

# Potential of Computer-Aided Diagnosis to Improve CT Lung Cancer Screening

Noah Lee, *Student Member, IEEE*, Andrew F. Laine, *Senior Member, IEEE*, Guillermo Márquez, Jeffrey M. Levsky, and John K. Gohagan

**Abstract**—The development of low-dose spiral computed tomography (CT) has rekindled hope that effective lung cancer screening might yet be found. Screening is justified when there is evidence that it will extend lives at reasonable cost and acceptable levels of risk. A screening test should detect all extant cancers while avoiding unnecessary workups. Thus optimal screening modalities have both high sensitivity and specificity. Due to the present state of technology, radiologists must opt to increase sensitivity and rely on follow-up diagnostic procedures to rule out the incurred false positives. There is evidence in published reports that computer-aided diagnosis technology may help radiologists alter the benefit–cost calculus of CT sensitivity and specificity in lung cancer screening protocols. This review will provide insight into the current discussion of the effectiveness of lung cancer screening and assesses the potential of state-of-the-art computer-aided design developments.

**Index Terms**—Computer-aided diagnosis, lung cancer screening, machine learning, receiver operating characteristics.

## I. INTRODUCTION

CANCER affects everyone—the young and old, the rich and poor, men, women, and children. Lung cancer is the leading cancer killer in the United States, with 1.3 million deaths worldwide annually [76]. Estimates for 2008 were for 215 020 new lung cancer cases and 161 840 deaths from lung cancer in the United States [78].<sup>1</sup> In 2009, the estimated new cases and deaths are 219 440 and 159 390, respectively.<sup>2</sup> More than 75% of lung cancers are diagnosed in advanced stages. The average five-year survival rate after lung cancer diagnosis is about 15%. If lung cancer is detected at its earliest stage, the five-year survival rate can reach 70% [14], [16], [19]. Approximately \$9.6 billion are spent in the United States each year for lung cancer treatment.<sup>3</sup> These figures call for effective cancer control and prevention strategies such as lung cancer screening programs.

Manuscript received February 18, 2009. First published October 16, 2009; current version published December 09, 2009.

N. Lee and A. F. Laine are with the Heffner Biomedical Imaging Lab, Department of Biomedical Engineering, Columbia University, New York, NY 10027 USA (e-mail: nl2168@columbia.edu; laine@columbia.edu).

G. Márquez is with the Early Detection Research Group, National Cancer Institute, Bethesda, MD 20892 USA (e-mail: marquezg@mail.nih.gov).

J. M. Levsky is with the Division of Cardiothoracic Imaging, Department of Radiology, Montefiore Medical Center and Albert Einstein College of Medicine, Bronx, NY 10467 USA (e-mail: jlevsky@montefiore.org).

J. K. Gohagan is with the Basic Prevention Sciences Research Group, National Cancer Institute, Bethesda, MD 20892 USA (e-mail: gohaganj@mail.nih.gov).

Digital Object Identifier 10.1109/RBME.2009.2034022

<sup>1</sup><http://www.cancer.gov/cancertopics/types/lung> (2008).

<sup>2</sup><http://www.cancer.gov/cancertopics/types/lung> (2009).

<sup>3</sup><http://progressreport.cancer.gov>.

Lung cancer screening is the presumptive identification of unrecognized malignant tissue in high-risk asymptomatic individuals. Screening may include medical examinations such as sputum cytology (SC) [15], [79], chest X-ray (CXR) [15], low-dose spiral computed tomography (LDCT) [8], [15], [31], [46], [68], gene expression tests [88], and accompanied computer-aided diagnosis (CADx) and detection (CADE) schemes [81]–[87]. Current CAD technology shows potential to improve CT lung cancer diagnosis, yet the question of whether state-of-the-art screening technology can decrease mortality rate is inconclusive [1], [3]–[5], [9]–[15], [20], [64], [67], [69]–[75], [77], [80], [86]. To address the mortality question by direct comparison of LDCT technology with CXR, several clinical trials are active or planned [1]. The largest of these and most advanced is the National Lung Screening Trial [6] (NLST) under the direction of U.S. National Cancer Institute in which the targeted 50 000 asymptomatic former or current heavy smokers were randomized to receive an initial and three annual screens by LDCT or CXR. Conclusive results are expected to be available by 2011. Outside the United States, randomized controlled trials include the ITALUNG trial,<sup>4</sup> NELSON,<sup>5</sup> and the UK Lung Cancer Screening Trial (UKLS).<sup>6</sup> Other clinical trials [90]<sup>7</sup> lack current gold standard clinical research guidelines to perform randomized controlled studies for objective evaluation of the superiority of a medical intervention for lung cancer screening.

In this paper, we take a top-down approach in discussing the potential of CAD for lung cancer screening. We hypothesize that state-of-the-art CAD technology has the potential to improve LDCT lung cancer screening, but the technology needs comparative and careful assessment with respect to clinically relevant performance measures [7], [103]–[105], [118]. We will provide insight into the current effectiveness of LDCT lung cancer screening and assess state-of-the-art CAD developments in commercial and academic research. In this context, this paper branches into three main components: 1) an overview of CAD and associated image-processing methodologies to aid the diagnostic decision process in lung cancer diagnosis; 2) the various evaluation criteria, in particular, high sensitivity and high specificity, in order to assess the potential of CAD for lung cancer screening; and 3) the integration of CAD into clinical practice.

<sup>4</sup><http://www.cspo.it>.

<sup>5</sup><http://www.nelsonproject.nl/>.

<sup>6</sup><http://www.hta.ac.uk/1752>.

<sup>7</sup><http://clinicaltrials.gov/ct2/show/NCT00963651>.

## II. HISTORICAL DEVELOPMENT OF CAD

The first concept of CAD was introduced half a century ago [17], [18], where Lusted talked about automated diagnosis of radiographs by computers in 1955 [77]. Early attempts to build a CAD system were initiated in the early 1950s. Several decades of research passed until this dream bore fruit in 1998 for the first commercial mammography CAD system [77] approved by the FDA. Large-scale systematic research began in 1980, but new automated systems were not immediately successful [89]. Development began from an initial concept of a fully automated computer diagnosis to a computer-aided diagnosis, where the human relies on the machine as a second reader. Since then, many CAD applications have been developed to help radiologists interpret images [77], [89]. CAD remains a major research subject, and many CAD systems and applications have been proposed. Currently, the major application of CAD involves breast cancer [71], lung cancer [90]–[98], colon cancer, and prostate cancer treatment. They have become part of the routine clinical work for detection of breast cancer in some clinics [89]. Starting from ad hoc and heuristic approaches [24], [25], [29], CAD technology moved to sophisticated machine-learning and data-mining techniques [82], [93], [98], [106]–[109]. In recent years, sophisticated machine-learning schemes have been developed [21]–[23], [90], [110] and entered the field of automated, semiautomated, and interactive CAD systems [82], [90]. Machine learning for CAD has become one of the principal research areas in medical imaging and diagnostic radiology. The reported literature [2], [24], [25], [27], [29], [91], [92], [98], [99] gives evidence that current CAD schemes as a second reader opinion often outperform manual grading performance of experts alone. Nishikawa and Doi provide an in-depth review of the historical and current developments of CAD from a clinical perspective [77], [89]. An in-depth review of CAD methodologies for lung cancer is described by Sluimer *et al.* [54] and Chan *et al.* [86].

## III. CAD OVERVIEW FOR CT LUNG CANCER SCREENING

CAD systems for lung tissue discrimination, nodule discrimination, and nodule characterization are increasingly being used as a *second reader* to aid the diagnostic decision process and to reduce the number of overlooked lung cancers. There is evidence in published reports that CAD technology may help radiologists alter the benefit–cost calculus of CT sensitivity and specificity in lung cancer screening protocols to the benefit of patients and radiologists alike [85]. Current CAD schemes include *lung tissue discrimination* [21]–[29], *nodule detection and classification* [34], [36], [39], [43]–[46], [82], [87], [97], *interstitial disease detection*, *differential diagnosis of interstitial disease*, *distinction between benign and malignant pulmonary nodules* [93], [94], and *estimation of malignancy potential* as well as *growth measurement* [55]. CAD in this context has improved since 2000, but major challenges persist in three areas: 1) *discrimination of lung tissue and regions of abnormality*; 2) *nodule detection and classification*; and 3) *nodule characterization and growth measurement*.

The pool of existing CAD systems and approaches is broad, ranging from hybrid image-processing systems [37] including registration [114], [115] and segmentation [116], [117] to lung

image retrieval and search. The field is advancing quickly, with new CAD schemes being developed and investigated for the task of lung cancer diagnosis and detection [106]–[109]. In what follows, we will give a brief snapshot of current CAD schemes for the three areas as well as present day and future areas of investigation that are being made.

