CHEST

Katharina Marten
Andreas Grillhösl
Tobias Seyfarth
Silvia Obenauer
Ernst J. Rummeny
Christoph Engelke

Computer-assisted detection of pulmonary nodules: evaluation of diagnostic performance using an expert knowledge-based detection system with variable reconstruction slice thickness settings

K. Marten (✉) · A. Grillhösl ·
T. Seyfarth · E. J. Rummeny ·
C. Engelke
Department of Radiology, Klinikum
rechts der Isar, Technical University
Munich,
Ismaningerstrasse 22,
81675 Munich, Germany
e-mail: katharina.marten@roe.med.
tu-muenchen.de
Tel.: +49-89-41402621
Fax: +49-89-41404834

S. Obenauer
Department of Radiology, Georg-
August University,
Robert-Koch-Strasse 40,
37075 Göttingen, Germany

Abstract The purpose of this study was to evaluate the performance of a computer-assisted diagnostic (CAD) tool using various reconstruction slice thicknesses (RST). Image data of 20 patients undergoing multislice CT for pulmonary metastasis were reconstructed at 4.0, 2.0 and 0.75 mm RST and assessed by two blinded radiologists (R1 and R2) and CAD. Data were compared against an independent reference standard. Nodule subgroups (diameter >10, 4–10, <4 mm) were assessed separately. Statistical methods were the ROC analysis and Mann-Whitney U test. CAD was outperformed by readers at 4.0 mm (Az = 0.18, 0.62 and 0.69 for CAD, R1 and R2, respectively; $P<0.05$), comparable at 2.0 mm (Az = 0.57, 0.70 and 0.69 for CAD, R1 and R2, respectively), and superior using 0.75 mm RST (Az = 0.80, 0.70 and 0.70 and sensitivity = 0.74, 0.53 and 0.53 for CAD, R1 and R2, respectively; $P<0.05$). Reader performances were significantly enhanced by CAD (Az = 0.93 and 0.95 for R1 + CAD and R2 + CAD, respectively, $P<0.05$). The CAD advantage was best for nodules <10 mm (detection rates = 93.3, 89.9, 47.9 and 47.9% for R1 + CAD, R2 + CAD, R1 and R2, respectively). CAD using 0.75 mm RST outperformed radiologists in nodules below 10 mm in diameter and should be used to replace a second radiologist. CAD is not recommended for 4.0 mm RST.

Keywords Lung neoplasms · Computers · Computed tomography

Introduction

Lung cancer represents the leading cause of cancer death in western countries, with a mean 5-year survival rate for all stages of 14%, which despite recent advances in diagnosis and treatment strategies has not changed significantly in the past 3 decades [1]. The prognosis of stage I cancer is more favorable, with 5-year mean survival rates of about 49%. However, only about 15% of lung cancers are diagnosed at this early stage [1]. CT is the most sensitive imaging modality for detection of early lung cancer [2–7]. The introduction of multislice CT (MSCT) scanners with short gantry revolution times and simultaneous acquisition of up to 16 0.625–0.75-mm slices has allowed for coverage of the entire chest by sub-millimeter isotropic volume acquisition within a single breathhold, enabling for detection of increasing quantities of small lung nodules. However, one MSCT examination of the chest delivers up to 750 axial images, which generate difficulties for chest radiologists, resulting in increasing false-negative rates for detection of small lesions [8]. Although thin-section CT allows for depiction of nodules as small as 1 mm [8], the detection rate of these lesions is generally low. In their study investigating nodules of 1–7 mm in size, Naidich and co-workers identified only 63% of all nodules, and in diameter ranges below 3.0 and 1.5 mm,

detection rates fell to 48 and 1%, respectively [9]. Particularly with ongoing MSCT lung cancer screening trials documenting large numbers of small pulmonary nodules —lesions below 4-mm diameter account for some 40% of all noncalcified lung nodules—the demand for automated assistance in the detection and work-up of nodular lesions becomes evident [10–14]. The purpose of this study is to evaluate the performance of a knowledge-based computer-assisted diagnosis (CAD) system in patients with multiple pulmonary nodules due to metastatic disease and to test the influence of various reconstruction slice thicknesses (RST) settings on the CAD performance in comparison and in combination with two experienced chest radiologists. Therefore, it tested the potential of CAD to improve nodule detection of radiologists using common RST settings of MSCT of the chest.

Materials and methods

Study design

This is a prospective observational study comparing the utility of an interactive CAD system prototype (ICAD, Siemens Medical Solutions, Erlangen, Germany) for multislice-CT-based detection of solid pulmonary nodules, with the performance of two experienced chest radiologists using various CT image RST. Multislice-CT (MSCT) scan images of patients being routinely evaluated for pulmonary metastatic disease were prospectively reconstructed using three slice thickness settings and evaluated by CAD and two experienced chest radiologists for solid pulmonary nodules. The results were compared against an independent reference standard. The resulting study data were not utilized for clinical decision making, and at our center no institutional review board approval is required for this type of study.

