研究成果の刊行に関する一覧表

著者氏名	論文タイトル名	発表誌名	巻号	ページ	出版年
Armato SG III, Altman MB, Wilkie J, Sone S, Li F, Doi K.	Automated lung nodule classification following automated nodule detection on CT: A serial approach.	Med Phys	30(6)	1188-1197	2003

Automated lung nodule classification following automated nodule detection on CT: A serial approach

Samuel G. Armato III,[a] Michael B. Altman, and Joel Wilkie
Department of Radiology, The University of Chicago, 5841 South Maryland Avenue, Chicago, Illinois 60637

Shusuke Sone
JA Azumi General Hospital, Nagano, Japan

Feng Li, Kunio Doi, and Arunabha S. Roy
Department of Radiology, The University of Chicago, 5841 South Maryland Avenue, Chicago, Illinois 60637

(Received 20 June 2002; accepted for publication 12 March 2003; published 30 May 2003)

We have evaluated the performance of an automated classifier applied to the task of differentiating malignant and benign lung nodules in low-dose helical computed tomography (CT) scans acquired as part of a lung cancer screening program. The nodules classified in this manner were initially identified by our automated lung nodule detection method, so that the output of automated lung nodule detection was used as input to automated lung nodule classification. This study begins to narrow the distinction between the "detection task" and the "classification task." Automated lung nodule detection is based on two- and three-dimensional analyses of the CT image data. Gray-level-thresholding techniques are used to identify initial lung nodule candidates, for which morphological and gray-level features are computed. A rule-based approach is applied to reduce the number of nodule candidates that correspond to non-nodules, and the features of remaining candidates are merged through linear discriminant analysis to obtain final detection results. Automated lung nodule classification merges the features of the lung nodule candidates identified by the detection algorithm that correspond to actual nodules through another linear discriminant classifier to distinguish between malignant and benign nodules. The automated classification method was applied to the computerized detection results obtained from a database of 393 low-dose thoracic CT scans containing 470 confirmed lung nodules (69 malignant and 401 benign nodules). Receiver operating characteristic (ROC) analysis was used to evaluate the ability of the classifier to differentiate between nodule candidates that correspond to malignant nodules and nodule candidates that correspond to benign lesions. The area under the ROC curve for this classification task attained a value of 0.79 during a leave-one-out evaluation. [DOI: 10.1118/1.1573210]

Key words: computer-aided diagnosis (CAD), lung cancer, lung nodules, computed tomography, image processing

I. INTRODUCTION

Low-dose helical computed tomography (CT) is gaining an international reputation as a viable lung cancer screening tool. Several studies have shown the increased number and stage shift of cancers detected at low-dose screening relative to screening with chest radiography[1-3] and relative to the onset of clinical symptoms. These same studies have demonstrated, however, that lung cancer screening promises to generate a large number of findings that deserve follow-up and scrutiny. Moreover, screening programs will produce large amounts of image data for radiologist interpretation, especially with the development of multislice scanners. This scenario poses two challenges for today's radiologist; namely, (1) a greater time commitment for the review of screening studies to identify all abnormal lesions and (2) an increased burden to distinguish between nodule features that likely differentiate malignant nodules from benign lesions both to reduce patient anxiety and to minimize unnecessary follow-up procedures and biopsies. The incorporation of computerized image analysis may address both of these challenges.

A diagnostic task in radiology generally has two components: first, the presence of an abnormality must be appreciated, and second, some notion of severity (e.g., potential malignancy) of the lesion must be evaluated. Research in computer-aided diagnosis (CAD) has followed this diagnostic paradigm. The automated "detection task," however, has traditionally been considered separate from the automated "classification task" (see Fig. 1), with researchers exploring the two components of CAD more or less independently.

Investigators have developed a number of computerized nodule detection techniques.[4-18] One of the first reported methods was that of Giger *et al.*,[4] who developed an automated nodule detection scheme based on geometric feature analysis in conjunction with multiple gray-level thresholding. Other investigators have followed with gray-level-thresholding-based methods for the automated detection of lung nodules.[8,12,14-17] Fuzzy clustering algorithms,[5] spatial

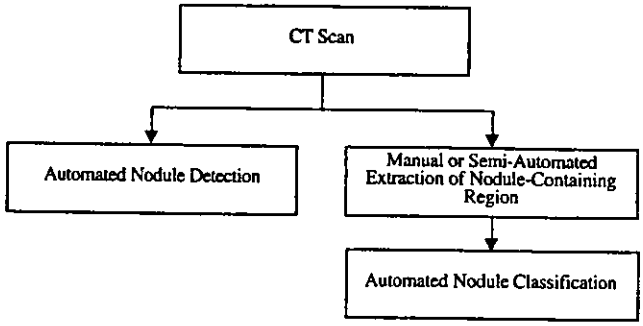

FIG. 1. Schematic illustrating the conventional approach to CAD in which the automated detection task is independent of the automated classification task, which typically begins with the manual indication of a nodule-containing region.

filtering,[7] template matching,[18] object-based deformation procedures,[9] morphological analysis,[11] and model-based techniques[6,13] have also been reported as viable means of detecting lung nodules.

Our automated lung nodule detection method[19–21,12] provided the detection results on which the present study was based. This method considers the three-dimensionality of the CT image data and identifies nodule candidates based on gray-level-thresholding techniques. Linear discriminant analysis is then used to categorize identified nodule candidates as either nodule or non-nodule. Recent results report an overall nodule detection sensitivity of 70% with 1.5 false-positive detections per section on a database of 43 diagnostic CT scans,[12] 89% sensitivity with 1.3 false positives per section on the subset of 20 cases (from the 43-case database) with, at most, two nodules,[12] and 84% overall detection sensitivity for lung nodules that represented lung cancers "missed" by radiologists during the clinical interpretation of the low-dose screening study with 1.0 false positives per section.[22] It should be noted that these previous studies all involved CT scans with 10-mm collimation and a 10-mm reconstruction interval.