### A. Lung Tissue and Regions of Abnormality Discrimination

The task of discriminating lung tissue and abnormal lung regions involves the analysis of large thoracic three-dimensional CT image datasets (see Fig. 1). Images containing diffuse abnormalities have been especially problematic in nodule screening due to partial volume effects and ambiguous image artifacts, making the distinction between nodule tissue and abnormal lung tissue difficult. Early approaches defined lung boundaries and used thresholding methods to differentiate between vessels and nodules [24]–[26]. Subsequent approaches attempted to remove ambiguous structures by comparisons between neighboring slices [27], while others applied true three-dimensional algorithms [28]. The notion of subtracting known anatomical structures to simplify the detection and classification task was suggested by the work of Mori *et al.* [29]. The authors described a method for the automated anatomical labeling of the tracheobronchial tree extracted from three-dimensional CT data and its application to virtual bronchoscopy. Proposed work in this area is manifold and provided discrimination results can be grouped to binary two-class discrimination [98] and finer multiclass discrimination [91] into respective tissue types.

Depeursigne *et al.* [91] presented a texture classification system for lung tissue multiclass classification into five different lung tissue patterns, (i.e., healthy, emphysema, ground glass, fibrosis, and micronodules). They used overcomplete wavelet frames combined with gray-level histogram features and obtained a classification accuracy of 92.5%. Classification was performed using k-nearest neighbor (KNN). In 2008, the authors reported a system [92] that integrated additional clinical context information to perform lung tissue classification with further 8% performance improvement compared to [91] using an optimized support vector machine (SVM).

Arzhaeva *et al.* [98] proposed a system for the localization of interstitial lesions in chest radiographs. The system used a two-class supervised classification approach to distinguish between normal and diseased texture. Texture analysis was performed by multiscale Gaussian filter banks, linear discriminant analysis (LDA), and an SVM classifier. They evaluated the method on 44 abnormal and eight normal cases with an area under the ROC curve (AUC) value of 78%.

Kato *et al.* [111] presented a bag of features approach for lung tissue multiclass classification in diffuse lung disease to classify disease patterns with inhomogeneous texture distributions within a region of interest (ROI). They use a scale-invariant feature transformation descriptor over many ROI samples for local feature extraction and to account for translation and rotation invariance. The authors report a classification accuracy of 92.8% using 1109 ROIs from 211 patients.

### B. Lung Nodule Discrimination

Lung *nodule discrimination* consists of two main components: a) *nodule detection* and b) *nodule classification*.



Fig. 1. From CT lung data to lung cancer diagnosis. (Top) CT lung dataset from the LIDC database with several hundred slices. (Middle) True positive nodules with different characteristics (solid, spiculated, and low contrast) surrounded in red. (Bottom) False positive nodules surrounded in yellow.

Recently, a comparative CAD assessment<sup>8</sup> was performed on the NLST data through standardized databases such as the Lung Image Database Consortium (LIDC) for the first time. The results of this assessment have not been published yet. The European counterpart for comparative CAD assessment is the NELSON and ANODE09 study.<sup>9</sup> Other openly available datasets include the database of Lung Test images from Motol Environment (Lung TIME) [113]. The Lung TIME database consists of 157 CT scans with 394 annotated nodules of various types, including solitary, regular, irregular, pleural, and vessel attached nodules.

In the last decades, a large body of research has been reported in the field of lung nodule detection and classification [2], [31]–[39], [43]–[46], [82], [87], [97], [112]. A central concern in nodule detection is the high rate of false positives when sensitivity is increased to detect subtle nodules. A nodule is deemed a false positive result if it led to a completely negative workup or more than 12 months of follow-up with no cancer diagnosis. Reducing false positive rates while maintaining high sensitivity is still a difficult problem (see Fig. 1). Techniques include LDA [34], rule-based approaches (a set of “if-then” statements) [38], combinations of these two [39], artificial neural networks (ANNs) [40], and maximum-margin based discriminators such as the SVM [92]. Novel methodologies for searching have been introduced, which include template matching for detection [36], unsupervised clustering techniques [39], and a local density maximum algorithm [41]. Methods to improve discrimination of nodules from lung tissue include subtraction of vessels by region-growing [30], knowledge-constrained routines based on anatomical models of the thorax

[31], and deformable models [32]. Various approaches have been taken to define the pleural interface and distinguish juxta-pleural nodules, including morphological image filtering [33], [34], curvature analysis of the pleural interface [35], and adaptive template-matching for the appropriate shape of a nodule given a location on the pleural wall [36].

Lee *et al.* [36] proposed a novel template-matching technique based on genetic algorithms and template matching for the detection of nodules. They evaluated their method on 557 sectional images with a detection rate of 72% and a false positive rate of 1.1 per sectional image.

Armato *et al.* [38] reported an extension of his earlier two- and three-dimensional automated lung CT analysis method [34] to segment lung volume on a section-by-section basis. A rule-based [42] approach combined with LDA was applied to reduce the number of nodule candidates. They evaluated their method on 43 CT scans with an AUC value of 90% and a nodule detection sensitivity of 70%. The false-positive detections per section were 1.5.

Li *et al.* [45] reported on a CAD scheme to help radiologists improve the detection of pulmonary nodules in chest radiographs by focusing on false positive reduction. They could reduce the number of false positives to 44.3% with a small increase in the number of true positives of 2.3%.

Katsuragawa *et al.* [2] described an automated method to distinguish benign and malignant solitary nodules. Fifty-five chest radiographs were discriminated using LDA and ANN for feature combination and classification. Comparisons with manual grading showed that LDA had an AUC value of 88.6%, whereas manual identification resulted in an AUC value of 85.4%.

Brown *et al.* [46] reported patient-specific models for detecting lung nodules for use in screening and follow-up surveillance. Baseline image data facilitated segmentation of subse-

<sup>8</sup><http://skynet.ohsu.edu/lungnodule09/>.

<sup>9</sup><http://anode09.isi.uu.nl/index.php>.

quent images so that changes in size and/or shape of nodules could be measured automatically. The system performed with an 86% detection rate and an average of 11 false positives per case on the baseline scans of 17 subjects. Follow-up scans performed with a detection rate of 81%. Brown *et al.* [62] also developed an automated system for detecting lung micronodules and applied it to data from 15 subjects with 77 lung nodules. Preliminary results indicated that the automated system considerably improved the radiologist's performance in micronodule detection but with a compensatory loss of specificity.

Gurcan *et al.* [39] developed a CAD system for lung nodule detection on CT images wherein the first-stage lung regions were identified by k-means clustering. After rule-based classification, LDA was used to further reduce the number of false positives. They used 1454 CT slices from 34 patients with 63 lung nodules and obtained a sensitivity of 84% with 5.48 false positives per slice.

Arimura *et al.* [43] reported a CAD system for nodule detection using a difference-image technique. They compared several rule-based schemes for identifying nodules. A massive-training ANN (MTANN) [44] reduced the false positives. The method was evaluated on a confirmed cancer database of 106 CT scans with 109 cancer lesions from 73 patients. They reported a sensitivity of 83% and 5.8 false positives per scan.

Suzuki *et al.* [82] developed a technique that used a multiple MTANN (multi-MTANN) for false-positive reduction. The investigators found that use of the trained multi-MTANN eliminated 68.3% of false-positive findings with a reduction of one true-positive result. The false-positive rate of the original CAD scheme was improved from 4.5 to 1.4 false positives per image, at an overall sensitivity of 81.3%, suggesting that this technique reduced the false-positive rate of the CAD scheme for nodule detection on chest radiographs while maintaining a high level of sensitivity.

Shiraishi *et al.* [87] investigated the effect of a CAD scheme on radiologist performance in the detection of lung cancers on chest radiographs. They combined two independent CAD schemes for the detection and classification of nodules into one new CAD scheme by use of a database of 150 chest images. Performance of the CAD scheme indicated that sensitivity in detecting lung nodules was 80.6%, with 1.2 false-positive results per image, and sensitivity and specificity for classification of nodules by use of the same database for training and testing the CAD scheme were 87.7% and 66.7%, respectively. The AUC value for detection of lung cancers improved significantly ( $P = .008$ ) from without (72.4%) to with CAD (77.8%). Shiraishi *et al.* [100] also developed a CAD system for detection of nodules in the lateral views of chest radiographs in order to improve the overall performance in combination with a CAD scheme for posterior–anterior (PA) views.

Murphy *et al.* [112] presented a large-scale evaluation study of automatic nodule detection in chest CT using local image features (shape index and curvedness) and two successive iterations of KNN classification for false-positive reduction. On 813 randomly selected scans, a sensitivity of 80% was achieved with an average of 4.2 false positives/scan. The same group participated in the ANODE09 benchmark and achieved top performance among six different CAD systems. Most of the work reported

for nodule detection and classification uses binary two-class decisions to discriminate nodules.