Study population, CT scans and their evaluation by radiologists

Multislice CT scans of the chest were obtained from 20 consecutive patients (mean age: 62.4 years; range: 29–84 years) at our center. All scans were performed using a 16-row MSCT unit (Sensation 16, Siemens Medical Solutions, Erlangen, Germany). The scan parameters used were tube voltage 120 kV, tube current 80 mAs at 0.5 s gantry revolution, 0.75-mm slice collimation and table feed of 19.2 mm/rot. Routinely, single phase peripheral intravenous power injection was performed using 80 ml nonionic contrast material of 300 mg/ml iodine concentration (Ultravist 300, Schering AG, Berlin, Germany) and subsequently 30 ml of normal saline solution at a flow of 2–3 ml/s. Image data were reconstructed using a lung filter kernel (B60f) at slice thickness settings of 4.0, 2.0 and

0.75 mm at 2.7-, 1.2- and 0.6-mm reconstruction increments (RI), respectively. The RST/RI settings were chosen to obtain image data comparable with standard single slice chest CT (RST = 4 mm; RI = 2.7 mm), standard chest MSCT with a limited numbers of resulting images (RST =2 mm; RI =1.2 mm) and MSCT with high isotropic resolution (RST =0.75 mm; RI =0.6 mm). Image data were stored and analyzed on a dedicated workstation not accessible for clinical routine. All CT image data were independently evaluated by two experienced chest radiologists who were blinded to the patient CT study and reconstruction protocol data. CT datasets were assessed using the interactive axial cine mode. Reader confidence for diagnosis of each pulmonary nodule was assessed and documented on a three-point scale (0= negative, 1= uncertain, 2= positive). The readers documented nodule position, diameter and solidity (solid, part solid and ground glass, as well as calcification). All image data were re-evaluated by the two radiologists in consensus with the CAD results.

CAD tool and CT evaluation by CAD

The CAD system (Siemens Medical Solutions, Erlangen, Germany; Siemens Corporate Research, Princeton, NJ) is a knowledge-based automatic lung nodule detection prototype comprising multiple "expert processing modules" to detect nodules with or without pleural or vascular attachment. Solitary nodules are recognized by an adaptive local histogram-based threshold-aided segmentation algorithm. After the structure is classified within a volume of interest (VOI), the chest wall is recognized and automatically removed if included in the VOI. After this, the potential nodule is extracted using a 3D-template method. Principle component analysis and classification tools ensure the segmentation of nodules in the presence of vessel attachment and the exclusion of structures recognized as non-nodules. An automatic bronchus recognition system allows for reduction of false-positive ratings simulated by bronchial wall thickening. The CAD segmentation algorithm has been described previously [15]. The CAD system offers a percentual confidence level for diagnosis of each lung nodule that was used for subsequent statistical analysis. Additionally, each nodule diameter is displayed on the output screen (Fig. 1). CAD image data evaluation was performed after evaluation by the two study readers.

Standard of reference and nodule inclusion criteria

The reference standard was established by consensus of two independent experienced chest radiologists who evaluated all datasets (0.75-, 2.0-, 4.0-mm datasets, Fig. 2) in consensus with knowledge of the results of the study