Automated nodule detection schemes indicate the spatial locations of suspected abnormalities but do not then endeavor to estimate the likelihood of malignancy. Research in the computerized *classification* of lung nodules as benign or malignant, however, has received attention from a number of investigators as well.[23–32] Cavouras *et al.*[23] selected the CT section that contained the center of a solitary pulmonary nodule in each of 95 patients. Twenty features were computed for each nodule based on the CT density matrix, and these feature values were input to a least-squares-minimum-density-distance classifier. In a jackknife evaluation, the system achieved an overall accuracy of 90.2% in the task of distinguishing between benign and malignant nodules. McNitt-Gray *et al.*[27,28] developed a semiautomated contouring tool to isolate lung nodules within high-resolution CT images. Based on the quantitative features of nodule density, density distribution, size, shape, and texture derived from these contours and the pixels they enclosed, a pattern classification approach was used to classify nodules as benign or malignant with an accuracy of 90.3% based on a jackknife analysis.[27] An extension of this work based on co-occurrence matrices and texture measures yielded an accuracy of 90.6%.[28] Henschke *et al.*[24] applied an artificial neural network developed for automatic target recognition to the task of lung nodule classification. For each nodule, four CT section images obtained during the biopsy procedure were used as input to the neural network, which computed a measure of similarity between a model and each image location. A cross-validation study resulted in the correct classification of all 14 malignant nodules and 11 of 14 benign nodules.

Kawata *et al.*[25] used interactively selected cubic regions containing lung nodules from high-resolution CT images as input to an automated nodule segmentation technique based on three-dimensional deformable surface models. Surface characteristics based on surface curvature measures and surrounding line patterns were computed for benign and malignant nodules. Kawata *et al.*[30] then included histogram and topological features along with surface characteristic features as input to a linear discriminant classifier. Leave-one-out analysis yielded an area under the receiver operating characteristic (ROC) curve of 0.97 in the task of classifying 128 lung nodules as benign or malignant. These same investigators later developed a hybrid classification approach that combined an unsupervised *k*-means clustering procedure with a linear discriminant classifier.[31] The initial clustering algorithm segregated nodules into two classes based predominantly on margin characteristics. A linear discriminant classifier was designed for each class separately based on internal structure features. The hybrid classification approach achieved an area under the ROC curve of 0.97 when applied to 248 lung nodules.

Minami *et al.*[32] reported an improvement in nodule classification with refined nodule segmentation techniques and the incorporation of nodule contrast enhancement information. Reeves *et al.*[29] explored shape metrics for the classification of solitary pulmonary nodules. Tozaki *et al.*[26] applied Gaussian curvature, thresholding, and edge-detection techniques to automatically extract pulmonary vessels and bronchi from high-resolution CT scans. The number of pulmonary vessels that appeared connected to interactively selected lung nodules and the angle that the vessels made with the surface of these nodules were shown to differ among benign and malignant nodules.

As these examples from the literature demonstrate, automated nodule classification schemes typically begin with manual or semiautomated extraction of a pulmonary structure known to represent an abnormality (Fig. 1). Thus, in an effort to focus development and testing of these techniques to the task of classifying lesions as benign or malignant, the initial challenge of automated detection is circumvented. This approach is quite practical and represents a very real clinical application of such automated methods in which a radiologist, after recognizing a suspicious lesion in a CT scan, might wish to select that lesion and have the computer provide some likelihood-of-malignancy estimate. An equally real clinical application of CAD will be a system that both indicates the location of (i.e., detects) and the likelihood of

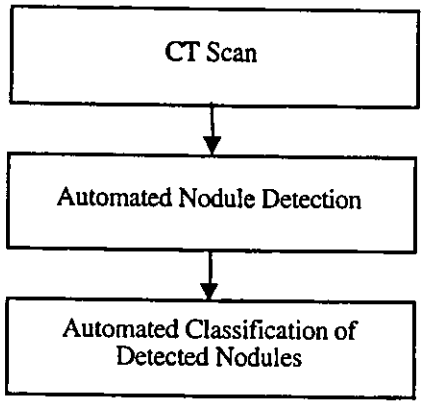

FIG. 2. Schematic illustrating the serial approach to automated lung nodule classification in which output from an automated nodule detection method provides the input for automated nodule classification.

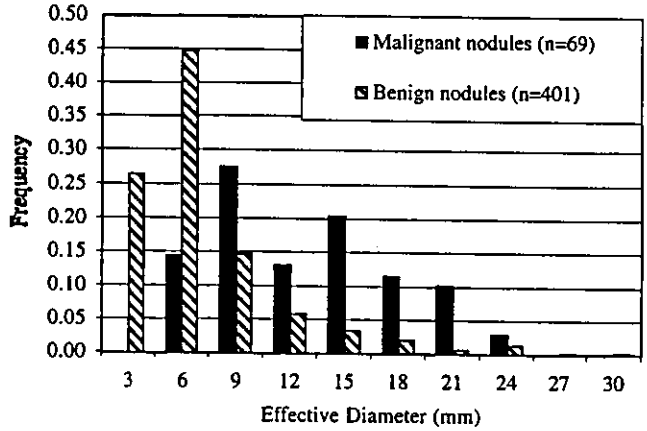

FIG. 3. Histogram representing the distributions of effective diameter of malignant and benign lung nodules in the database.

malignancy of (i.e., classifies) a lung nodule (Fig. 2). Although such an approach has been reported previously by Jiang *et al.*[33] for clustered microcalcifications in mammograms, in this study we present the beginnings of such a combined approach for lung nodules in CT.

II. MATERIALS

The database used in this study contained 393 noninfused, low-dose thoracic helical CT scans acquired from distinct patients (172 females, 221 males) who voluntarily participated in a lung cancer screening program in Nagano, Japan between 1996 and 1998.[2,34] Informed consent was obtained from all patients. The age range of the patients at the time of each scan was 30–88 years (mean: 63 years). The CT examinations were performed on a mobile CT scanner (CT-W950SR; Hitachi Medical, Tokyo, Japan) and were acquired with a low-dose protocol of 25 mA (232 scans) or 50 mA (161 scans), 2:1 helical pitch, 120 kVp, and 10-mm collimation with a 10-mm reconstruction interval.[34] The pixel size was 0.586 mm for 278 scans and 0.684 mm for 115 scans with a 512×512-pixel image matrix size. After sections representing anatomy inferior to the lung bases were manually excluded from each scan, the number of remaining sections per scan ranged from 21 to 33 (mean: 28 sections per scan). Image data were transferred from optical disk storage media to the research computer (SGI Onyx; Silicon Graphics, Inc., Mountain View, CA) used for this study.

