Published in final edited form as:

Acad Radiol. 2016 February ; 23(2): 186–191. doi:10.1016/j.acra.2015.10.014.

After Detection: The Improved Accuracy of Lung Cancer Assessment Using Radiologic Computer-aided Diagnosis

Guy J. Amir, MD, MPH and Harold P. Lehmann, MD, PhD

Division of Health Sciences Informatics, Johns Hopkins University, 2024 East Monument Street, Suite 1-200, Baltimore, MD 21205, USA

Abstract

Rationale and Objectives—The aim of this study was to evaluate the improved accuracy of radiologic assessment of lung cancer afforded by computer-aided diagnosis (CADx).

Materials and Methods—Inclusion/exclusion criteria were formulated, and a systematic inquiry of research databases was conducted. Following title and abstract review, an in-depth review of 149 surviving articles was performed with accepted articles undergoing a Quality Assessment of Diagnostic Accuracy Studies (QUADAS)-based quality review and data abstraction.

Results—A total of 14 articles, representing 1868 scans, passed the review. Increases in the receiver operating characteristic (ROC) area under the curve of .8 or higher were seen in all nine studies that reported it, except for one that employed subspecialized radiologists.

Conclusions—This systematic review demonstrated improved accuracy of lung cancer assessment using CADx over manual review, in eight high-quality observer-performance studies. The improved accuracy afforded by radiologic lung-CADx suggests the need to explore its use in screening and regular clinical workflow.

Keywords

Computer-aided; imaging; medical; lung; cancer

INTRODUCTION

Lung cancer is the second leading cause of death in the United States and among the top 10 worldwide. More Americans die each year from lung cancer than from breast, prostate, and colorectal cancers combined. Annually, lung cancer kills more men than prostate cancer and more women than breast cancer (1).

Whereas overall cancer incidence rates are declining, lung cancer incidence rates among women are rising. Between 1960 and 1990, deaths from lung cancer among women increased over 400%. It is the second most common cancer among African American men and kills more African Americans than any other cancer. Five-year survival ranges from

70% for stage I disease to less than 5% for stage IV disease. As of 2014, overall 5-year survival is 17%, with only 15% diagnosed at the localized stage (2).

In this paper, “CADE” is defined as computer-aided detection and “CADx” as computer-aided diagnosis, and unless otherwise specified, will refer to both radiographic and computed tomography (CT) scans of the lungs. These are software outputs that a radiologist supervises/evaluates while viewing the image himself before final assessment. A CADE system detects abnormal nodules, without discriminating malignant from benign. CADx involves further interpretation as to likelihood of cancer. Unsupervised CADx (UCAD) is a standalone reader, a potential technological evolution (cf. Fig 1).

If diagnosed at the early stage (1A, < 3 cm), curative resection of stage I non-small cell lung cancer affords the survival rate of 70–80% (3). Screening current/former heavy smokers aged 55–74 years with low-dose CT by The National Cancer Institute’s National Lung Screening Trial (53,454 participants) demonstrated that those who received low-dose CT scans had a 20% lower risk of dying from lung cancer than participants who received chest X-rays (4). Chest radiography (CXR) screening remains controversial (e.g. see Hoop et al. (5)). Because all the above studies excluded CADx, improved outcome is possible. Because of the radiologist workload (6) and false-positive rate after biopsy of up to 50% (7), CADx continues to be investigated.

To our knowledge, there has never been a systematic review focused on CADx (CXR + CT) diagnostic accuracy. The novelty in our paper is in the exhaustive search we performed.

The goal of CADE is to differentiate true nodules from normal lung structures with the key outcome being sensitivity and false-positive rate that are reported per scan (patient). The comparative gold standard for CADE is a consensus panel. CADx studies are about correctly classifying/assessing detected nodules as high risk of malignancy (actionable) or low risk of malignancy (no immediate workup indicated, i.e. nonactionable). The comparative gold standard reference for CADx used by the studies is biopsy.

There are different diagnostic CADE/x models described in the literature. Both CADE and CADx employ at a minimum, a two-phase approach (cf. Fig 2). The first phase is segmentation with subtraction/difference image processing using gray-level picture thresholding techniques. The focus of the second phase is feature extraction to reduce false positives and involves predictive modeling techniques such as support vector machines (employing Gaussian, linear, or polynomial kernels), artificial neural networks (ANNs), cluster analysis, Bayesian Wavelet Snake, or other techniques.