Fig. 1 Surface of CAD showing a results section containing position data, confidence level score and size measurements. The full axial slice display shows recognized structures marked by *arrows* (here nodule confirmed by the reference standard in right posterior upper lobe, *arrow*). The 3D display at the *top left* shows in this case a nodule with vascular attachment
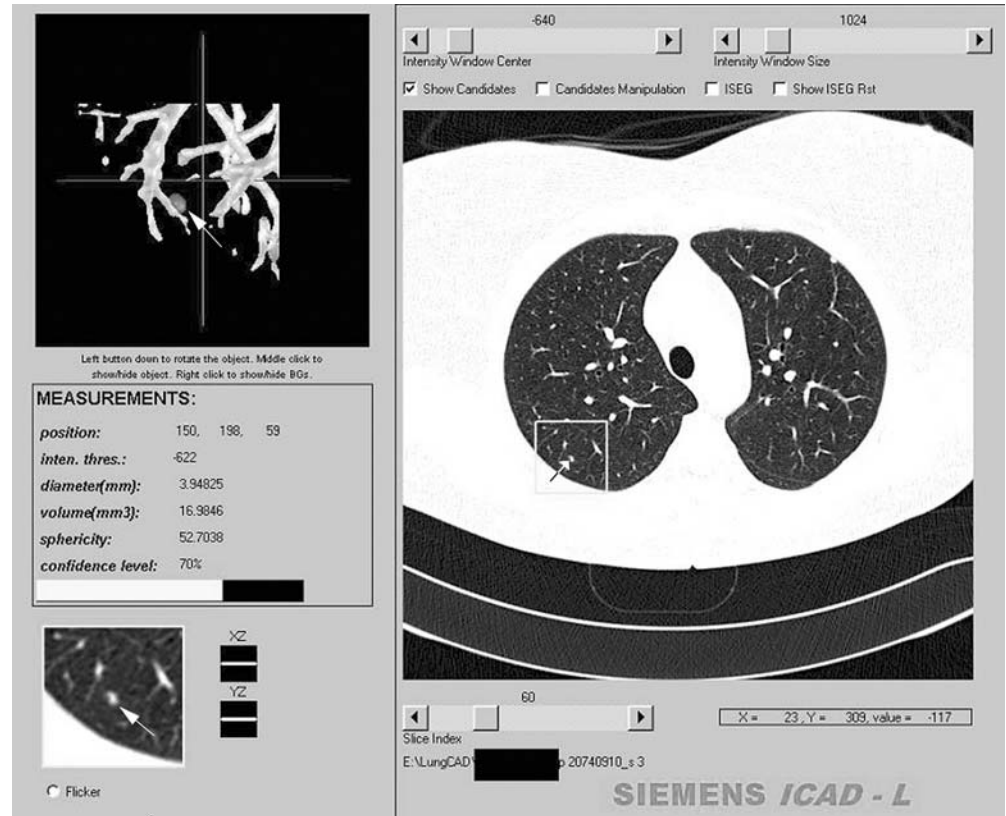
radiologists and CAD. The reference standard containing all nodules found on 0.75-mm slices was chosen for comparison of performance of CAD at different slice thicknesses. For comparison of CAD with reader performances for each of the reconstruction protocols, a portion of the reference standard containing all nodules detectable on each individual RST (4.0, 2.0 and 0.75 mm) was used as ground truth reference. Hereafter, results pertaining to the total number of nodules contained in the reference standard are referred to as absolute detection rates or absolute performances, and results pertaining to the reduced reference standard of nodules detectable in one particular RST setting are referred to as relative detection rates or relative performance. The nodule diameters were adapted in consensus from CAD measurement data or in cases of failed CAD detection revaluated manually or adapted from the study reader data. For assessment of the potential clinical significance of CAD performance in patients with incidental lung nodule findings, all nodules in our study were grouped according to a malignancy-probability and a resulting diameter-related clinical actionability equivalent (high actionability equivalent: nodule diameter ≥10.0 mm, intermediate actionability equivalent: nodule diameter of 4–9.9 mm and low actionability equivalent: nodule diameter <4 mm). Criteria for the diagnosis of a pulmonary nodule were the presence of a well-demarcated, solid, part-solid or ground glass spherical to irregular structure that, if solid, should have a density in the range of soft tissue or above or otherwise be well above the surrounding lung parenchyma (Fig. 2). Linear and subpleural densities that were attributable to pleural adhesions without nodular aspect were excluded from the reference standard. Bronchial wall thickening, thickening of interlobular septa and linear or reticular interstitial changes were excluded, as were areas of air space consolidation and regions with movement artifacts that would not allow for safe differentiation between pulmonary nodules and bronchial or vascular structures.
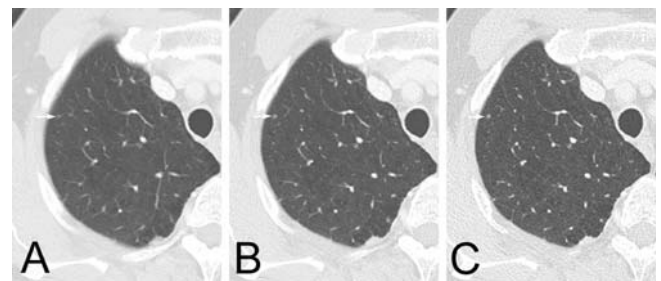

Fig. 2 A 2.5-mm diameter subpleural nodule of the right apical upper lobe segment in **A** 4.0-mm, **B** 2.0-mm and **C** 0.75-mm RST image data. On 4.0-mm RST the nodule was not recognized by the review panel and therefore not in the reference standard

Statistical methods

Statistical calculations were performed using a spreadsheet-based software package (StatsDirect release 2.3.4). Statistical evaluation of CAD and radiologist sensitivities and false-positive ratings per scan was performed by ROC analysis. StatsDirect performs the calculation of the area under the ROC by a non-parametric method analogous to the Wilcoxon/Mann-Whitney test [16, 17]. Two-tailed confidence intervals were calculated using DeLong's variance estimate [18]. Interobserver agreement between readers and CAD and each reader were evaluated by the method described by Markus, Bland and co-workers [19]. Median diameter differences between the various detected and nondetected nodule groups were assessed by Mann-Whitney U test. All tests were performed two-tailed with P values less than 0.05, indicating statistical significance.