These 393 scans were obtained from a larger database of studies from the Nagano screening program and represented all cases from the screening program that contained only confirmed malignant lung nodules and/or confirmed benign lung nodules. In other words, at least one lung nodule was present in each of the 393 cases, and the malignancy status of every nodule in all scans was known. The 393-case database contained a total of 470 lung nodules, of which 69 were malignant, as determined by biopsy, and 401 were benign, as determined by biopsy or by follow-up over a period of two years. The effective diameters of the lung nodules were determined from bounding boxes fit around the nodules by an experienced chest radiologist (Feng Li) and ranged from 3 to 30 mm (mean: 8.9 mm). Effective diameter was computed as the mean bounding box dimension (one-half the sum of the short and long axis lengths). When the nodule was present in more than one section, the bounding box with the greatest area was used for the effective diameter calculation. Figure 3 presents the distributions of effective diameters of the malignant and benign nodules in the database. The mean effective diameter of the 69 malignant nodules was 13.9±5.1 mm, and the mean effective diameter of the 401 benign nodules was 8.0±4.3 mm. It should be noted that the difference between the mean effective diameters of malignant and benign nodules was statistically significant ($P<0.01$).

III. METHODS

The automated *classification* of lung nodules as malignant or benign is initiated by automated lung nodule *detection* (Fig. 2), so that nodule classification takes as input the output of the automated detection method. This serial approach of detection followed by classification represents one potential implementation of the fully automated analysis of CT scans for lung cancer.

A. Automated lung nodule detection

Details of the automated lung nodule detection method have been previously reported.[12] To summarize here, automated nodule detection involves two-dimensional processing followed by three-dimensional analysis and feature extraction (Fig. 4). A single gray-level threshold is selected to isolate the patient from background in each section image. Analyses of gray-level histograms constructed from pixels within the patient on each section are used to segment the lung regions on a section-by-section basis. The composite of these segmented lung regions is used to represent the segmented lung volume, in which a three-dimensional multiple-gray-level-thresholding ("region melting") technique is used to identify initial lung nodule candidates based on a maximum volume requirement. Since a nodule is defined radiologically as any well-demarcated, soft-tissue focal opacity with a diameter less than 3 cm,[35] the automated method designates a structure identified through the region melting tech-

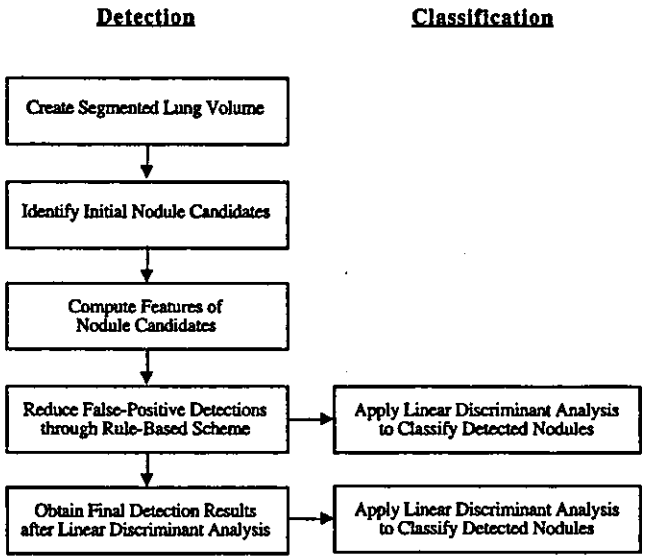

FIG. 4. Overall method for the automated detection and classification of lung nodules. The automated classification of nodule candidates as benign or malignant may occur at one of two stages in the detection method, either after the rule-based scheme or after linear discriminant analysis.

nique as a nodule candidate if its volume is less than the volume of a model 3-cm-diameter sphere. Figures 5 and 6 show original CT section images containing a biopsy-confirmed malignant nodule and a confirmed benign nodule, respectively, along with the initial lung nodule candidates identified in the corresponding section. Magnified views of the nodules and the candidates that correspond to the nodules are also shown. Note that although nodule candidates are identified in three-dimensions, only a single section is shown in the figures, so that the two regions in Fig. 5(d), for example, actually belong to the same three-dimensional nodule candidate.

The categorization of nodule candidates as "nodule" or "non-nodule" proceeds with rule-based and linear discriminant analyses using features computed for each nodule candidate. The rule-based classifier is used to eliminate "obvious" non-nodules based on maximum eccentricity, gray-level standard deviation, and the average area of a candidate in all sections in which it appears. Linear discriminant analysis[36] is applied to a nine-dimensional feature vector computed for the remaining candidates to distinguish further between nodule candidates that correspond to actual nodules and those that correspond to non-nodule structures. These features include six three-dimensional features: (1) the mean gray level of the candidate, (2) the gray-level standard deviation, (3) the gray-level threshold at which the candidate was identified, (4) volume, (5) sphericity, and (6) the radius of the sphere of equivalent volume. The first three features are gray-level based, while the second three features are morphology based. Three two-dimensional morphological features are used also: (1) maximum eccentricity, (2) maximum compactness, and (3) maximum circularity. Automated lung nodule detection results then may be obtained by specifying an appropriate linear discriminant analysis output threshold that yields an acceptable nodule-detection sensitivity and false-positive rate.