CADx involves those systems that use the data acquired from the second phase beyond detection in a third phase to classify nodules as actionable or nonactionable. Rule-based methods, ANN, discriminant analysis, and other classifier techniques are employed in this third phase. CADx represents the entire scheme as depicted on the left in Figure 2.

The goals of this systematic review were to ascertain whether and by how much CADx improves the accuracy of lung cancer assessment over that of radiologists working without

the technology. The modalities included are radiography, low-dose computed tomography (LDCT), and high-resolution computed tomography (HRCT).

MATERIALS AND METHODS

The design of this systematic review was informed by Cochrane guidelines (8). We aimed to include all peer-reviewed journal articles that contained original data on the diagnostic performance of CADx. Databases searched included PUBMED, BioMed Central, the Cochrane Library, CINAHL/CINAHL PLUS, EMBASE, IEEE Xplore, INSPEC, JHSearch, and Web of Science.

Inclusion criteria were comparative evaluations of CADx for lung cancer and the use of CXR or CT. To increase sensitivity of the search, we included “detection” in the search algorithm. Search strategies for each reference database followed the following pattern:

Computer-aided (Detection OR Diagnosis) (Lung Neoplasm) OR (Lung Nodule) (Radiography OR CT)

Whenever possible, controlled vocabularies were utilized. When not available, synonyms were tested to increase sensitivity (see online supplement for all search strategies). For example, using MeSH for PubMed, the inclusion criteria became:

Computer-aided (Detection OR Diagnosis) AND (“Lung Neoplasms”[MeSH] OR Lung Nodule) AND (“Radiography”[MeSH] OR “Tomography, X-Ray Computed” [MeSH]) NOT review[PT]

Additionally, a manual search of the articles’ references was conducted to extract further eligible articles. Exclusion criteria were applied after reading titles and abstracts (if needed), and again after surviving articles were read, were any of:

1. Nonoriginal data
2. In title, nothing related to the thoracic region nor computer-aided systems
3. Lack of quantitative data related to CADx, e.g. solely CADE
4. Absence of any quantitative data related to accuracy
5. Absence of reference to either CT or CXR modalities
6. Duplicates

Abstraction forms included a quality scoring form based on the validated Quality Assessment of Diagnostic Accuracy Studies (QUADAS) scale for rating studies of diagnostic accuracy (9) and a content form with design and result tables for statistical analysis. Design data and Accuracy data tables were assembled. Specific factors abstracted from each paper were design data, verification tests, algorithms employed, and statistical data to evaluate improved accuracy. Different accuracy measures were sought, including sensitivity per scan at cutoff, false-positive reading/scan (test positive outputs for normal lung structure or benign lesions), accuracy $(TP+TN)/(P+N)$, where TP = true-positive, TN = true-negative, P = total-positive, N = total-negative, and receiver operating characteristic (ROC)-area index A_z , (area under the ROC equal to the probability that a system will rate a

randomly chosen positive instance higher than a randomly chosen negative one). A pilot study of 10 articles was used to improve the quality and content article abstraction forms and to ensure inclusion of all relevant variables.

RESULTS

From the reference databases, 444 articles were obtained by the unified search strategies. After title and abstract review (first review), 295 were excluded with over 60% duplicates. The full texts of the surviving 149 articles were reviewed (cf. Fig 3).

After the in-depth and reference review, 135 articles failed exclusion criteria #3 (lack of quantitative data), and 28 references failed #6 (duplicates), thus leaving 14 accepted articles (10–23) that were analyzed for quality and content as shown in Tables 1 and 2.

Nine studies used observer-performance with ROC analyses, meaning a radiologist's accuracy was evaluated first without CADx and then with CADx, and ROC curves were obtained by varying CADx cut-off values while measuring true-positive-fraction as a function of false-positive-fraction. The five remaining studies were unsupervised studies, all utilizing pathological diagnosis as verification. QUADAS scores ranged from 7 to 18 on a 21-point scale, with the most common methodological deficit being lack of a separate accuracy comparisons.

The nine observer-performance studies averaged 10.3 participant radiologists per study. Eight of the observer-performance studies showed significant accuracy improvement ($P = .002-.020$). Although the radiologists in one study (18) did not show significant improvement, these were thoracic specialists.

Several of the lung CT studies included similar authorship with the possibility of minor overlap of some of the data sets such as Ref. 13,14 and Ref. 16,20. However, in all cases, the studies were different, with clearly unique results as depicted in Table 2. Again, duplicate studies were eliminated by the title/abstract and full reviews.