Results

Interobserver agreement and nodule detection rates

A total of 135 nodules were included by the review panel in the reference standard. On 2.0- and 4.0-mm RST image data 129 and 59 nodules, respectively, were considered identifiable. Minor breathing or other pulmonary movement artifacts were recorded by the study readers in 4 of 20 patients and small areas of atelectasis or consolidation were recorded in three patients each. Emphysema or other pathologic diffuse low pulmonary attenuation was not recorded. In no case was the scan quality judged to be more substantially reduced than to allow for assessment of pulmonary nodules. The vast majority (97.0%) of the detected nodules was solid. A small minority of nodules (4.4%) was calcified. Therefore, testing for influence of nodule solidity on their detection by CAD and the readers was not performed. Reader confidence was generally good to excellent: the rates of indeterminate reader 1 and 2 ratings were 5.5 and 6.2%, 7.1 and 7.9% and 10.6 and 1.7% for 0.75-, 2.0- and 4.0-mm RST settings, respectively. Agreement between readers 1 and 2 was very good with all RST settings (k=0.89, 0.88 and 0.91 for 0.75-, 2.0- and 4.0-mm slice thickness settings, respectively). However, agreement between the readers and CAD was generally poor (k=0.39–0.40, 0.36–0.38 and 0.15–0.36 for 0.75-, 2- and 4-mm RST settings, respectively, Table 1). CAD had a lower lung nodule detection rate than the readers in 4.0-mm RST image data (absolute detection rates for CAD, readers 1 and 2 d=20.0, 35.5 and 38.5%, respectively), and with the use of 2.0- and 0.75-mm RST settings, it detected more nodules than the readers (absolute detection rates for CAD, readers 1 and 2 d=65.9, 47.4 and 47.4% for 2.0 mm, and 76.3, 52.3 and 52.3% for 0.75 mm RST settings, respectively, Table 2). Using 0.75-mm slices, CAD detected 45.1% more nodules than the

Table 1 Pairwise interobserver agreement

Observers	RST (mm)	k
Reader 1–2	4.0	0.905
Reader 1–CAD	4.0	0.156
Reader 2–CAD	4.0	0.357
Reader 1–2	2.0	0.882
Reader 1–CAD	2.0	0.359
Reader 2–CAD	2.0	0.379
Reader 1–2	0.75	0.887
Reader 1–CAD	0.75	0.409
Reader 2–CAD	0.75	0.398

k, pairwise interobserver agreement for positive observations; CI, confidence interval; RST, reconstruction slice thickness.

readers and with support of CAD readers 1 and 2 detected again 22.3 and 18.4% more nodules compared to CAD alone. With 2.0-mm RST CAD detected 39.1% more nodules than readers 1 and 2 and with support of CAD readers 1 and 2 detected again 28.1 and 34.8% more nodules than CAD, respectively. On 4.0-mm RST image data sets CAD detected 43.8 and 48.1% less nodules than readers 1 and 2. With support of CAD, readers 1 and 2 improved their nodule detection rates only by an additional 10.2 and 7.7%, respectively (Table 2).

Influence of reconstruction slice thickness on detected nodule size by CAD and readers

Using 0.75-mm-thick reconstructed slices, no significant difference in size between nodules detected or missed by CAD could be demonstrated (diameters of detected nodules =5.28±3.14 mm; diameters of non-detected nodules =4.3±1.7 mm, P=0.373), whereas nodules detected by the readers were significantly larger than their missed nodules (diameters of detected nodules =5.5±3.0 mm; diameters of non-detected nodules =3.6±1.6 mm; P<0.0001, Table 3). In addition, nodules missed by the readers were significantly smaller than nodules missed by CAD (P=0.007). Similarly, significant size differences were found with the use of 2.0-mm RST datasets (Table 3). With 4.0-mm RST nodules detected by CAD were larger than missed nodules (diameters of detected nodules =7.21±2.74 mm; diameters of non-detected nodules =5.53±3.03 mm, P=0.028). The same was observed for nodules detected by the readers (diameters of detected nodules =6.97±2.99 mm; diameters of non-detected nodules =4.82±1.61 mm; P=0.002). There was no significant size difference between nodules missed by CAD and those missed by the readers (P=0.786, Table 3). The detection rates of nodules with low and intermediate clinical diameter-related actionability equivalent were higher with CAD than with the readers and rose substantially with smaller RST settings (Table 4): the detection rates for nodules of these size ranges were best

Table 2 Nodule detection by CAD and readers

Observer	RST (mm)	n in all scans	n detectable at spec. RST	n detected	Total n not detected	FP
CAD	4.0	135	59	27	108	84
CAD	2.0	135	129	89	46	18
CAD	0.75	135	135	103	32	11
Reader 1	4.0	135	59	48	87	0
Reader 1	2.0	135	129	64	71	2
Reader 1	0.75	135	135	71	64	5
Reader 2	4.0	135	59	52	83	0
Reader 2	2.0	135	129	64	71	5
Reader 2	0.75	135	135	71	64	3
Reader 1 + CAD	4.0	135	59	54	81	0
Reader 1 + CAD	2.0	135	129	114	21	3
Reader 1 + CAD	0.75	135	135	126	9	2
Reader 2 + CAD	4.0	135	59	56	79	0
Reader 2 + CAD	2.0	135	129	120	15	3
Reader 2 + CAD	0.75	135	135	122	13	1

spec., specific; RST, reconstruction slice thickness; FP, false-positive.

with support of readers 1 and 2 by CAD using 0.7 mm RST (absolute detection rates for high, intermediate and low diameter-related actionability equivalent =100, 98.3–100 and 82.0–86.9%, respectively).