### C. Nodule Characterization by Malignancy Potential

The characterization of nodules by their malignancy potential involves the analysis of nodule candidates into different nodule type categories such as subtlety, texture, margin, sphericity, calcification, internal structure, lobulation, spiculation, and malignancy. A major application of CAD for lung CT is the classification of nodules by likelihood of malignancy using automated feature analyses algorithms [47], [96]. Here, the common approach has been to calculate many features by which to measure nodules and attempt to find correlations between particular features (e.g., size, shape, attenuation) and histological-confirmed cancers. Promising results have been demonstrated using classifiers based on classical nodule texture features [48]. More recently, fractal analysis of lung-nodule interfaces [49] and LDA of multiple features [50] have shown promise. Other important efforts in distinguishing benign and malignant nodules are measurement of size change over time [35], [46] and quantification of nodule uptake of intravenously administered contrast enhancement [51]. The solitary pulmonary nodule is a commonly encountered finding that might represent lung cancer. Morphological characteristics including lesion size, contour, edge, calcification, nodule density, and contrast enhancement can help differentiate malignant from benign nodules. Temporal change in lung nodule size raises concern for malignancy, while size stability is traditionally considered an indicator of benignity [52].

Yankelevitz *et al.* [55] sought to determine the accuracy of LDCT volumetric measurements of small pulmonary nodules to assess growth and malignancy via three-dimensional image extraction and isotropic resampling. The synthetic nodule studies revealed that volume could be measured accurately to within  $\pm 3\%$ .

Ko *et al.* [35] developed a CAD system that automatically identified nodules from chest CT, quantified nodule diameter, and estimated temporal change in size. High correlation between the algorithm and thoracic radiologists on change in nodule size was achieved (Spearman rank correlation coefficient = 0.932). The automated nodule detection system identified 86% of 370 nodules in 16 studies from eight patients with known nodules.

Li *et al.* [95] evaluated a system to investigate whether a CAD scheme can assist radiologists in distinguishing small benign from small malignant nodules on LDCT data. The dataset used consisted of 28 primary lung cancers (6–20 mm) and 28 benign nodules. Cancer cases included nodules with pure ground-glass opacity, mixed ground-glass opacity, and solid opacity. The AUC of the CAD scheme alone was 83.1% for distinguishing benign from malignant nodules. The average AUC value for radiologists was improved with the aid of the CAD scheme from 78.5% to 85.3% ( $p = 0.016$ ). The radiologists' diagnostic performance with the CAD scheme was more accurate than that of the CAD scheme alone ( $p < 0.05$ ) and that of radiologists alone. Li *et al.* [93] also described the current status of the development and evaluation of CAD schemes for the detection and characterization of lung nodules

in thin-section CT. They also reviewed a number of observer performance studies, in which it was attempted, to assess the potential for clinical usefulness of CAD schemes for nodule detection and characterization in thin-section CT.

Petkovska *et al.* [94] studied whether conventional nodule densitometry or contrast enhancement maps of indeterminate lung nodules can distinguish benign from malignant nodules. Conventional nodule densitometry was performed to obtain the maximum difference in mean enhancement values for each nodule from a circular ROI. The ROC curve for higher values of enhancement indicated malignancy, which had an AUC value of 76%. The visually scored magnitude of enhancement was found to be less effective in distinguishing malignant from benign lesions, with an AUC value of 62%. The visually scored pattern of enhancement was found to be more effective with an average AUC value of 79%.

Zhen *et al.* [114] proposed a new tumor growth measure for pulmonary nodules, which could account for tumor deformation using nonrigid registration combined with nodule detection and -segmentation. They proposed an adaptive doubling time measure and reported comparative results to the standard doubling time growth-rate measure. Results were based on two successive scans for ten benign and nine malignant nodule datasets.

#### D. Machine Learning for Lung Cancer Screening

In recent years, the machine-learning community developed sophisticated tools and learning paradigms to address the issue of CAD schemes that show clinically relevant performance measures. Feature selection methods and temporal learning schemes are being employed successfully for the task of nodule characterization and growth measurement. Recently, Vapnik *et al.* [123] proposed a new framework called learning with hidden information that would enable the integration of hidden information that could further improve CAD technology for lung cancer diagnosis.

In Barreno *et al.* [106], the authors described a theoretical analysis on how to combine classifiers with an optimal decision rule and optimal ROC curve. The combination of different CAD schemes also found interest in comparative CAD studies such as the ANODE09 study. The issue of unbalanced class distribution in medical diagnostic applications and different class importance needs to be addressed when developing CAD schemes for effective lung cancer screening. Most standard classification methods, however, are designed to maximize the overall accuracy and cannot incorporate different costs to different classes explicitly. Liu and Tan *et al.* [107] proposed a method to directly maximize the weighted specificity and sensitivity of the ROC curve. They reported excellent generalization properties with the ability to assign different error costs to different classes to account for the difference in the importance of the class distribution. Mozer *et al.* [108] took an approach of constrained optimization to obtain a reduced solution space that directly models the problem domain and has relevant performance characteristics on a specific target region of the ROC curve. They showed significant performance improvements in the domain of telecommunications that could also benefit the application domain of lung cancer screening.

Dreiseitle [109] proposed a learning scheme for training multiclass classifiers by maximizing the volume under the ROC surface, which could benefit multiclass lung nodule characterization tasks. Rather than having a binary decision on the malignancy of a test case, a multiclass grading on the malignancy decision would further provide additional measures that could improve lung cancer screening.

The mentioned CAD schemes were not applied directly to the domain of lung cancer screening, yet they provide theoretical justification and potential to be effective tools to improve CT lung cancer screening. One has to examine carefully the reported results and their applicability to lung cancer screening.

#### IV. CLINICALLY RELEVANT SENSITIVITY AND SPECIFICITY

We emphasize that for cancer screening and performance assessment of available CAD schemes, the focus of attention should be put on clinically relevant performance measures. In the context of cancer screening, the sensitivity-specificity calculus of CAD systems is an essential factor when it comes to treatment cost and patient outcome. One should consider that validation of false-positives and ground-truth generation is still a very timid approach. Problems of inter- and intraobserver variability (see Fig. 2) and manual time-intensive grading call for minimally invasive methodologies to obtain ground truth information that would further alter the sensitivity-specificity calculus of CAD and its potential acceptance. Minimally invasive lung cancer surgery such as thoracoscopic lobectomy or video-assisted thoracic surgery (VATS), or even noninvasive surgeries such as the CyberKnife method [121], are advances towards this direction.

To provide clinically relevant definitions for *sensitivity* and *specificity* [103]–[105], we follow the definition in [100] and [101] and point out how these diagnostic performance measures should be interpreted for cancer screening. *Sensitivity* and *specificity* measure the number of false positives and false negatives and are useful in evaluating the effectiveness of screening methods. Alternative terms are the *true-positive rate* (TPR) and the *false-positive rate* (FPR). The terms “positive” and “negative” are used to refer to the presence or absence of lung cancer.

Sensitivity and specificity are defined as follows. The sensitivity of a screening test is its ability to detect those individuals with cancer. It is computed by taking the number of true positives (TPs) and dividing it by the total number of cancer cases (TP + FN). The specificity of a test is its ability to identify those individuals who actually do not have cancer. It is computed by dividing the true negative (TN) by the sum of the TN and FP cases. From these probabilities, one can compute confidence intervals [102] and ROC curves that summarize diagnostic performance for comparative assessment. However, the majority of published research does not provide confidence intervals even though they could be obtained from the algorithms.

##### A. Detection Theory and ROC

The ROC curve was first developed by electrical engineers and radar engineers during World War II for detecting enemy objects in battlefields and was soon introduced in psychology to account for perceptual detection of signals [17], [18]. The use of ROC in medicine to assess diagnostic test performance was