In terms of the technologies in Figure 2, all systems utilized segmentation and feature extraction with several other predictive modeling techniques. Seven studies employed ANN/Massive Training Artificial Neural Network (MTANN) in the assessment algorithms (11–13,16,18–20), six studies limited to linear classifiers (10,15,17,21–23), and one employed a decision tree (14). There were insufficient data to discern differences in diagnostic accuracy as a function of the different categories of algorithms.

DISCUSSION

This systematic review, unique in its focus on lung cancer CADx, combines 14 studies to conclude that there is important improvement in accuracy afforded by CADx to the assessments by general radiologists.

Among the 14 studies, the accuracy of lung cancer assessment using CADx has been demonstrated in eight observer-performance studies, all with QUADAS scores above 70%.

The single study (18) that did not show significant improvement was limited because it involved thoracic specialists as the supervisors rather than general radiologists. The same study did however show significant improvement among a subgroup of resident radiologists ($P < .009$).

All classifier algorithms except for the decision tree (15) employed elements of statistical machine learning and computer vision.

This systematic review is unique in its focus on lung-CADx. Chan et al. (24), in their related lung review, produced a meta-analysis of lung cancer and pulmonary embolism, and despite the breadth of technologies included, they exclude radiography and do not use systematic methods such as inclusion/exclusion criteria, methodological quality assessment, and so on. Another review (25) of 18 breast studies and only 3 lung studies used QUADAS but lacks exhaustive search (lacked radiography and excluded critical lung studies). We have demonstrated improved accuracy of lung cancer computer-aided diagnosis and suggest that it is ready for broader screening.

CADx has direct implications for other cancers. CAdE, the precursor technology, has gained significant traction for mammography, achieving Food and Drug Administration approval in 1998. In 2006, Hologic, Inc. reported having sold over 3000 ImageChecker Mammography CAD systems world-wide. Reports indicate a 20% gain in early breast cancer detection (26,27). Although CADx is applicable to numerous organ systems, mammo-CAdE followed by colon and lung have led current research efforts. The future is bright for CADx applications to other organ systems.

In addition, new, nonradiologic screening modalities different from those studied here are available in the marketplace. A predictive model using nine volatile organic compounds delivered sensitivity and specificity above 80% for primary lung cancer (28). Emerging proteomic CADx technology may one day find application to lung cancer screening as they have with ovarian cancer (29).

CONCLUSION

We completed a unique and exhaustive systematic review of research of relevant medical and engineering databases for radiologic assessment of lung cancer afforded by CADx.

The improved accuracy of supervised lung-CADx systems over traditional reads, as shown in eight studies with high QUADAS scores, suggests that inclusion of CADx in strategies for lung cancer screening studies and regular clinical workflow may be warranted. This recommendation applies to general radiologists. Advances in accuracy must be achieved before these systems can significantly augment the performance of thoracic subspecialty-trained radiologists.

ACKNOWLEDGEMENT

The authors thank NLM (training grant T15LM007452) for financial support.