Influence of reconstruction slice thickness on absolute CAD and reader performances and comparison of relative observer performances at various reconstruction slice thickness settings

Absolute CAD performance pertaining to 135 nodules of the reference standard was best using 0.75-mm reconstructed sections (Az =0.8, 95% CI =0.71–0.89), with 73.9% sensitivity (95% CI =0.66–0.81) and 0.79 false-positive ratings per scan (Table 5). Due to a fall of the total

nodule detection rate from 76.2 to 66.0% and an increasing total false-positive rate of 1.05 per scan, the CAD performance deteriorated significantly using 2.0-mm sections (Az =0.57, 95% CI =0.46–0.69), with a sensitivity of 59% (95% CI =0.5–0.67) and 1.1 false-positive ratings per scan. With a massive decrease of the absolute nodule detection rate to 20% and a growing total false-positive rate, performance declined again significantly using 4.0-mm sections (Az =0.18, 95% CI =0.13–0.23). Compared to 2.0-mm sections, CAD sensitivity was significantly reduced to 4.4% (95% CI =0.02–0.09) with 4.5 false-positive ratings per scan (Table 5). However, the changes of absolute reader performance pertaining to 135 nodules of the reference standard were dissimilar: although the absolute performances of both readers were slightly improved by diminution of the RST values from 4.0 through

Table 3 Descriptive statistics of detected and undetected nodules

Observer	RST (mm)	Detection	Group index	n missed/detected	Min (mm)	Max (mm)	MD (mm)	SD (mm)	SEM (mm)	Comp.	p	Comp.	p
CAD	4.0	Yes	1	27	2.9	14	7.214	2.74	0.457	1, 2	0.028	2, 8	0.786
CAD	4.0	No	2	32	1.6	11.5	5.526	3.032	0.696				
CAD	2.0	Yes	3	89	2	14	4.858	2.572	0.263	3, 4	0.679	4, 10	0.01
CAD	2.0	No	4	30	1	10	5.061	2.542	0.457				
CAD	0.75	Yes	5	103	1.8	14	5.281	3.144	0.316	5, 6	0.373	6, 12	0.007
CAD	0.75	No	6	32	1	8	4.345	1.714	0.308				
1 and 2	4.0	Yes	7	100	1.6	14	6.969	2.994	0.312	7, 8	0.002	1, 3	<0.0001
1 and 2	4.0	No	8	18	2	7.8	4.815	1.612	0.361				
1 and 2	2.0	Yes	9	128	1	14	6.213	2.889	0.259	9, 10	<.0001	1, 5	0.0001
1 and 2	2.0	No	10	130	2	7.8	3.663	1.267	0.111				
1 and 2	0.75	Yes	11	142	1	14	5.506	3.04	0.216	11, 12	<.0001	3, 5	0.858
1 and 2	0.75	No	12	128	1.8	9.1	3.592	1.578	0.142				

RST, reconstruction slice thickness; Min., minimum diameter; Max., maximum diameter; MD, mean diameter; SD, standard deviation; SEM, standard error of mean; comp., group comparison; P, P value of Mann-Whitney U test.

Table 4 Size-related actionability equivalent of detected nodules

Observer	D (mm)	Actionability equivalent	RS	0.75 mm	RS (2 mm)	2.0 mm	RS (4 mm)	4.0 mm
CAD	>9.9	High	15	12	12	6	12	6
CAD	4–9.9	Intermediate	58	43	61	39	39	18
CAD	<4	Low	61	47	56	44	8	3
Reader 1	>9.9	High	15	14	12	12	12	12
Reader 1	4–9.9	Intermediate	58	40	61	41	39	28
Reader 1	<4	Low	61	17	56	11	8	7
Reader 2	>9.9	High	15	14	12	12	12	12
Reader 2	4–9.9	Intermediate	58	40	61	40	39	32
Reader 2	<4	Low	61	17	56	12	8	7
Reader 1 + CAD	>9.9	High	15	15	12	12	12	12
Reader 1 + CAD	4–9.9	Intermediate	58	58	61	53	39	34
Reader 1 + CAD	<4	Low	61	53	56	49	8	7
Reader 2 + CAD	>9.9	High	15	15	12	12	12	12
Reader 2 + CAD	4–9.9	Intermediate	58	57	61	58	39	35
Reader 2 + CAD	<4	Low	61	50	56	50	8	8

D, diameter range; RS, reference standard; 0.75, 2.0, 4.0 mm, RST settings.