B. Automated lung nodule classification

The results of the automated nodule detection process described in the previous section serve as the basis for automated lung nodule classification. Features of the detected nodule candidates were merged through linear discriminant analysis for the purpose of nodule classification. These features included the nine features used for the detection task and three additional features: minimum compactness, effective diameter, and location along the z axis. Classification

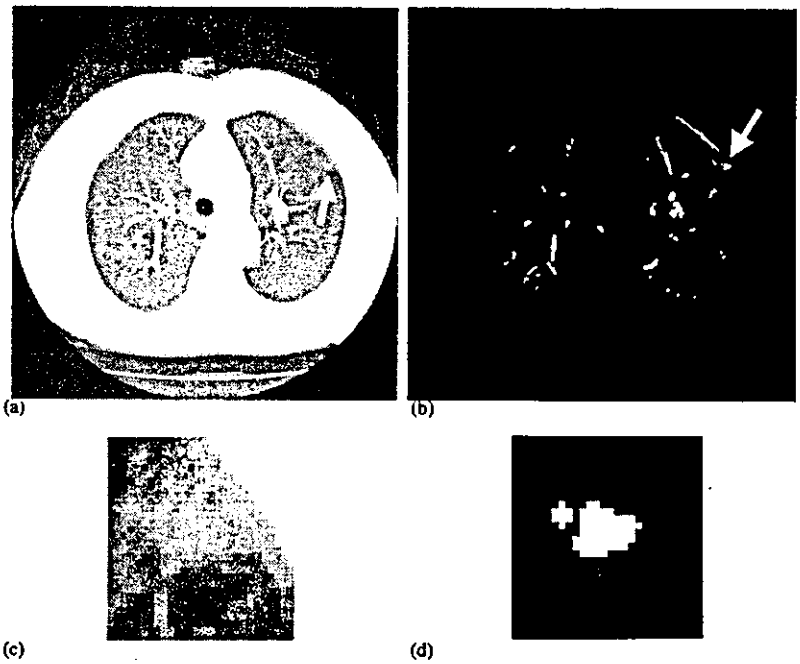

FIG. 5. (a) Original image containing a biopsy-confirmed malignant lung nodule (arrow). (b) Binary image depicting automatically determined initial lung nodule candidates, including one corresponding to the malignant lesion (arrow). (c) Magnified view of the nodule in (a). (d) Magnified view of the corresponding nodule candidate in (b). Since nodule candidates are identified in three-dimensions, the two regions in (d) actually belong to the same three-dimensional nodule candidate. Automated classification proceeds on the basis of these detection results.

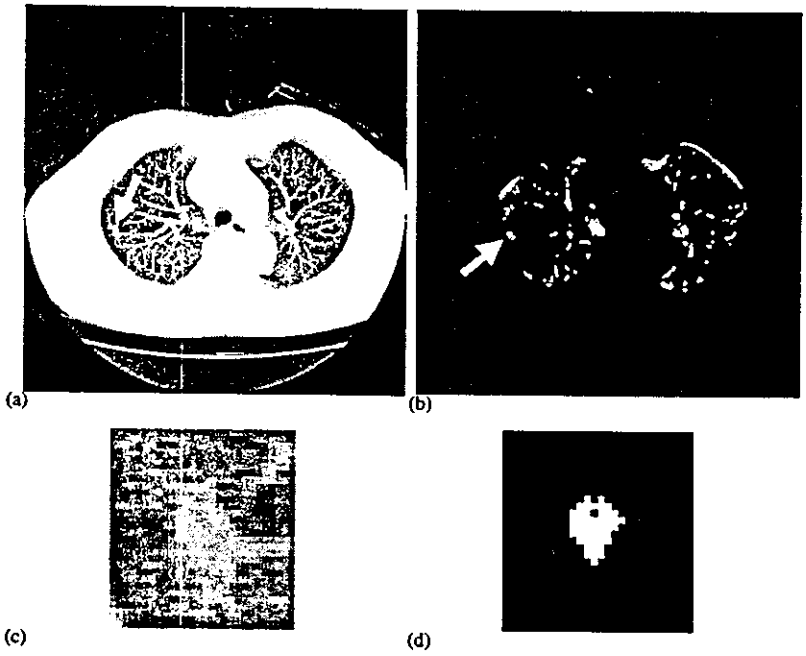

FIG. 6. (a) Original image containing a confirmed benign lung nodule (arrow). (b) Binary image depicting automatically determined initial lung nodule candidates, including one corresponding to the benign lesion (arrow). (c) Magnified view of the nodule in (a). (d) Magnified view of the corresponding nodule candidate in (b). Automated classification proceeds on the basis of these detection results.

was performed on the basis of the pixels contained within the nodule candidates themselves, not within regions of interest constructed around the candidates.

The set of nodule candidates generated by the detection method comprises three subsets, as illustrated in Fig. 7: nodule candidates that correspond to actual malignant nodules, candidates that correspond to actual benign nodules, and candidates that do not correspond to nodules (i.e. false positives). Our purpose in this initial study was to investigate the performance of subsequent linear discriminant analysis in the differentiation of automatically detected lung nodule candidates that correspond to malignant nodules from those that correspond to benign nodules. Accordingly, computer-detected candidates that corresponded to actual nodules (subsets A and B in Fig. 7) were manually separated from computer-detected candidates that represented false-positive detections (subset C) based on knowledge of the locations of "true" nodules so that computer-detected candidates that represented false-positive detections were ignored for this analysis. This separation was performed at two different stages in the automated detection method (Fig. 4): (1) after the rule-based scheme was applied and (2) after the linear discriminant analysis (i.e., the "final" detection results). A separate analysis was performed to evaluate the ability of the subsequent linear discriminant analysis to distinguish between nodule candidates that corresponded to malignant nodules and all other nodule candidates (i.e., those that corresponded to benign nodules *and* those that represented false positives), as depicted in Fig. 7. No manual separation of appropriate nodule candidate subsets was required for this evaluation. The essential aspects of this serial approach are that the nodules to be classified were initially identified through the automated detection method and that the features used for classification were derived from the candidates as isolated from the surrounding anatomy by the automated detection method.

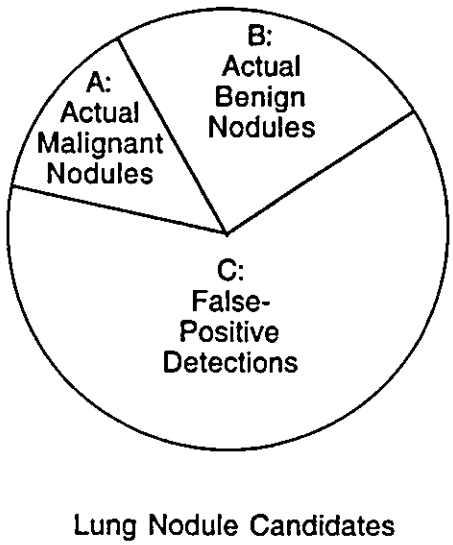

FIG. 7. A schematic of the set of lung nodule candidates that is obtained at either one of two stages in the automated nodule detection method (see Fig. 4): (1) after application of the rule-based scheme or (2) after application of the linear discriminant analysis used to obtain the final detection results. Automated classification of these candidates then is performed to differentiate between (1) members of subset A and members of subset B or (2) members of subset A and members of the union of subsets B and C.