REFERENCES

1. Fry WA, Menck HR, Winchester DP. The National Cancer Data Base report on lung cancer. *Cancer*. 1996; 77:1947–1955. [PubMed: 8646697]
2. American Cancer Society's 2014 Cancer Facts & Figures Annual Report. Available at: <http://www.cancer.org/Research/CancerFactsStatistics/CancerFactsFigures2014/cancer-facts-and-figures-2014.pdf>.
3. Flehinger BJ, Kimmel M, Melamed MR. The effect of surgical treatment on survival from early lung cancer. Implications for screening. *Chest*. 1992; 101:1013–1018. [PubMed: 1313349]
4. Aberle DR, Adams AM, et al. National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med*. 2011; 365:395–409. [PubMed: 21714641]
5. de Hoop B, Schaefer-Prokop C, Gietema HA, et al. Screening for lung cancer with digital chest radiography: sensitivity and number of secondary work-up CT examinations. *Radiology*. 2010; 255:629–637. [PubMed: 20413773]
6. McLeod N, Montane G. The radiologist assistant: the solution to radiology workforce needs. *Emerg Radiol*. 2010; 17:253–256. [PubMed: 17033795]
7. Swensen SJ, Viggiano RW, Midthun DE, et al. Lung nodule enhancement at CT: multicenter study. *Radiology*. 2000; 214:73–80. [PubMed: 10644104]
8. van Tulder M, Furlan A, Bombardier C, et al. Updated method guidelines for systematic reviews in the Cochrane collaboration back review group. *Spine*. 2003; 28:1290–1299. [PubMed: 12811274]
9. Whiting P, Rutjes AW, Reitsma JB, et al. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol*. 2003; 3:25. [PubMed: 14606960]
10. Armato SG 3rd, Altman MB, Wilkie J, et al. Automated lung nodule classification following automated nodule detection on CT: a serial approach. *Med Phys*. 2003; 30:1188–1197. [PubMed: 12852543]
11. Arimura H, Katsuragawa S, Suzuki K, et al. Computerized scheme for automated detection of lung nodules in low-dose computed tomography images for lung cancer screening. *Acad Radiol*. 2004; 11:617–629. [PubMed: 15172364]
12. Li F, Arimura H, Suzuki K, et al. Computer-aided detection of peripheral lung cancers missed at CT: ROC analyses w/without localization. *Radiology*. 2005; 237:684–690. [PubMed: 16244277]
13. Li Q, Li F, Suzuki K, et al. Computer-aided diagnosis in thoracic CT. *Semin Ultrasound CT MR*. 2005; 26:357–363. [PubMed: 16274004]
14. Shah SK, McNitt-Gray MF, De Zoysa KR, et al. Solitary pulmonary nodule diagnosis on CT: results of an observer study. *Acad Radiol*. 2005; 12:496–501. [PubMed: 15831424]
15. Shah SK, McNitt-Gray MF, Rogers SR, et al. Computer-aided diagnosis of the solitary pulmonary nodule. *Acad Radiol*. 2005; 12:570–575. [PubMed: 15866129]
16. Suzuki K, Li F, Sone S, et al. Computer-aided diagnostic scheme for distinction between benign and malignant nodules in thoracic low-dose CT by use of massive training artificial neural network. *IEEE Trans Med Imaging*. 2005; 24:1138–1150. [PubMed: 16156352]
17. Shiraishi J, Abe H, Li F, et al. Computer-aided diagnosis for the detection and classification of lung cancers on chest radiographs: ROC analysis of radiologists' performance. *Acad Radiol*. 2006; 13:995–1003. [PubMed: 16843852]
18. Awai K, Murao K, Ozawa A, et al. Pulmonary nodules: estimation of malignancy at thin-section helical CT—effect of computer-aided diagnosis on performance of radiologists. *Radiology*. 2006; 239:276–284. [PubMed: 16467210]
19. Chen H, Wang XH, Ma DQ, et al. Neural network-based computer-aided diagnosis in distinguishing malignant from benign solitary pulmonary nodules by computed tomography. *Chin Med J*. 2007; 120:1211–1215. [PubMed: 17697569]
20. Suzuki K. A supervised “lesion-enhancement” filter by use of a massive-training artificial neural network (MTANN) in computer-aided diagnosis (CAD). *Phys Med Biol*. 2009; 54:S31–S45. [PubMed: 19687563]

21. Kusano S, Nakagawa T, Aoki T, et al. Efficacy of computer-aided diagnosis in lung cancer screening with low-dose spiral computed tomography: receiver operating characteristic analysis of radiologists' performance. *Jpn J Radiol.* 2010; 28:649–655. [PubMed: 21113748]
22. Way T, Chan HP, Hadjiiski L, et al. Computer-aided diagnosis of lung nodules on CT scans: ROC study of its effect on radiologists' performance. *Acad Radiol.* 2010; 17:323–332. [PubMed: 20152726]
23. Lee KH, Goo JM, Park CM, et al. Computer-aided detection of malignant lung nodules on chest radiographs: effect on observers' performance. *Korean J Radiol.* 2012; 13:564–571. [PubMed: 22977323]
24. Chan HP, Hadjiiski L, Zhou C, et al. Computer-aided diagnosis of lung cancer and pulmonary embolism in computed tomography—a review. *Acad Radiol.* 2008; 15:535–555. [PubMed: 18423310]
25. Eadie LH, Taylor P, Gibson AP. A systematic review of computer-assisted diagnosis in diagnostic cancer imaging. *Eur J Radiol.* 2012; 81:e70–e76. [PubMed: 21345631]
26. Freer TW, Ulissey MJ. Screening mammography with computer-aided detection: prospective study of 12,860 patients in a community breast center. *Radiology.* 2001; 220:781–786. [PubMed: 11526282]
27. Warren Burhenne LJ, Wood SA, D'Orsi CJ, et al. Potential contribution of computer aided detection to the sensitivity of screening mammography. *Radiology.* 2000; 215:554–562. [PubMed: 10796939]
28. Phillips M, Cataneo RN, Cummin AR, et al. Detection of lung cancer with volatile markers in the breath. *Chest.* 2003; 123:2115–2123. [PubMed: 12796197]
29. Yu JK, Zheng S, Tang Y, et al. An integrated approach utilizing proteomics and bioinformatics to detect ovarian cancer. *J Zhejiang Univ Sci B.* 2005; 6:227–231. [PubMed: 15754417]