2.0 to 0.75 mm, the influence on the Az values was not statistically significant (0.62–0.70 and 0.69–0.70 for readers 1 and 2, respectively, $P>0.05$). However, the false-positive rates per scan were increased (from 0 to 2.6 and 0 to 2.1 for readers 1 and 2, respectively) with decreasing RST values (Table 5).

The analyses of relative CAD performances in comparison to the two radiologists on 4.0-, 2.0- and 0.75-mm-thick slices pertaining to the nodules documented on the images at each RST setting are listed in Table 6. In 0.75-mm reconstructed sections, CAD outperformed the two readers (Az=0.802, 95% CI=0.71–0.89 for CAD vs. Az=0.64 and 0.69, 95% CI=0.52–0.76 and 0.58–0.81 for readers 1 and 2, respectively). Sensitivity of CAD was significantly higher than reader sensitivities (CAD sensitivity=73.9%, 95% CI=0.66–0.81; vs. both readers equally sensitivity=53%, 95% CI=0.44–0.62). False-positive CAD counts per scan were higher than those of the readers (CAD FP/scan=0.79 vs. 0.26 and 0.21 for readers 1 and 2, respectively). Using 2.0-mm-thick reconstructed slices, no statistically significant differences were found between the relative CAD and reader performances, although the readers performed slightly superiorly to CAD (CAD Az=0.59, 95% CI=0.47–0.69, and Az=0.70, 95% CI=0.60–0.80 equally for readers 1 and 2). CAD sensitivity at 2.0-mm RST was slightly better than that of the readers (CAD sensitivity=60.3%, 95% CI=0.51–0.69 and sensitivities=48.9% and 51.2%, 95% CI=0.40–0.58 and 0.42–0.60 for readers 1 and 2, respectively), which

Table 5 ROC analysis of absolute CAD and reader performances

Observer	RST (mm)	n analyzed	Az	Cut-off point	95% CI	Sensitivity	95% CI	FP/scan
CAD	4.0	135	0.181	80.2	0.128–0.234	0.044	0.016–0.094	4.474
CAD	2.0	135	0.574	70.2	0.462–0.687	0.59	0.501–0.674	1.053
CAD	0.75	135	0.802	60.2	0.711–0.894	0.739	0.656–0.811	0.789
Reader 1	4.0	135	0.622	0.004	0.494–0.749	0.407	0.324–0.495	0
Reader 1	2.0	135	0.695	1.004	0.597–0.793	0.478	0.391–0.566	0.105
Reader 1	0.75	135	0.704	1.004	0.640–0.768	0.53	0.442–0.617	0.263
Reader 2	4.0	135	0.685	0.004	0.569–0.802	0.393	0.310–0.480	0
Reader 2	2.0	135	0.692	1.004	0.594–0.791	0.5	0.413–0.588	0.158
Reader 2	0.75	135	0.696	1.004	0.632–0.761	0.53	0.442–0.617	0.211
Reader 1 + CAD	4.0	135	0.692	2	0.652–0.734	0.382	0.302–0.467	0
Reader 1 + CAD	2.0	135	0.903	2	0.844–0.963	0.881	0.813–0.938	0.158
Reader 1 + CAD	0.75	135	0.928	2	0.873–0.983	0.925	0.867–0.964	0.105
Reader 2 + CAD	4.0	135	0.720	2	0.678–0.762	0.418	0.350–0.501	0
Reader 2 + CAD	2.0	135	0.938	2	0.888–0.988	0.896	0.835–0.938	0.158
Reader 2 + CAD	0.75	135	0.950	2	0.907–0.993	0.911	0.849–0.953	0.053

RST, reconstructed slice thickness; Az, Wilcoxon estimate of area under ROC curve; FP, false-positive.

Table 6 ROC analysis of relative CAD and reader performances

Observer	RST (mm)	n analyzed	Az	Cut-off point	95% CI	Sensitivity	95% CI	FP/scan
CAD	4	59	0.597	69.0	0.506–0.688	0.600	0.465–0.724	4.474
Reader 1	4	59	0.958	0.004	0.916–1	0.917	0.816–0.972	0
Reader 2	4	59	0.942	0.004	0.894–0.989	0.883	0.774–0.952	0
CAD	2	129	0.585	70.2	0.473–0.697	0.603	0.514–0.687	1.053
Reader 1	2	129	0.702	1.004	0.604–0.799	0.489	0.400–0.577	0.105
Reader 2	2	129	0.700	1.004	0.602–0.798	0.512	0.423–0.600	0.158
CAD	0.75	135	0.802	61.0	0.711–0.894	0.739	0.656–0.811	0.789
Reader 1	0.75	135	0.637	1.004	0.516–0.757	0.530	0.442–0.617	0.263

Az, Wilcoxon estimate of area under ROC curve; RST, reconstructed slice thickness; FP, false-positive. ROC results are comparable only within the same RST setting.

was outweighed by the increasing false-positive CAD rates per scan (FP/scan=1.1 for CAD and 0.11 and 0.16 for readers 1 and 2, respectively). Using 4.0-mm reconstructed sections, the readers performed significantly superiorly to the CAD system, which was largely attributable to high false-positive and false-negative rates of CAD (Az of CAD=0.60, 95% CI=0.51–0.69 and Az=0.96 and 0.94, 95% CI=0.92–1 and 0.89–0.99 for readers 1 and 2, respectively). As a result, sensitivities and false-positive ratings per scan were significantly better with the readers than with CAD (sensitivity of CAD=0.60, 95% CI=0.47–0.72 and sensitivities=0.917 and 88.3%, 95% CI=0.82–0.97 and 0.77–0.95 for readers 1 and 2, respectively; FP/scan=4.47 for CAD and 0 for readers 1 and 2, respectively).