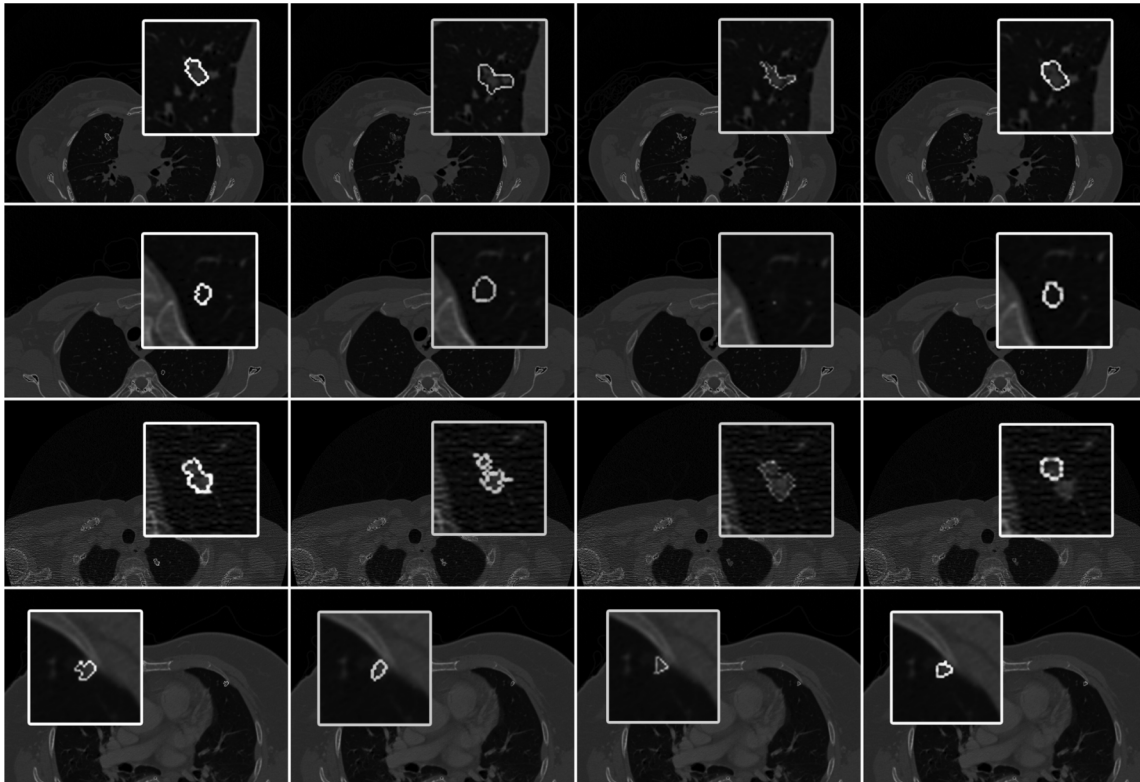


Fig. 2. Lung nodules and inter-reader variability. Axial slices of four different nodules (rows 1–4). Nodule regions of interest have been zoomed up for improved visualization. The columns denote the different lung nodule outlines provided by four different experts (columns 1–4). Columns 2 and 3 show examples of expert variability in providing lung nodule boundary outlines. Rows 1–3 show more central nodules, whereas row 4 shows an example of subpleural nodules.

first described by Lusted in 1971. Recent years have seen an increase in the use of ROC in the machine-learning community, due in part to the realization that simple classification accuracy is often a poor metric for measuring performance.

ROC curves are useful for organizing and visualizing classifier performance. The ROC curve is a plot of the TPR on the vertical axes versus the FPR on the horizontal axes for multiple classification rules. The farther up and to the left the ROC curve is located in ROC space, the better the performance of the screening test [103]. The likelihood ratio (LR) can be computed from the tangent slope at a cut-point of the ROC curve. The AUC, also known as the *c*-statistic, is a measure of the overall diagnostic accuracy of the test. Its value ranges from 0.5 to 1, where 0.5 denotes random chance and an AUC of 1 denotes perfect discrimination. Though ROC offers various advantages in the evaluation of CAD systems, several constraints limit the effective use of ROC for clinically relevant CAD evaluation in lung cancer screening. Issues of dissimilar or missing a priori probabilities, the lack of confidence intervals, and severe class imbalances need to be addressed in a statistically clinically relevant manner.

### B. Alternative Approaches in CAD Assessment

Baker *et al.* provided guidelines for clinically relevant FPR and TPR measures [104] as well as special ROC methods for cancer screening [105], [118]. Baker *et al.* also noted that since the underlying prevalence of cancer in average risk populations is very low, the FPR should be very small for acceptable cancer screening of asymptomatic people. Current practice when

comparing the performance of diagnostic tests is to compare the areas under the entire ROC curves. However, since the false-positive rate must be very small for effective cancer screening to be acceptable, most of the ROC curve is not relevant for comparing the performance of cancer screening tests. He proposes alternative ROC methods for cancer screening that base inference on the areas under the small part of the ROC curves near the false-positive rates that are relevant for cancer screening [105], [118]. Further, evaluation of CAD methods for lung cancer screening should not solely be based on the ROC curve alone. In assessing CAD performance for lung cancer screening, one should acknowledge that the time to event is the actual object under analysis, i.e., mortality [119]. ROC analysis methods have limitations when it comes to censored time data requiring specific statistical considerations to adjust for confounding factors. Solutions to this problem include statistical calibration methods such as clinical reclassification by performing risk stratification [118] or standard Kaplan–Meier or Cox regression analyses [122]. Cook [120] notes that in comparing different models, one would prefer those that stratify individuals correctly into various categories that are better calibrated. ROC statistics alone do not provide whether one model is better at classifying individuals, or if individual risk estimates differ between two models. Several ROC paradigms such as FROC, AFROC, JAFROC, LROC, and MRMC-ROC have been proposed to overcome the limitations of the standard ROC approach, yet most of the validation work reported so far is based on standard ROC analysis.

## V. INTEGRATION OF CAD TO IMPROVE SCREENING PARADIGMS

While CT lung cancer screening is being evaluated against CXR in clinical trials [6], CAD has also found use in other imaging domains to improve sensitivity and specificity in lung cancer screening. The potential of CAD for lung cancer screening becomes evident not only in CT but also in existing screening paradigms such as CXR, CT single photon emission computed tomography (SPECT) [124], and CT positron emission tomography (PET) [124], [125]. Several reports [61] show productive use of CAD by radiologists reading chest films.

Aoyama [50] developed an automated computerized “second opinion” scheme to assist radiologists in discriminating between benign and malignant solitary pulmonary nodules on traditional CXR. ROC analysis of their system showed slightly improved performance with CAD assistance.

From 1996 to 2001, CAD schemes were presented as interactive demonstrations on computer workstations at each annual scientific assembly of the Radiological Society of North America. Abe *et al.* in 2003 [56] summarized the results of these demonstrations. These informal experiments demonstrated increased sensitivity and statistically improved ROC areas among radiologists who participated. The same group demonstrated in a controlled experiment that radiologists improved their performance using CAD aids to distinguish benign from malignant nodules on chest radiographs [57].

CAD for CXR is currently investigated in clinical trials by Meziane *et al.* [99] to determine whether the technology can improve practical early detection of lung cancer. The trial is designed to determine whether CXR and CAD can help to identify hard-to-detect lung cancers at an early stage when they are most treatable, leading to improved patient survival rates. The study will involve 9000 subjects with a five-year follow-up. The current retrospective study is a multiple-reader multiple-cases study involving six expert chest radiologists, six general radiologists, and six pulmonologists to assess the usefulness of CAD. Each reader will read 200 cases of suspected lung cancer, of which about 100 had cancer and 100 did not. All cases were CT confirmed. All cancers are biopsy or surgically confirmed.

At present, there are two CAD application domains for which the FDA has given premarket approval: i) detection of breast cancer via mammography [77] and ii) detection of signs consistent with lung cancer on chest radiographs [90]. In parallel, commercial developments have undergone evaluation in radiology settings to assess the radiologists’ performance in detecting malignant lung lesions with and without the use of CAD. Commercial software to assist radiologists in reading CT images of the lungs typically focus on decreasing the frequency of missed nodules in large volume of image data produced by CT scans [22].

Wormanns *et al.* [53] assessed the performance of a Siemens LungCare CT CAD system against a radiologist-generated database of CT imaged nodules. CAD and individual radiologists participating in the evaluation performed similarly.

Beyer *et al.* [63] investigated the ability of a Siemens CAD system to localize pulmonary nodules previously identified on CT exams. Since high-resolution CT is a typical follow-up exam to a suspicious screening low-dose CT, accurately identifying the suspicious lesion on a follow-up CT is critical. A nodule

was considered correctly localized if the marking box drawn by the software was visible on at least one slice together with the nodule and the center of the nodule was localized inside the marking box. The software correctly located 164 of 190 of all lung nodules (86.3%).

Kakeda *et al.* [83] assessed the usefulness of a new commercially available CAD system with an automated method of detecting nodules. For patients with cancer, 45 cases with solitary lung nodules were used. For healthy patients, 45 cases were selected. Eight radiologists participated in observer performance studies and interpreted the original radiographs and CAD output images using a sequential testing method. The average AUC increased significantly from 92.4% without to 98.6% with CAD output images.

Sakai *et al.* [84] evaluated the usefulness of a commercially available CAD system on digital chest radiographs of patients with stage T1 lung carcinoma. Fifty patients underwent surgery for primary lung cancer, and 50 normal cases were selected. Eight radiologists participated in observer tests and interpreted soft copy images without and with CAD output. The overall detectability of lung cancer cases with CAD system was 74% (37/50), and the false-positive rate was 2.28 (114/50) false positives per case for normal cases. The mean AUC value increased significantly from 89.6% without CAD output to 92.3% with CAD output ( $P = 0.018$ ). The main cause of the improvement in performance is attributable to changes from false negatives without CAD to true positives with CAD (19/31, 61%). Moreover, improvement in the location of the tumor was observed in 1.5 cases, on average, for radiology residents.