Figure 1. Past and possible future evolution of computers in diagnosis. CAdE, computer-aided detection; CAdx, computer-aided diagnosis; Time, chronological time, assuming continuous technical advancements; UCAD, unsupervised CAdx;

Figure 2. Computational phases of CADe/CADx. CADe, computer-aided detection; CADx, computer-aided diagnosis.

Figure 3.
Study flow diagram.

TABLE 1

Description of Lung-CADx Studies

Study/Year	Population	Type	Gold Standard	QS	#R	#SCANS	Mode/ST	ALG
1 (10)/2003	CA/cont	UCAD	path	8	N/A	393	LDCT/10 mm	LDA
2 (11)/2004	CA only	UCAD	path	12	N/A	106	LDCT/10 mm	MTANN
3 (12)/2005	CA/cont	OP	path	15	14	27	LDCT/10 mm	MTANN
4 (13)/2005	CA/cont	OP	path	15	16	56	HRCT	MTANN
5 (14)/2005	CA/cont	OP	path	16	8	28	LDCT/3 mm	DT
6 (15)/2005	CA/cont	UCAD	path	10	N/A	81	LDCT/3 mm	LDA
7 (16)/2005	CA/cont	UCAD	path	11	N/A	415	LDCT/10 mm	MTANN
8 (17)/2006	CA/cont	OP	path	18	9	48	CXR	LDA
9 (18)/2006	CA/cont	OP	path	17	10	33	HRCT	ANN
10 (19)/2007	CA/cont	OP	path	18	9	200	LDCT/8 mm	ANN
11 (20)/2009	CA only	UCAD	path	9	N/A	69	LDCT/10 mm	MTANN
12 (21)/2010	CA/cont	OP	path	16	11	60	LDCT/10 mm	LDA
13 (22)/2010	CA/cont	OP	path + hx	15	6	152	LDCT/2 mm	LDA
14 (23)/2012	CA/cont	OP	path + CT	15	10	200	CXR	LDA

ALG, algorithm; ANN, artificial neural network; CA/cont, cancer control; CADx, computer-aided diagnosis; CXR, chest radiography; DT, Decision Tree; LDA, linear discriminant analysis; LDCT, low-dose computed tomography; HRCT, high-resolution CT; MT, massive training; MTANN, massive training artificial neural network; OP, observer-performance; QS, Quality Assessment of Diagnostic Accuracy Studies Score; # R, number of raters; #SCANS, number of scans total in the study; ST, slice thickness; UCAD, unsupervised computer diagnosis; PATH, pathology review of biopsy; hx, history; N/A, not applicable.

TABLE 2

Lung-CADx Accuracy Measures

Study	Sensitivity/CADx Alone	FP/Scan	ACC	A _Z Human	A _Z Machine	A _Z Both	P Value
1 (10)	.84	1.0	NR	N/A	.79	N/A	N/A
2 (11)	.83	5.8	NR	N/A	NR	N/A	N/A
3 (12)	.87	3.0	NR	.763	NR	.854	.002
4 (13)	.90	6.5	NR	.785	.831	.853	.016
5 (14)	.91 (.67/sp)	NR	81%	.68	NR	.81	.020
6 (15)	NR	NR	NR	N/A	.92	N/A	N/A
7 (16)	1.00 (.48sp)	NR	NR	N/A	.882	N/A	N/A
8 (17)	.81 (.70sp)	1.2	NR	.724	NR	.778	.008
9 (18)	.72	NR	76%	.910	.795	.944	.190
10 (19)	.93	NR	93%	.85	NR	.94	.014
11 (20)	.84	.5	NR	N/A	NR	N/A	N/A
12 (21)	NR	NR	NR	.864	NR	.924	.010
13 (22)	NR	NR	NR	.833	NR	.853	.010
14 (23)	.87	1.9	NR	N/A	NR	N/A	N/A

ACC, machine accuracy calculation (TP + TN)/(P + N), true-positive, true-negative, total-positive, total-negative; A_Z, receiver operating characteristic-area index; CADx, computer-aided diagnosis; FP, false-positive; NR, not reported; sp, specificity; , difference.