National Cancer Institute. *Acad Radiol* 2003; 10:798-802.
  42. National Cancer Institute. NIH guide archive ending the week of April 14, 2000. Available at: <http://grants1.nih.gov/grants/guide/rfa-files/RFA-CA-01-001.html>. Accessed September 26, 2003.
  43. Clarke LP, Croft BY, Staab E, Baker H, Sullivan DC. National Cancer Institute initiative: lung image database resource for imaging research. *Acad Radiol* 2001; 8:447-450.
  44. Austin JH, Mueller NL, Friedman PJ, et al. Glossary of terms for CT of the lungs: recommendations of the Nomenclature Committee of the Fleischner Society. *Radiology* 1996; 200:327-331.
  45. Tuddenham WJ. Glossary of terms for thoracic radiology: recommendations of the nomenclature committee of the Fleischner Society. *AJR Am J Roentgenol* 1984; 143:509-517.
  46. Travis WD, Colby TV, Corrin B, Shimamoto Y, Brambilla E, Sobin LH. *Histological typing of lung and pleural tumors: World Health Organization international histological classification of tumours series*. Berlin, Germany: Springer-Verlag, 1999.
  47. Yankelevitz DF, Gupta R, Zhao B, Henschke CI. Small pulmonary nodules: evaluation with repeat CT—preliminary experience. *Radiology* 1999; 212:561-566.
  48. Hillman B. Economic, legal, and ethical rationales for the ACRIN national lung screening trial of CT screening for lung cancer. *Acad Radiol* 2003; 10:349-350.
  49. Dodd LE, Wagner RF, Armato SG 3rd, 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* 2004; 11:462-475.
  50. Nishikawa RM, Yarusso LM. Variations in measured performance of CAD schemes due to database composition and scoring protocol. *Proc SPIE* 1998; 3338:840-844.
  51. Tourassi GD, Floyd CE Jr. The effect of data sampling on the performance evaluation of artificial neural networks in medical diagnosis. *Med Decis Making* 1997; 17:186-192.
  52. West J, Fitzpatrick JM, Wang MY, et al. Comparison and evaluation of retrospective intermodality brain image registration techniques. *J Comput Assist Tomogr* 1997; 21:554-566.