Li *et al.* [81] examined the sensitivity and specificity of a commercially available CAD system for identifying lung cancers previously missed on chest radiographs by radiologists. A CAD nodule detection program was applied to 34 PA digital chest radiographs. All 34 radiographs showed a nodular lung cancer that was apparent in retrospect but had not been mentioned in the report. Two radiologists identified these radiologist-missed cancers on the chest radiographs and graded them for visibility, location, subtlety, and actionability. The CAD program had an overall sensitivity of 35% (12 of 34 cancers), identifying seven (30%) of 23 very subtle and five (45%) of 11 relatively obvious radiologist-missed cancers ( $p = 0.21$ ) and detecting two (25%) of eight missed not actionable and ten (38%) of 26 missed actionable cancers ( $p = 0.33$ ). The CAD program made an average of 5.9 false-positive marks per radiograph.

## VI. CURRENT CHALLENGES AND FUTURE DIRECTIONS

Is spiral CT technology poised for productive application to lung cancer screening? CAD is one technology that potentially could affect the sensitivity-specificity calculus of spiral CT for lung cancer screening. From commercial [22], [53], [63], [81], [83], [84] and academic [50], [57], [58], [60], [61], [87], [98] research developments, we certainly see that CAD technology has the potential to improve the performance of lung cancer diagnosis in CT and in CXR. CAD for lung CT has improved over the past decade and is now offered in early stage commercial systems aimed at reducing the reading burden on radiologists while decreasing the chance of missing significant nodules. Yet several technological challenges persist.



TABLE I  
SUMMARY OF CAD ALGORITHMS. (BTL = BRONCHIAL TREE LABELING, IL = INTERSTITIAL LESIONS, TD = TISSUE DISCRIMINATION, ND = NODULE DETECTION, NC = NODULE CLASSIFICATION, RA = RADIOLOGIST ASSIST, RAC = RADIOLOGIST ASSIST COMMERCIAL SOFTWARE, SE = SENSITIVITY, SP = SPECIFICITY)

Application	Method (Sample Size)	Measure 1	Measure 2
IL [101]	LDA, SVM (44 abnormal, 8 normal)	N/A	A(z)=0.78
TD [94]	Wavelets, KNN (N/A)	N/A	92.5%
TD [95]	SVM (N/A)	N/A	93%
ND [36]	Template matching (15 abnormal, 5 normal)	SE = 72% of nodules	SP=1.1 FPs/image
ND [38]	Thresholding (43 scans)	SE = 70%	SP=1.5 FPs/image
ND [39]	Rule-based classifier, LDA (34 patients, 63 nodules)	SE = 84%	SP=1.74 FPs/image
ND [43]	Difference imaging (73 abnormal, 21 normal)	SE = 88%	SP=1.9 FPs/scan
ND [45]	3D multi-template matching (N/A)	Increased FN by 2.3%	FPs/scan reduction by 44.3%
ND [35]	Size/shape detection (16 patients, 370 nodules)	SE=86%	System v. RA, CorCoeff = 93.2%
ND [85]	Multi-MTANN (N/A)	SE=81.3%	1.4 FPs/image
NC [98]	LDA (28 abnormal, 28 normal)	A(z)=83.1%	With CAD A(z) =85.3%
NC [2]	LDA, ANN (33 abnormal, 22 normal)	Az=0.886 with LDA	Az=0.872 with ANN
NC [97]	Contrast enhancement MAP (29 patients)	A(z) high enhancement=76%	Average A(z)=79%
RA [50]	LDA, ANN (33 abnormal, 22 normal)	A(z)=0.854 without CAD	A(z)=0.872 with CAD
RA [61]	Temporal subtraction (10 abnormal, 10 normal)	A(z)=0.920 without CAD	A(z)=0.980 with CAD
RA [59]	RA with and without CAD (50 cases)	N/A	N/A
RA [60]	RA with and without CAD (18 cases)	N/A	N/A
RA [62]	RA with and without CAD (15 patients, 77 nodules)	SE = 91% without CAD	SE=95% with CAD
RAC [53]	Siemens Lung CAD system (457 nodules)	SE = 79%	FPs 7.2% higher than RA
RAC [86]	RA with and without CAD (45 abnormal, 45 normal)	A(z)=92.4% without CAD	A(z)=98.6% with CAD
RAC [84]	RA with and without CAD (34 cases)	SE=35%	5.9 FPs/image

On the image processing and CAD methodology level discrimination of nodules attached to vessels or the pleural surface as well as low-contrast nodules still poses a challenge. Methods for detecting ground glass nodules and highly spiculated nodules with irregular boundaries still lack desired sensitivity and specificity performance.

Validation studies of the proposed CAD systems are still primitive lacking confidence intervals and validation statistics that provide clinically relevant performance measures for lung cancer screening. From the temporal snapshot presented in Section III comparative effectiveness research is hard to perform (see Table I) due to inconsistencies of reported validation results, small sample sizes, and experimental setup. The difficulty of obtaining ground truth is still a timid and invasive approach not to mention the issues of inter-reader variability between experts. However, since 2000, progress has been made to realize image databases [23], [65] to allow comparative assessment of various CAD schemes and coherent validation metrics. NELSON and ImageCLEF are examples of this promising research direction. In this regard, researchers are investigating the combination of CAD systems for optimized decision-making in lung cancer diagnosis. In the future, as new multimodal imaging techniques start to emerge, the extension of CAD for lung cancer screening to multimodal temporal data from SPECT/CT or PET/CT are promising directions.

## VII. SUMMARY

In this paper, we summarized the literature on CAD approaches to improve lung cancer diagnosis through lung tissue discrimination, nodule detection/classification, and nodule

characterization. CAD applications have shown high sensitivity but must also demonstrate high specificity to avoid cost-intensive, inconvenient, or even harmful follow-up procedures to rule out misclassified lung lesions. However, current protocols for nodule detection report many false positives, requiring substantial improvements in the technology and its applications prior to clinical application of CAD in the practice of lung cancer screening.

## ACKNOWLEDGMENT

The authors would like to thank A. Stein and L. Lamerato for providing important contextual information of CAD technology and valuable discussions.

## REFERENCES

- [1] P. M. Marcus, R. M. Fagerstrom, P. C. Prorok, J. K. Gohagan, and B. S. Kramer, "Screening for lung cancer with helical CT scanning," *Clin. Pulmon. Med.*, vol. 9, pp. 323–329, 2002.
- [2] M. Aoyama, Q. Li, S. Katsuragawa, H. MacMahon, and K. Doi, "Automated computerized scheme for distinction between benign and malignant solitary pulmonary nodules on chest images," *Med. Phys.*, vol. 29, pp. 701–8, 2002.
- [3] "Lung cancer screening: Recommendation statement," *Ann. Intern. Med.*, vol. 140, pp. 738–9, 2004.
- [4] S. Sone *et al.*, "Mass screening for lung cancer with mobile spiral computed tomography scanner," *Lancet*, vol. 351, pp. 1242–5, 1998.
- [5] C. I. Henschke *et al.*, "Early lung cancer action project: Overall design and findings from baseline screening," *Lancet*, vol. 354, pp. 99–105, 1999.
- [6] D. S. Gierada, T. K. Pilgrim, M. Ford, R. M. Fagerstrom, T. R. Church, H. Nath, K. Garg, and D. C. Strollo, "Lung cancer: interobserver agreement on interpretation of pulmonary findings at low-dose CT screening," *Radiol.*, vol. 246, no. 1, pp. 265–272, 2008.
- [7] B. S. Kramer and O. W. Brawley, "Cancer screening," *Hematol. Oncol. Clin. North Amer.*, vol. 14, pp. 831–48, 2000.