IV. RESULTS

A. Detection

As the precursor to automated classification, the automated nodule detection method was evaluated first (Table I). The 393-case database contained 470 actual lung nodules, of which 14.7% ($n=69$) were malignant and 85.3% ($n=401$) were benign. The initial set of lung nodule candidates gen-

TABLE I. Composition of the nodule candidate sets that result at different stages of the automated nodule detection method. The results after linear discriminant analysis are based on a leave-one-out analysis. Since some actual nodules were represented by two nodule candidates, the numbers before the "\" indicate the number of nodule candidates in the category, while the numbers after the "\" represent the number of actual nodules to which those candidates correspond.

	Number of actual nodules (total 470)		
	Malignant (total 69)	Benign (total 401)	False-positive candidates
Initial set of nodule candidates	416 (88.5%)		212 236
	66 (95.7%)	350 (87.3%)	
Nodule candidates after rule-based scheme	410\405 (86.2%)		101 935
	67\66 (95.7%)	343\339 (84.5%)	
Nodule candidates after linear discriminant analysis	335\332 (70.6%)		15 266
	59\58 (84.1%)	276\274 (68.3%)	

erated by the detection method (prior to both the rule-based and linear discriminant analyses) captured 416 of the 470 actual lung nodules (88.5%). Sixty-six of these 416 nodules (15.9%) corresponded to actual malignant nodules and 350 (84.1%) corresponded to actual benign nodules, so that 95.7% (66/69) of the actual malignant nodules were included among these initial candidates and 87.3% (350/401) of the actual benign nodules were included. The initial nodule candidate set also included 212,236 structures that represented non-nodules (i.e., initial false positives).

Implementation of the rule-based scheme eliminated 52.0% of the false positives ($n = 110{,}301$) with the loss of 11 true-positive nodules, so that 405 of the 470 (86.2%) actual lung nodules corresponded to nodule candidates that passed the rule-based scheme. These 405 lung nodules, however, were represented by 410 separate nodule candidates, since five actual lung nodules each were associated with structures that corresponded to two distinct nodule candidates. A single nodule may be "divided" by the automated method into two distinct structures if the automated three-dimensional region-labeling algorithm fails to recognize the spatial relationship between a region associated with an actual nodule in one section image and a region associated with the same actual nodule in an adjacent section image. This situation occurred in only 1.2% of the actual nodules that remained nodule candidates after the rule-based scheme was implemented (5 out of 405).

Accordingly, the first set of lung nodule candidates evaluated by the automated classification scheme, the candidates that remained after the rule-based scheme (see Fig. 4), contained 410 candidates that corresponded to actual nodules, including 67 candidates that corresponded to 66 malignant nodules (95.7% of the total 69 malignant nodules contained in the database) and 343 that corresponded to 339 benign nodules (84.5% of the total 401 benign nodules contained in the database), and 101 935 candidates that represented false positives (Table I).

In the final stage of the automated detection method, the nine-dimensional feature vectors for the nodule candidates that passed the rules were merged by using linear discriminant analysis to further eliminate false-positive detections. A leave-one-out analysis was performed in which all nodule candidates except one were used to train the linear discriminator, which was then applied to the remaining nodule candidate. This process was repeated until all nodule candidates had been used as the "left out" candidate. The output of this linear discriminant analysis may be successively thresholded to generate an ROC curve that depicts the ability of linear discriminant analysis to differentiate nodule candidates that correspond to actual nodules from those that correspond to non-nodules, as shown in Figure 8. The area under the ROC curve (A_z) for this task was 0.90. At one specific operating point along this curve (see Fig. 8), the number of false positives was reduced to 15,266 at an overall nodule detection sensitivity of 70.6% (332/470). Accordingly, the second set of lung nodule candidates evaluated by the automated classification scheme was the final set of detection results at this specific performance level. This set contained 335 candidates that corresponded to actual nodules, including 59 candidates that corresponded to 58 malignant nodules (84.1% of the total 69 malignant nodules contained in the database) and 276 candidates that corresponded to 274 benign nodules (68.3% of the total 401 benign nodules contained in the database), and 15 266 candidates that represented false-positive detections (Table I).

B. Classification

The set of nodule candidates generated at either one of two stages in the automated detection method consisted of the three subsets shown in Fig. 7. ROC analysis[37] was used to evaluate the ability of the linear discriminant classifier to

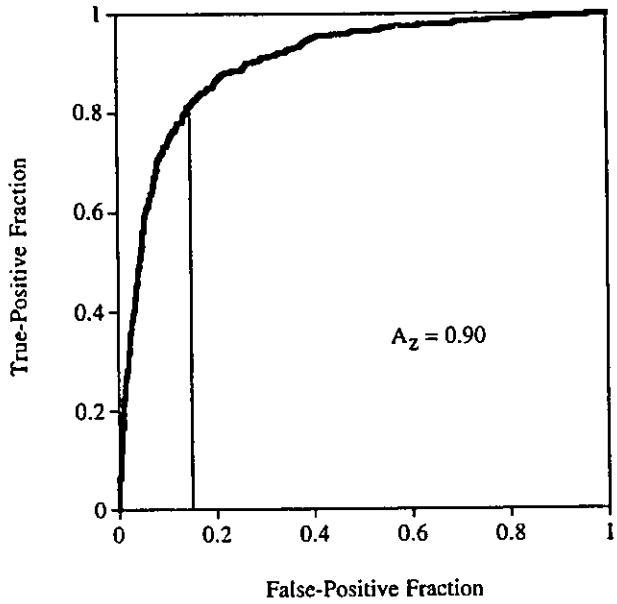

FIG. 8. ROC curve representing the ability of linear discriminant analysis employed during automated detection to differentiate between candidates that correspond to actual nodules and candidates that correspond to nonnodules based on the nine-dimensional feature vectors of nodule candidates that pass the rule-based scheme. These results are based on a leave-one-out analysis. The vertical line indicates the operating point at which the final detection results used for this study were obtained.

differentiate among these subsets. A_z values were computed based on output values from the linear discriminant analysis (see Table II).