Influence of reconstruction slice thickness on absolute reader performance supported by CAD

The results of a consensus reading of our study radiologists with support of CAD are given in Tables 2, 4, 5. By increasing nodule detection rates with simultaneous reduction of false-positive rates in comparison to independent CAD and readers on 0.75-mm RST data (detection rates=93.3, 90.4, 76.3, 52.6 and 52.6% for CAD + reader 1, CAD + reader 2, CAD, reader 1 and reader 2, respectively; false-positive rate/scan=0.11, 0.05, 0.79, 0.26 and 0.21 for CAD + reader 1, CAD + reader 2, CAD, reader 1 and reader 2, respectively), the performances of readers 1 and 2 could be substantially or significantly improved by support of CAD over the individual reader or independent CAD values (Az=0.93, 0.95, 0.80, 0.70 and 0.70 for CAD + reader 1, CAD + reader 2, CAD, reader 1 and reader 2, respectively). Supported by CAD, readers 1 and 2 identified 312.8 and 294.1% more pulmonary nodules of low diameter-related actionability equivalent (nodule diameter <4 mm) than without CAD, 45.0 and 42.5% additional nodules of intermediate diameter-related actionability equivalent, respectively (nodule diameter 4.0–9.9 mm), and a total of 94.7 and 87.7% additional nodules of low or intermediate diameter-related actionability equivalent (nodule diameter <10 mm, Table 4). Similar results were obtained in 2.0-mm slices, where the absolute performance

of the readers with CAD support were significantly superior to independent CAD and reader results (Table 5). Using 4.0-mm RST, the readers' absolute performances could be only marginally improved without the advantage reaching statistical significance.

Discussion

Computer-aided detection has been proposed as a management solution to ever-increasing amounts of clinical MSCT data. In the chest, CAD has been employed for detection of pulmonary embolism, for quantification of interstitial lung disease and for automatic measurement of nodule volume [20–24]. Due to the inherently high contrast between soft tissues and lung tissues, it represents a particularly promising tool for optimized automated detection of pulmonary nodules from MSCT datasets. Several CAD approaches are currently undergoing clinical evaluation with preliminary evidence that CAD may be suited to guide the radiologist to suspicious lesions [20, 25–31]. Mathematical models for computer-aided detection of pulmonary nodules can be broadly divided into two categories: density-based approaches using the high density interval between the nodule and the pulmonary parenchyma employ techniques such as multiple thresholding [7, 25, 26], region-growing [20], locally adaptive thresholding in combination with region-growing [27] and fuzzy clustering [28] for nodule identification. A priori knowledge of small nodules is applied to reduce false-positive results from the detected nodule candidates [8]. Geometric model-based systems, by contrast, use the relatively compact nodule shape in order to create the models for nodule identification. For this purpose, approaches such as "N-Quoit filter" [29], template-matching [30], object-based deformation [31] and the anatomy-based generic model [32] have been developed. In addition analysis of curved surface morphology [33, 34] and subtraction of bronchovascular structures from the chest CT images [35] have been used for nodule detection.

Current literature data on CAD performance is based on small numbers of nodules investigated with CAD prototype systems, and the broad variety of technological ap-

proaches impedes the comparison of the detection performances based on the above-mentioned algorithms. Few studies have been performed employing CAD in more than 50 nodules. Zhao et al. evaluated a CAD tool based on multiple thresholding in combination with feature extraction and classification [8]. The authors investigated 266 simulated nodules 2–7 mm in size that were placed onto eight normal chest CT scans and found a sensitivity of 84.2% with five false-positive findings per scan. Wormanns et al. used a CAD method based on region-growing designed for automated detection of >5-mm-sized nodules [34]. However, in this study, sensitivity of CAD was only 38%. Armato et al. applied a multiple thresholding-based CAD tool in 187 pulmonary nodules (3.1–27.8 mm in size) and found a sensitivity of 70% with three false-positive findings per section [36]. In another study investigating 50 nodules, 38 of them representing missed lung cancers, overall sensitivity was 80% with an average of one false-positive finding per section [37]. Lung cancers were even detected with 84% sensitivity, but still a false-positive rate of 1.0 per section. However, it is problematic to compare the performance of different published CAD systems from literature data, because different algorithms for automated detection and, most importantly, no standardized nodule database were used. Moreover, although early data on the performance of CAD systems appear promising, a major issue, the influence of RST on CAD performance with a view to replacement of clinical readers, remains unaddressed, to the best of our knowledge.