- [8] L. Berlin, "Liability of performing CT screening for coronary artery disease and lung cancer," *Amer. J. Roentgenol.*, vol. 179, pp. 837–42, 2002.
- [9] S. J. Swensen *et al.*, "Screening for lung cancer with low-dose spiral computed tomography," *Amer. J. Respir. Crit. Care Med.*, vol. 165, pp. 508–13, 2002.
- [10] S. J. Swensen *et al.*, "Lung cancer screening with CT: Mayo clinic experience," *Radiology*, vol. 226, pp. 756–61, 2003.
- [11] C. I. Henschke *et al.*, "CT screening for lung cancer: Suspiciousness of nodules according to size on baseline scans," *Radiology*, vol. 231, pp. 164–8, 2004.
- [12] R. MacRedmond *et al.*, "Screening for lung cancer using low dose CT scanning," *Thorax*, vol. 59, pp. 237–41, 2004.
- [13] J. Gohagan *et al.*, "Baseline findings of a randomized feasibility trial of lung cancer screening with spiral CT scan vs chest radiograph: The lung screening study of the National Cancer Institute," *Chest*, vol. 126, pp. 114–21, 2004.
- [14] B. J. Flehinger, M. Kimmel, and M. R. Melamed, "Survival from early lung cancer: Implications for screening," *Chest*, vol. 101, pp. 1013–1018, 1992.
- [15] L. L. Humphrey, S. Teutsch, and M. Johnson, "Lung cancer screening with sputum cytologic examination, chest radiography, and computed tomography: An update for the U.S. Preventive Services Task Force," *Ann. Intern. Med.*, vol. 140, pp. 740–53, 2004.
- [16] J. C. Nesbitt *et al.*, "Survival in early-stage lung cancer," *Ann. Thorac. Surg.*, vol. 60, pp. 466–472, 1995.
- [17] W. W. Peterson, T. G. Birdsall, and W. C. Fox, "The theory of signal detectability," *Trans. Prof. Group Inform. Theory*, vol. 4, 1954.
- [18] J. A. Swets, "ROC analysis applied to the evaluation of medical imaging techniques," *Invest. Radiol.*, vol. 14, pp. 109–21, 1979.
- [19] R. Sah *et al.*, "Results of surgical treatment of stage I and II lung cancer," *J. Cardiovasc. Surg.*, vol. 37, pp. 169–172, 1996.
- [20] R. S. Fontana *et al.*, "Lung cancer screening: The mayo program," *J. Occup. Med.*, vol. 28, pp. 746–750, 1986.
- [21] A. P. Reeves and W. J. Kostis, "Computer-aided diagnosis of small pulmonary nodules," *Semin. Ultrasound CT MR*, vol. 21, pp. 116–28, 2000.
- [22] A. P. Reeves and W. J. Kostis, "Computer-aided diagnosis for lung cancer," *Radiol. Clin. North Amer.*, vol. 38, pp. 497–509, 2000.
- [23] M. Freedman, "Improved small volume lung cancer detection with computer-aided detection: Database characteristics and imaging of response to breast cancer risk reduction strategies," *Ann. NY Acad. Sci.*, vol. 1020, pp. 175–89, 2004.
- [24] M. Kavuru *et al.*, "Three-dimensional imaging of the lung in vivo: Work in progress," *J. Dig. Imag.*, vol. 4, pp. 139–142, 1991.
- [25] K. Suzuki, E. Kohda, M. Izutsu, N. Shiraga, and K. Hiramatsu, "3D image, MIP and MPR for evaluation of spiral CT data of the chest," in *Proc. CAR'95*, 1995, pp. 1250–1251.
- [26] C. Schaefer, M. Prokop, and M. Galanski, "Minimum and maximum intensity projections for evaluation of spiral CT data of the chest," in *Proc. Adv. CT III*, Berlin, Germany, 1994, pp. 269–275.
- [27] M. L. Giger, K. T. Bae, and H. MacMahon, "Computerized detection of pulmonary nodules in computed tomography images," *Invest. Radiol.*, vol. 29, pp. 459–65, 1994.
- [28] J. Hasegawa, K. Mori, and J. Toriwaki, "Automated extraction of lung cancer lesions from multi-slice chest CT images by using three-dimensional image processing," *Trans. IEICE Jpn.*, pp. 1587–94, 1993.
- [29] K. Mori *et al.*, "Automated anatomical labeling of the bronchial branch and its application to the virtual bronchoscopy system," *IEEE Trans. Med. Imag.*, vol. 19, pp. 103–14, 2000.
- [30] P. Croisille *et al.*, "Pulmonary nodules: Improved detection with vascular segmentation and extraction with spiral CT," *Radiology*, vol. 197, pp. 397–401, 1995.
- [31] M. S. Brown *et al.*, "Method for segmenting chest CT image data using an anatomical model: Preliminary results," *IEEE Trans. Med. Imag.*, vol. 16, pp. 828–39, 1997.
- [32] D. Y. Kim, J. H. Kim, S. M. Noh, and J. W. Park, "Pulmonary nodule detection using chest CT images," *Acad. Radiol.*, vol. 44, pp. 252–7, 2003.
- [33] S. Toshioka *et al.*, "Computer aided diagnosis system for lung cancer based on helical CT images," *Proc. SPIE*, vol. 3034, pp. 975–984, 1997.
- [34] S. G. Armato *et al.*, "Computerized detection of pulmonary nodules on CT scans," *Radiographics*, vol. 19, pp. 1303–11, 1999.
- [35] J. P. Ko and M. Betke, "Chest CT: Automated nodule detection and assessment of change over time—preliminary experience," *Radiology*, vol. 218, pp. 267–73, 2001.
- [36] 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. Imag.*, vol. 20, pp. 595–604, 2001.
- [37] T. Okhumura *et al.*, "Image processing for computer-aided diagnosis of lung cancer screening system by CT (LSCT)," in *Proc. SPIE*, 1998, pp. 1314–1322.
- [38] S. G. Armato, M. L. Giger, and H. MacMahon, "Automated detection of lung nodules in CT scans: Preliminary results," *Med. Phys.*, vol. 28, pp. 1552–61, 2001.
- [39] M. N. Gurcan *et al.*, "Lung nodule detection on thoracic computed tomography images: Preliminary evaluation of a computer-aided diagnosis system," *Med. Phys.*, vol. 29, pp. 2552–8, 2002.
- [40] K. Suzuki, S. G. Armato, F. L. S. Sone, and K. Doi, "Massive training artificial neural network (MTANN) for reduction of false positives in computerized detection of lung nodules in low-dose computed tomography," *Med. Phys.*, vol. 30, pp. 1602–17, 2003.
- [41] B. Zhao, G. Gamsu, M. S. Ginsberg, L. Jiang, and L. H. Schwartz, "Automatic detection of small lung nodules on CT utilizing a local density maximum algorithm," *J. Appl. Clin. Med. Phys.*, vol. 4, pp. 248–60, 2003.
- [42] P. H. Winston, *Artificial Intelligence*. New York: Addison-Wesley, 1992.
- [43] H. Arimura *et al.*, "Computerized scheme for automated detection of lung nodules in low-dose computed tomography images for lung cancer screening," *Acad. Radiol.*, vol. 11, pp. 617–29, 2004.
- [44] S. G. Armato *et al.*, "Lung cancer: Performance of automated lung nodule detection applied to cancers missed in a CT screening program," *Radiology*, vol. 225, pp. 685–92, 2002.
- [45] Q. Li, S. Katsuragawa, and K. Doi, "Computer-aided diagnostic scheme for lung nodule detection in digital chest radiographs by use of a multiple-template matching technique," *Med. Phys.*, vol. 28, pp. 2070–6, 2001.
- [46] M. S. Brown *et al.*, "Patient-specific models for lung nodule detection and surveillance in CT images," *IEEE Trans. Med. Imag.*, vol. 20, pp. 1242–50, 2001.
- [47] D. Cavouras, P. Prassopoulos, and N. Pantelidis, "Image analysis methods for solitary pulmonary nodule characterization by computed tomography," *Eur. J. Radiol.*, vol. 14, pp. 169–72, 1992.
- [48] 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," *Comput. Med. Imag. Graph.*, vol. 23, pp. 339–48, 1999.
- [49] S. Kido, K. Kuriyama, M. Higashiyama, T. Kasugai, and C. Kuroda, "Fractal analysis of small peripheral pulmonary nodules in thin-section CT: Evaluation of the lung-nodule interfaces," *J. Comput. Assist. Tomogr.*, vol. 26, pp. 573–8, 2002.
- [50] M. Aoyama *et al.*, "Computerized scheme for determination of the likelihood measure of malignancy for pulmonary nodules on low-dose CT images," *Med. Phys.*, vol. 30, pp. 387–94, 2003.
- [51] K. Minami *et al.*, "Computerized characterization of contrast enhancement patterns for classifying pulmonary nodules," *IEEE Trans. Med. Imag.*, vol. 2, pp. 897–900, 2001.
- [52] D. Shaham and L. Guralnik, "The solitary pulmonary nodule: Radiologic considerations," *Semin. Ultrasound CT MR*, vol. 21, pp. 97–115, 2000.
- [53] D. Wormanns, F. Beyer, S. Diederich, K. Ludwig, and W. Heindel, "Diagnostic performance of a commercially available computer-aided diagnosis system for semi-automatic detection of pulmonary nodules: Comparison with single and double reading," *Fortschr. Röntgenstr.*, vol. 176, pp. 953–8, 2004.
- [54] I. Sluimer, A. Schilham, M. Prokop, and B. van Ginneken, "Computer analysis of computed tomography scans of the lung: A survey," *IEEE Trans. Med. Imag.*, vol. 4, pp. 385–405, 2006.
- [55] D. F. Yankelevitz, A. P. Reeves, W. J. Kostis, B. Zhao, and C. I. Henschke, "Small pulmonary nodules: Volumetrically determined growth rates based on CT evaluation," *Radiology*, vol. 217, pp. 251–6, 2000.
- [56] H. Abe *et al.*, "Computer-aided diagnosis in chest radiography: Results of large-scale observer tests at the 1996–2001 RSNA scientific assemblies," *Radiographics*, vol. 23, pp. 255–65, 2003.
- [57] J. Shiraiishi *et al.*, "Computer-aided diagnosis to distinguish benign from malignant solitary pulmonary nodules on radiographs: ROC analysis of radiologists' performance—Initial experience," *Radiology*, vol. 227, pp. 469–74, 2003.
- [58] K. Kubota *et al.*, "The results in the clinical trial of CAD system for lung cancer using helical CT images," *IEEE Trans. Med. Imag.*, vol. 1, pp. 313–316, 2001.