Feature values for the 410 post-rule-based-scheme nodule candidates that corresponded to actual nodules (67 malignant, 343 benign) were analyzed. The same individual features and feature combinations investigated for automated detection[12] were evaluated for their ability to discriminate between these malignant and benign nodule candidates. The A_z values are reported in Table II. Also shown in Table II is the A_z value that results when several additional features are combined with selected original features; an A_z value of 0.79 was attained based on the features (1) radius of the sphere of equivalent volume, (2) minimum compactness, (3) maximum compactness, (4) gray-level threshold at which the candidate was identified, (5) effective diameter, and (6) location along the z axis identified through an exhaustive search of feature combinations in conjunction with a leave-one-out training/testing approach using all nodule candidates in the database. Based on the CLABROC program (Charles E. Metz, Ph.D., The University of Chicago), this A_z value was statistically significantly greater ($P<0.05$) than the highest A_z value achieved with a single feature (i.e., the radius of the sphere of an equivalent volume feature).

While the set of nodule candidates that emerges from the rule-based stage of the detection method represents the highest yield of actual nodules, any practical application of the detection method would incorporate linear discriminant analysis to further reduce the number of false positives. The nodule candidates that remained at one particular operating point of the linear discriminant analysis employed by the detection method (82% sensitivity, 85% specificity, as shown in Fig. 8) were obtained. Feature values for the 335 such candidates that corresponded to actual nodules (59 malignant, 276 benign) were analyzed. The same individual features and feature combinations investigated for automated detection were evaluated for their ability to discriminate be-

TABLE II. Areas under the ROC curve (A_z) representing the ability of individual nodule candidate features and various combinations of candidate features to differentiate nodule candidates that correspond to malignant nodules from those that correspond to benign nodules. These results are based on a leave-one-out analysis.

	A_z		
	After rule-based scheme	After rules and LDA	
Feature(s)	Malignant versus benign	Malignant versus benign	Malignant versus other
Radius of the sphere of equivalent volume	0.73	0.74	0.68
Volume	0.71	0.72	0.67
Mean gray level	0.67	0.66	0.53
Gray-level standard deviation	0.65	0.62	0.59
Sphericity	0.61	0.61	0.62
Threshold at which candidate identified	0.61	0.63	0.50
Maximum compactness	0.58	0.61	0.55
Maximum eccentricity	0.52	0.55	0.60
Maximum circularity	0.51	0.57	0.56
All features (9)	0.75	0.74	0.82
3-D features (6)	0.75	0.74	0.75
Morphology-based features (6)	0.74	0.75	0.79
3-D morphology-based features (3)	0.74	0.75	0.73
Gray-level-based features (3)	0.66	0.64	0.57
2-D features (3)	0.61	0.63	0.63
Additional features combined with selected original features (see the text)	0.79	0.79	0.79

tween these malignant and benign nodule candidates. The A_z values are reported in Table II along with the A_z value that results when several additional features are combined with selected original features; an A_z value of 0.79 was attained based on the features (1) the radius of the sphere of equivalent volume, (2) minimum compactness, (3) maximum compactness, (4) gray-level threshold at which the candidate was identified, (5) effective diameter, and (6) location along the z-axis identified through an exhaustive search of feature combinations in conjunction with a leave-one-out training/testing approach using all nodule candidates in the database that remained at the particular operating point of the linear discriminant analysis. Based on the CLABROC program, this A_z value was statistically significantly greater ($P < 0.05$) than the highest A_z value achieved with a single feature (i.e., the radius of the sphere of equivalent volume feature).

The previous analyses required the manual separation of candidates that correspond to actual nodules from the set of lung nodule candidates (i.e., with reference to Fig. 7, nodule candidates that belonged to subsets A and B were segregated from nodule candidates that belonged to subset C). To circumvent this intervention, we investigated the ability of the method to differentiate candidates that correspond to actual malignant nodules from all other nodule candidates (i.e., candidates that correspond to actual benign nodules and candidates that correspond to false positives) that remained at the 82% sensitivity/85% specificity operating point of the detection method (see Fig. 8). The original nine features used for automated detection were evaluated for their ability to discriminate between the resulting 59 candidates that corresponded to malignant nodules and the 15,542 other nodule candidates (276 candidates that corresponded to benign nodules and 15 266 candidates that represented non-nodules). An A_z value of 0.82 was attained for this classification task (see Table II).

V. DISCUSSION

In clinical practice, nodule classification would not be performed, either by humans or computers, based on low-dose CT scans acquired with 10-mm collimation and 10-mm reconstruction interval, as was the case with the database used in this study. A suspected abnormal finding on such low-dose CT scans would typically be more fully assessed from a targeted, high-resolution diagnostic CT scan at the appropriate anatomic level. Differential diagnosis would depend then on this follow-up scan. The serial approach to nodule classification would be more relevant clinically to lung cancer screening examinations performed with multislice CT scanners at high axial resolution (less than, for example, 2.5 mm); such examinations are currently used by radiologists at some centers for both the detection and differential diagnosis of nodular lesions. Another relevant clinical application for this serial approach is the diagnostic CT scan, either obtained as a follow-up after a thicker-section low-dose screening study or as an independent study not related to lung cancer in which a nodule may be present as an incidental finding. Although the thick-section CT scans used do not represent the preferred clinical application of the serial approach to lung nodule classification, the present study nevertheless demonstrates the validity and potential utility of the method.

This initial study focused predominantly on features that were used for the detection task. Other features more appropriate specifically for the classification task will be investigated in future work. The ability to more accurately classify detected nodule candidates will depend greatly on the axial voxel dimension of the acquired CT data; as thinner-section CT scans are analyzed, the characteristics of both benign and malignant nodules will be more faithfully captured, a broader base of nodule candidate features will become available, and the discriminating ability of automated classification methods is expected to improve. We have already begun to show in other work that the *detection* task is facilitated by CT scans with higher axial resolution.

It is not surprising that the volume-based features demonstrated the greatest potential for accurate classification of nodules in this database. As indicated earlier, the actual malignant nodules and the actual benign nodules in the database exhibited a statistically significant difference in mean effective diameter. This "truth" as captured by a radiologist, however, will not necessarily translate into the nodules as extracted by the computer. The essential point of this study is that the lung nodule candidates as identified by the automated detection method contained information that allowed for their subsequent classification as malignant or benign. The importance of other features may emerge when the method is applied to CT scans acquired with higher axial resolution and when specific methods to more accurately segment lung nodule candidates are developed and validated.

Clearly, the separation of nodule candidates that correspond to actual nodules is a step that would be eliminated in a fully automated serial approach to nodule classification. Our experiment to distinguish nodule candidates that correspond to malignant nodules from all other nodule candidates (benign nodules and false positives) was an effort to attain such complete automation. The fully automated process of lung cancer analysis will depend, in part, on the ability of the detection method initially to reduce further the number of false-positive detections. The level of performance indicated by the A_z value of 0.82 determined in the "malignant vs other" classification mode would not be deemed clinically acceptable, since a relatively large number of false-positive detections would be erroneously classified "malignant." The "malignant vs other" mode yields a higher A_z value (0.82) than the "malignant vs benign" mode (0.79), a result that is likely due to the larger number of structures more easily classified as "nonmalignant" when the detection false positives are included in the set for partitioning, although it should be noted that the sets of features that result in the highest A_z value for each classification mode differ (see Table II).

VI. CONCLUSIONS

We have developed a serial approach to the analysis of lung nodules in CT scans that merges the automated detection task and the automated classification task. Lung nodule candidates identified through our automated detection method were used as input to a linear discriminant analysis that classified the automated detection results as malignant or benign. The overall serial classification method attained an A_z value of 0.79 in distinguishing between nodule candidates that corresponded to actual malignant nodules and nodule candidates that corresponded to actual benign nodules. This approach represents an important step toward the fully automated analysis of lung nodules in CT scans.

ACKNOWLEDGMENTS

The authors would like to thank Maryellen L. Giger, Ph.D. and Yulei Jiang, Ph.D. for their insightful comments on the nature of this study. This work was supported in part by USPHS Grant No. CA83908 and funding from The University of Chicago Cancer Research Center. S.G. Armato and K. Doi are shareholders in R2 Technology, Inc. (Sunnyvale, CA). K. Doi is a shareholder in Deus Technology, Inc. (Rockville, MD).

[a]Corresponding author. Samuel G. Armato III, Ph.D., Department of Radiology, MC 2026, The University of Chicago, 5841 S. Maryland Ave., Chicago, Illinois 60637. Telephone: 773-834-3044; fax: 773-702-0371; electronic mail: s-armato@uchicago.edu

[1]M. Kaneko, K. Eguchi, H. Ohmatsu, R. Kakinuma, T. Naruke, K. Suemasu, and N. Moriyama, "Peripheral lung cancer: Screening and detection with low-dose spiral CT versus radiography," Radiology **201**, 798–802 (1996).

[2]S. Sone *et al.*, "Mass screening for lung cancer with mobile spiral computed tomography scanner," Lancet **351**, 1242–1245 (1998).

[3]C. I. Henschke *et al.*, "Early lung cancer action project: Overall design and findings from baseline screening," Lancet **354**, 99–105 (1999).

[4]M. L. Giger, K. T. Bae, and H. MacMahon, "Computerized detection of pulmonary nodules in computed tomography images," Invest. Radiol. **29**, 459–465 (1994).

[5]K. Kanazawa, M. Kubo, N. Niki, H. Satoh, H. Ohmatsu, K. Eguchi, and N. Moriyama, "Computer assisted lung cancer diagnosis based on helical images," in *Image Analysis Applications and Computer Graphics: Proceedings of the 3rd International Computer Science Conference*, edited by R. T. Chin, H. H. S. Ip, A. C. Naiman, and T.-C. Pong (Springer-Verlag, Berlin, 1995), pp. 323–330.

[6]W. J. Ryan, J. E. Reed, S. J. Swensen, and P. F. Sheedy, Jr., "Automatic detection of pulmonary nodules in CT," in *Proceedings Computer Assisted Radiology*, edited by H. U. Lemke, M. W. Vannier, K. Inamura, and A. G. Farman (Elsevier Science, Amsterdam, 1996), pp. 385–389.

[7]T. Okumura, T. Miwa, J. Kako, S. Yamamoto, M. Matsumoto, Y. Tateno, T. Iinuma, and T. Matsumoto, "Image processing for computer-aided diagnosis of lung cancer screening system by CT (LSCT)," Proc. SPIE **3338**, 1314–1322 (1998).

[8]M. Fiebich, C. Wietholt, B. C. Renger, S. G. Armato III, K. R. Hoffmann, D. Wormanns, and S. Diederich, "Automatic detection of pulmonary nodules in low-dose screening thoracic CT examinations," Proc. SPIE **3661**, 1434–1439 (1999).

[9]S.-L. Lou, C.-L. Chang, K.-P. Lin, and T.-S. Chen, "Object-based deformation technique for 3-D CT lung nodule detection," Proc. SPIE **3661**, 1544–1552 (1999).

[10]H. Satoh, Y. Ukai, N. Niki, K. Eguchi, K. Mori, H. Ohmatsu, R. Kakinuma, M. Kaneko, and N. Moriyama, "Computer aided diagnosis system for lung cancer based on retrospective helical CT image," Proc. SPIE **3661**, 1324–1335 (1999).

[11]H. Taguchi, Y. Kawata, N. Niki, H. Satoh, H. Ohmatsu, R. Kakinuma, K. Eguchi, M. Kaneko, and N. Moriyama, "Lung cancer detection based on helical CT images using curved surface morphology analysis," Proc. SPIE **3661**, 1307–1314 (1999).