- [59] K. Awai *et al.*, "Pulmonary nodules at chest CT: Effect of computer-aided diagnosis on radiologists' detection performance," *Radiology*, vol. 230, pp. 347–52, 2004.
- [60] K. Marten *et al.*, "Computer-assisted detection of pulmonary nodules: Performance evaluation of an expert knowledge-based detection system in consensus reading with experienced and inexperienced chest radiologists," *Eur. Radiol.*, vol. 14, pp. 1930–8, 2004.
- [61] S. Kakeda *et al.*, "Improved detection of lung nodules by using a temporal subtraction technique," *Radiology*, vol. 224, pp. 145–51, 2002.
- [62] M. S. Brown *et al.*, "Lung micronodules: Automated method for detection at thin-section CT—Initial experience," *Radiology*, vol. 226, pp. 256–62, 2003.
- [63] F. Beyer *et al.*, "Clinical evaluation of a software for automated localization of lung nodules at follow-up CT examinations," *Rofo*, vol. 176, pp. 829–36, 2004.
- [64] C. I. Henschke *et al.*, "Computed tomography screening for lung cancer," *Clin. Chest Med.*, vol. 23, pp. 49–57, 2002.
- [65] L. P. Clarke, B. Y. Croft, E. Staab, H. Baker, and D. C. Sullivan, "National Cancer Institute Initiative: Lung image database resource for imaging research," *Acad. Radiol.*, vol. 8, pp. 447–50, 2001.
- [66] S. G. Armato *et al.*, "Lung image database consortium: Developing a resource for the medical imaging research community," *Radiology*, vol. 232, pp. 739–48, 2004.
- [67] A. K. Ganti and J. L. Mulshin, "Lung cancer screening," *Oncologist*, vol. 11, no. 5, pp. 481–487, 2006.
- [68] I. M. Burger, N. E. Kass, J. H. Sunshine, and S. S. Siegelman, "The use of CT for screening: A national survey of radiologists' activities and attitudes," *Radiology*, vol. 248, pp. 160–168, 2008.
- [69] J. Murtagh *et al.*, "CT and MRI for selected clinical disorders: A systematic review of economic evaluations," *Technol. Rep.*, no. 68.
- [70] C. I. Henschke *et al.*, "International early lung cancer action program investigators Survival of patients with stage I lung cancer detected on CT screening," *New Eng. J. Med.*, vol. 355, no. 17, pp. 1763–1771, 2006.
- [71] M. Unger, "A pause progress reassessment in lung cancer screening," *New Eng. J. Med.*, vol. 355, no. 17, pp. 1822–1824, 2006.
- [72] "No authors listed. Expert panel does not yet endorse CT screening for lung cancer," *Health Technol. Trends*, vol. 19, no. 2, pp. 4–5, 2007.
- [73] C. Black *et al.*, "Population screening for lung cancer using computed tomography is there evidence of clinical effectiveness?," *System. Rev. Lit., Thorax*, vol. 62, no. 2, pp. 131–138, 2007.
- [74] W. C. Black and J. A. Baron, "CT screening for lung cancer: Spiraling into confusion?," *J. Amer. Med. Assoc.*, vol. 297, no. 9, pp. 995–997, 2007.
- [75] W. M. Alberts, "Diagnosis and management of lung cancer executive summary," *ACCP Evidence-Based Clinical Practice Guidelines (2nd Ed.)*, *Chest*, vol. 132, no. 3, pp. 1S–19S, 2007.
- [76] World Health Organization, Cancer 2008.
- [77] R. M. Nishikawa, "Current status and future directions of computer-aided diagnosis in mammography," *Comput. Med. Imag. Graph.*, vol. 31, pp. 224–235, 2007.
- [78] National Cancer Institute, A snapshot of lung cancer 2006.
- [79] T. C. Kennedy, Y. Miller, and S. Prindiville, "Screening for lung cancer revisited and the role of sputum cytology and fluorescence bronchoscopy in a high-risk group," *Chest*, vol. 117, no. 4, pp. 725–795, 2000.
- [80] P. B. Bach *et al.*, "Computed tomography screening and lung cancer outcomes," *J. Amer. Med. Assoc.*, vol. 297, no. 9, pp. 953–961, 2007.
- [81] F. Li *et al.*, "Lung cancers missed on chest radiographs: Results obtained with a commercial computer-aided detection program," *Radiology*, vol. 246, no. 1, pp. 273–280, 2008.
- [82] K. Suzuki *et al.*, "False-positive reduction in computer-aided diagnostic scheme for detecting nodules in chest radiographs by means of massive training artificial neural network," *Acad. Radiol.*, vol. 12, no. 2, pp. 191–201, 2005.
- [83] S. Kakeda *et al.*, "Improved detection of lung nodules on chest radiographs using a commercial computer-aided diagnosis system," *Amer. J. Roentgenol.*, vol. 182, no. 2, pp. 505–510, 2004.
- [84] S. Sakai *et al.*, "Computer-aided nodule detection on digital chest radiography: Validation test on consecutive T1 cases of resectable lung cancer," *J. Dig. Imag.*, vol. 19, no. 4, pp. 376–382, 2006.
- [85] J. J. Chen and C. S. White, "Use of CAD to evaluate lung cancer on chest radiography," *J. Thorac Imag.*, vol. 23, no. 2, pp. 93–6, 2008.
- [86] H. P. Chan *et al.*, "Computer-aided diagnosis of lung cancer and pulmonary embolism in computed tomography—A review," *Acad. Radiol.*, vol. 15, pp. 535–555, 2008.
- [87] J. Shiraishi *et al.*, "Computer-aided diagnosis for the detection and classification of lung cancers on chest radiographs ROC analysis of radiologists' performance," *Acad. Radiol.*, vol. 13, no. 8, pp. 995–1003, 2006.
- [88] S. Wolff, "Screening of lung cancer samples using gene expression profiling," *Nature Gen.*, vol. 27, no. 96, 2001.
- [89] K. Doi, "Computer-aided diagnosis in medical imaging: Historical review, current status and future potential," *Comput. Med. Imag. Graph.*, vol. 31, no. 4–5, pp. 198–211, 2007.
- [90] R. R. Bharat *et al.*, "Lungcad: A clinically approved, machine learning system for lung cancer detection," in *Proc. ACM KDD '07*, 2007, pp. 1033–1037.
- [91] A. Depeursinge *et al.*, "Lung tissue classification using wavelet frames," in *Proc. 29th Annu. Int. Conf. IEEE EMBS*, 2007, pp. 6259–6262.
- [92] A. Depeursinge *et al.*, "Lung tissue classification in HRCT data integrating the clinical context," in *Proc. 21st IEEE Int. Symp. Comput.-Based Med. Syst.*, 2008, pp. 542–547.
- [93] L. Qiang, "Recent progress in computer-aided diagnosis of lung nodules on thin-section CT," *Comput. Med. Imag. Graph.*, vol. 31, no. 4–5, pp. 248–257, 2007.
- [94] I. Petkovska *et al.*, "Pulmonary nodule characterization: A comparison of conventional with quantitative and visual semi-quantitative analysis using contrast enhancement maps original research," *Eur. J. Radiol.*, vol. 59, no. 2, pp. 244–252, 2006.
- [95] F. Li *et al.*, "Radiologists' performance for differentiating benign from malignant lung nodules on high-resolution CT using computer-estimated likelihood of malignancy," *Amer. J. Roentgenol.*, vol. 183, pp. 1209–1215, 2004.
- [96] Y. Kawata *et al.*, "Hybrid classification approach of malignant and benign pulmonary nodules based on topological and histogram features," in *Proc. MICCAI*, 2000, pp. 297–306.
- [97] J. Shiraishi *et al.*, "Computer-aided diagnosis for improved detection of lung nodules by use of PA and lateral chest radiographs," *Acad. Radiol.*, vol. 14, no. 1, pp. 28–37, 2007.
- [98] Y. Arzhaeva *et al.*, "Computer-aided detection of interstitial abnormalities in chest radiographs using a reference standard based on computed tomography," *Med. Phys.*, vol. 34, no. 12, pp. 4798–4809, 2007.
- [99] "Cleveland clinic leading clinical program to improve early-stage lung cancer detection," in *Proc. 93rd Sci. Assemb. Annu. Meeting Radiol. Soc. North Amer. (RSNA)*, 2007.
- [100] C. H. Yarbro, M. Goodman, and M. H. Frogge, *Cancer Nursing: Principles and Practice*. Boston, MA: Jones & Bartlett, 2005.
- [101] D. G. Altman and J. M. Bland, "Diagnostic tests 1: Sensitivity and specificity," *Br. Med. J.*, vol. 308, pp. 1552–1552, 1994.
- [102] M. J. Gardner and D. G. Altman, "Calculating confidence intervals for proportions and their differences," in *Statistics With Confidence*, M.J. Gardner and D.G. Altman, Eds. London, U.K.: BMJ, 1989, pp. 28–33.
- [103] S. G. Baker, "Identifying combinations of cancer markers for further study as triggers of early intervention," *Biometrics*, vol. 56, pp. 1082–1087, 2000.
- [104] S. G. Baker, B. S. Kramer, and S. Srivastava, "Markers for early detection of cancer: Statistical guidelines for nested case-control studies," *BMC Med. Res. Methodol.*, vol. 2, no. 4, 2002.
- [105] S. G. Baker and P. F. Pinsky, "A proposed design and analysis for comparing digital and analog mammography: Special receiver operating characteristic methods for cancer screening," *J. Amer. Statist. Assoc.*, vol. 96, no. 454, pp. 421–428, 2001.
- [106] M. Barreno, A. A. Cardenas, and J. D. Tygar, "Optimal ROC curve for a combination of classifiers," in *Proc. NIPS*, 2007, pp. 57–64.
- [107] Z. Liu and M. Tan, "ROC-based utility function maximization for feature selection and classification with applications to high-dimensional protease data, biometrics," *J. Int. Biometr. Soc.*, vol. 64, no. 4, pp. 1155–1161, 2008.
- [108] M. C. Mozer, R. Dodier, M. D. Colagrosso, C. G. Salcedo, and R. Wolniewicz, "Prodding the ROC curve: Constrained optimization of classifier performance," in *Proc. NIPS*, 2001, pp. 1409–1415.
- [109] S. Dreiseitl, "Training multiclass classifiers by maximizing the volume under the ROC surface," in *Proc. EUROCAST*, 2007, vol. 4739, pp. 878–885.
- [110] M. M. Dundar, G. Fung, B. Krishnapuram, and R. B. Rao, "Multiple-instance learning algorithms for computer-aided detection," *IEEE Trans. Biomed. Eng.*, vol. 55, pp. 1015–1021, 2008.
- [111] N. Kato, M. Fukui, and T. Isozaki, "Bag-of-features approach for improvement of lung tissue classification in diffuse lung disease," in *Proc. SPIE Med. Imag. Comput.-Aided Diagnos.*, 2009, vol. 7260, pp. 72600C–72600C.