[12]S. G. Armato III, M. L. Giger, and H. MacMahon, "Automated detection of lung nodules in CT scans: Preliminary results," Med. Phys. **28**, 1552–1561 (2001).

[13]M. S. Brown, M. F. McNitt-Gray, J. G. Goldin, R. D. Suh, J. W. Sayre, and D. R. Aberle, "Patient-specific models for lung nodule detection and surveillance in CT images," IEEE Trans. Med. Imaging **20**, 1242–1250 (2001).

[14]L. Fan, C. L. Novak, J. Qian, G. Kohl, and D. P. Naidich, "Automatic detection of lung nodules from multi-slice low-dose CT images," Proc. SPIE **4322**, 1828–1835 (2001).

[15]M. Fiebich, D. Wormanns, and W. Heindel, "Improvement of method for computer-assisted detection of pulmonary nodules in CT of the chest," Proc. SPIE **4322**, 702–709 (2001).

[16]M. N. Gurcan, B. Sahiner, N. Petrick, H.-P. Chan, E. A. Kazerooni, P. N. Cascade, and L. M. Hadjiiski, "Lung nodule detection on thoracic computed tomography images: Preliminary evaluation of a computer-aided diagnosis system," Med. Phys. **29**, 2552–2558 (2002).

[17]J. P. Ko and M. Betke, "Chest CT: Automated nodule detection and assessment of change over time—preliminary experience," Radiology **218**, 267–273 (2001).

[18]Y. Lee, T. Hara, H. Fujita, S. Itoh, and T. Ishigaki, "Automated detection of pulmonary nodules in helical CT images based on an improved template-matching technique," IEEE Trans. Med. Imaging **20**, 595–604 (2001).

[19]S. G. Armato III, M. L. Giger, J. T. Blackburn, K. Doi, and H. MacMahon, "Three-dimensional approach to lung nodule detection in helical CT," Proc. SPIE **3661**, 553–559 (1999).

[20]S. G. Armato III, M. L. Giger, C. J. Moran, J. T. Blackburn, K. Doi, and H. MacMahon, "Computerized detection of pulmonary nodules on CT scans," Radiographics **19**, 1303–1311 (1999).

[21]S. G. Armato III, M. B. Altman, and P. J. LaRivière, "Automated detection of lung nodules in CT: Effect of image reconstruction algorithm," Med. Phys. **30**, 461–472 (2003).

[22]S. G. Armato III, F. Li, M. L. Giger, H. MacMahon, S. Sone, and K. Doi, "Lung cancer: Performance of automated lung nodule detection applied to cancers missed in a CT screening program," Radiology **225**, 685–692 (2002).

[23]D. Cavouras, P. Prassopoulos, and N. Pantelidis, "Image analysis methods for solitary pulmonary nodule characterization by computed tomography," Eur. J. Radiol. **14**, 169–172 (1992).

[24]C. I. Henschke, D. F. Yankelevitz, I. Mateescu, D. W. Brettle, T. G. Rainey, and F. S. Weingard, "Neural networks for the analysis of small pulmonary nodules," Clin. Imaging **21**, 390–399 (1997).

[25]Y. Kawata, N. Niki, H. Ohmatsu, R. Kakinuma, K. Eguchi, M. Kaneko, and N. Moriyama, "Quantitative surface characterization of pulmonary nodules based on thin-section CT images," IEEE Trans. Nucl. Sci. **45**, 2132–2138 (1998).

[26]T. Tozaki, Y. Kawata, N. Niki, H. Ohmatsu, R. Kakinuma, K. Eguchi, M. Kaneko, and N. Moriyama, "Pulmonary organs analysis for differential diagnosis based on thoracic thin-section CT images," IEEE Trans. Nucl. Sci. **45**, 3075–3082 (1998).

[27]M. F. McNitt-Gray, E. M. Hart, N. Wyckoff, J. W. Sayre, J. G. Goldin, and D. R. Aberle, "A pattern classification approach to characterizing solitary pulmonary nodules imaged on high resolution CT: Preliminary results," Med. Phys. **26**, 880–888 (1999).

[28]M. F. McNitt-Gray, N. Wyckoff, J. W. Sayre, J. G. Goldin, and D. R. Aberle, "The effects of co-occurrence matrix based texture parameters on the classification of solitary pulmonary nodules imaged on computed tomography," Computerized Med. Imaging Graphics **23**, 339–348 (1999).

[29]A. P. Reeves, W. J. Kostis, D. F. Yankelevitz, and C. I. Henschke, "Three-dimensional shape characterization of solitary pulmonary nodules from helical CT scans," in *Computer Assisted Radiology and Surgery (CARS'99)*, edited by H. U. Lemke, M. W. Vannier, K. Inamura, and A. G. Farman (Elsevier Science, Amsterdam, 1999), pp. 83–87.

[30]Y. Kawata *et al.*, "Quantitative analysis of internal texture for classification of pulmonary nodules in three-dimensional thoracic images," Proc. SPIE **3979**, 863–871 (2000).

[31] Y. Kawata *et al.*, "Computer-aided differential diagnosis of pulmonary nodules based on a hybrid classification approach," Proc. SPIE **4322**, 1796–1806 (2001).
[32] K. Minami *et al.*, "Computerized characterization of contrast enhancement patterns for classifying pulmonary nodules," Proc. SPIE **4322**, 1936–1943 (2001).
[33] Y. Jiang, R. M. Nishikawa, and J. Papaioannou, "Dependence of computer classification of clustered microcalcifications on the correct detection of microcalcifications," Med. Phys. **28**, 1949–1957 (2001).
[34] S. Sone *et al.*, "Results of three-year mass screening programme for lung cancer using mobile low-dose spiral computed tomography scanner," Br. J. Cancer **84**, 25–32 (2001).
[35] W. J. Tuddenham, "Glossary of terms for thoracic radiology: Recommendations of the Nomenclature Committee of the Fleischner Society," Am. J. Rhinol. **143**, 509–517 (1984).
[36] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis* (Prentice–Hall, Englewood Cliffs, NJ, 1992).
[37] C. E. Metz, "ROC methodology in radiologic imaging," Invest. Radiol. **21**, 720–733 (1986).