- [112] K. Murphy *et al.*, "A large-scale evaluation of automatic pulmonary nodule detection in chest CT using local image features and k-nearest-neighbor classification," *Med. Image Anal.*, vol. 13, no. 5, pp. 757–770, 2009.
- [113] M. Dolejsi, J. Kybic, M. Polovincak, and S. Tuma, "The lung time: Annotated lung nodule dataset and nodule detection framework," *Proc. SPIE Med. Imag. Comput.-Aided Diagnos.*, vol. 7260, 2009.
- [114] Y. J. Zheng, C. Kambhamettu, T. Bauer, and K. Steiner, "Accurate estimation of pulmonary nodule's growth rate in CT images with nonrigid registration and precise nodule detection and segmentation," in *Proc. MMBIA*, 2009, pp. 101–108.
- [115] A. El-Baz, G. Gimelfarb, R. Falk, and M. A. El-Ghar, "Automatic analysis of 3D low dose CT images for early diagnosis of lung cancer," *Pattern Recognit.*, vol. 42, no. 6, pp. 1041–1051, 2009.
- [116] E. M. van Rikxoort, B. de Hoop, S. van de Vorst, M. Prokop, and B. van Ginneken, "Automatic segmentation of pulmonary segments from volumetric chest CT scans," *Med. Imag.*, vol. 28, no. 4, pp. 621–630, 2009.
- [117] J. T. Pu *et al.*, "A computational geometry approach to automated pulmonary fissure segmentation in CT examinations," *Med. Imag.*, vol. 28, no. 5, pp. 710–719, 2009.
- [118] S. G. Baker, "The central role of receiver operating characteristic (ROC) curves in evaluating tests for the early detection of cancer," *J. Nat. Cancer Inst.*, vol. 95, no. 7, pp. 511–515, 2003.
- [119] S. Kjetil, "Receiver-operating characteristic (ROC) curve analysis in diagnostic, prognostic and predictive biomarker research," *J. Clin. Pathol.*, vol. 62, no. 1, pp. 1–5, 2009.
- [120] N. R. Cook, "Statistical evaluation of prognostic versus diagnostic models: Beyond the ROC curve," *Clin. Chem.*, vol. 54, no. 1, pp. 17–23, 2008.
- [121] W. Hara, S. G. Soltys, and I. C. Gibbs, "Cyberknife robotic radiosurgery system for tumor treatment," *Expert Rev. Anticancer Ther.*, vol. 7, no. 11, pp. 1507–1515, 2007.
- [122] G. Tripepi, J. J. Kitty, W. D. Friedo, and Z. Carmine, "Diagnostic methods 2: Receiver operating characteristic (ROC) curves," *Kidney Int.*, vol. 76, pp. 252–256, 2009.
- [123] V. Vapnik, A. Vashist, and N. Pavlovitch, "Learning using hidden information (learning with teacher)," in *Proc. Int. Joint Conf. Neural Netw.*, Atlanta, GA, Jun. 2009, pp. 14–19.
- [124] K. K. Wong *et al.*, "Hybrid SPECT-CT and PET-CT imaging of differentiated thyroid carcinoma," *Br. J. Radiol.*, vol. 82, pp. 860–876, 2009.
- [125] H. Schoeder and M. Goenen, "Screening for cancer with PET and PET/CT: Potential and limitations," *J. Nucl. Med.*, vol. 48, no. 1, pp. 4–18, 2007.



**Noah Lee** (S'05) received the B.S. and M.S. degrees in 2004 from the Computer Science Department of Beuth University of Applied Sciences for Technology Berlin, Germany, and the M.S. and M.Phil. degrees in biomedical engineering from Columbia University, New York, NY, in 2006.

He is currently a graduate research assistant in the Biomedical Engineering Department at Columbia University. His research interests are in medical image analysis, computer-aided diagnosis, and machine learning. He is currently exploring image

segmentation algorithms for cardiac, ophthalmic, and oncologic imaging applications.



**Andrew F. Laine** (S'82–M'83–SM'07) received the D.Sc. degree from Washington University, School of Engineering and Applied Science in Computer Science, St. Louis, MO, in 1989 and the B.S. degree from Cornell University, Ithaca, NY.

He is Director of the Heffner Biomedical Imaging Laboratory in the Department of Biomedical Engineering at Columbia University in New York City and is Professor of Biomedical Engineering and Radiology (Physics). His research interests include quantitative image analysis; cardiac functional imaging; ultrasound and MRI, retinal imaging, intravascular imaging and biosignal processing.

Dr. Laine served as: Chair of Technical Committee (TC-BIIP) on Biomedical Imaging and Image Processing for the EMBS (2006–2009); and on the IEEE ISBI (International Symposium on Biomedical Imaging) steering committee since 2006 and Program Chair for the IEEE EMBS annual conference in 2006 (New York City), Program Co-Chair for IEEE ISBI in 2008 (Paris, France); Vice President of Publications for the IEEE Engineering in Medicine and Biology Society since 2008. He is a Fellow of AIMBE.



**Guillermo Márquez** received the B.S. and Ph.D. degrees in biomedical engineering in 1995 and 2001 from Texas A&M University, College Station, TX. His doctoral research involved the development of an optical probe for cancer detection using multiwavelength oblique incidence reflectometry.

He is a program director and project officer in the Early Detection Research Group in the Division of Cancer Prevention of the National Cancer Institute (NCI). His main interest are in imaging modalities for cancer detection and screening, digital pathology,

and informatics.



**Jeffrey M. Levsy** is a graduate of the Medical Scientist Training Program at Albert Einstein College of Medicine, Bronx, NY, and received the M.D. and Ph.D. (anatomy/structural biology) degrees in 2004. He subsequently trained in the Diagnostic Radiology Residency at Montefiore Medical Center, Bronx, NY.

He is currently an Assistant Professor of Cardiothoracic Imaging in the Department of Radiology at Montefiore/Einstein with a research program involving imaging and clinical outcomes.



**John K. Gohagan** is chief of the Basic Prevention Science Research Group in the Division of Cancer Prevention, National Cancer Institute. Prior to this position, he was chief of Early Detection Branch in the Division of Cancer Prevention, NCI. Prior to coming to the National Institutes of Health, he was Professor of Engineering and Applied Science, Associate Professor of Radiology, and Associate Professor of Preventive Medicine, Washington University, St. Louis, MO.