We evaluated the prototype of a knowledge-based interactive computer-aided detection system for automatic detection of pulmonary nodules in MSCT image data. The system was designed to detect three different kinds of solid pulmonary nodules: solitary nodules, nodules attached to the chest wall and nodules attached to vessels. By determining the influence of the RST on CAD and in parallel on reader performance, using all nodules identifiable on 0.75-mm slices as the standard of reference, we obtained a robust overall comparison of system performance. The best independent CAD absolute performance was achieved using 0.75-mm sections, with sensitivity values and a false-positive rate per scan well in the range of results reported by other investigators. We could improve these results significantly by combination of CAD with single radiologists with excellent sensitivity and false-positive values, which in the light of our radiologists' very good interobserver agreement and their high confidence levels would suggest a role of CAD in combination with one reader for clinical assessment of pulmonary nodules. CAD as an independent modality and for support of our readers had a significantly superior performance with the use of 0.75-mm than with 2.0-mm thick sections, suggesting a benefit by restricting its application to thin-collimation chest MSCT. In addition, CAD outperformed our readers in the detection of nodules of low to intermediate diameter-related actionability equivalent using 0.75-mm RST. It had

comparable relative and absolute performances using 2.0-mm and significantly poorer relative and absolute performances than the radiologists using 4.0-mm RST. This deficit using 4.0-mm-thick slices could not be substantially abridged by consensus of one of our readers with CAD. Consequently, we recommend avoiding the application of CAD with image data of such high RST setting. In this study, we considered it valuable to include the 4.0-mm section thickness in order to demonstrate the effect of increasing slice thickness to a single-slice non-high resolution CT standard (or "rationalized" MSCT archive image quality standard) on reader and CAD performances.

As evidenced by the poor interobserver agreements between CAD and the readers and the significant improvement of their absolute performances in consensus, CAD detected different nodules (significantly higher numbers of smaller nodules) than the readers, which is in good agreement with another study documenting that a radiologist-based model for establishing a reference standard of pulmonary nodules in CT data missed a substantial number of pulmonary nodules when applied to thin-section MSCT data [38]. The best relative reader performance pertaining to RST-specific detectable nodules was achieved using 4.0-mm sections, which deteriorated significantly with decreasing slice thicknesses, the reason being the larger number of small lesions missed on thin sections and the moderate increase of false-positives, whereas CAD's relative performance against the readers improved when 0.75-mm sections were used. Interestingly, the CAD detection performance using thin RST settings was independent from nodule size: when 0.75- and 2.0-mm-thick slices were evaluated, there was no significant difference in the size of nodules detected or missed. However, when 4.0-mm-thick sections were evaluated by CAD, the detected nodules were significantly larger than the missed nodules, presumably due to a nodule recognition malfunction with increasing volume averaging. Conversely, as expected, the detected nodules were significantly larger than the undetected nodules with 0.75- and 2.0-mm-thick sections when evaluated by the two readers, suggesting a substantial role of combined reader and CAD application in pulmonary nodules below 4.0-mm diameter, also in nodules between 4.0- and 9.9-mm diameter.

Our study included patients with pulmonary metastasis, undergoing clinical staging and follow-up CT examinations of the chest. All scans were performed with "regular" radiation doses as indicated above, and no low-dose CT scans were performed. For this reason, it remains to be determined how well our study results can be translated into a lung cancer screening setting using unenhanced low-dose thin-slice MSCT. The study design was not tailored to investigate the influence of pulmonary abnormality or artifacts on CAD performance, which should be subject to further studies. In our collective, the vast majority of nodules was solid, and more work is required to

evaluate CAD in partially solid lesions for definition of its potential role in lung cancer screening collectives. It should be noted that our results pertain to experienced readers. The effect of reader experience on combined performance with CAD has been investigated previously [39].

In conclusion, we found an excellent performance of CAD in 0.75-mm-thick reconstructed sections, where it significantly outperformed experienced readers, in the detection of nodules of small and intermediate diameters up to 9.9 mm. We observed considerable rates of false-positive findings compared to human readers. This and the inability of a CAD system to assess etiology or clinical significance prohibit the stand-alone use of this technology as a reader of clinical data sets, which is in keeping with results of other ongoing studies [40, 41]. However, we recommend CAD as a replacement of a second radiologist for detection of some 300% of additional pulmonary nodules below 4-mm diameter and of some 90% additional nodules below 10 mm using 0.75-mm slices. The use of CAD with 4.0-mm RST image data should be avoided. If our results can be maintained with the use of low-dose MSCT in lung cancer screening trials, the use of CAD could double the detection rate of early stage cancers. In addition, it could aid in the quantitative assessment of treatment response of pulmonary metastatic disease.

References

1. Ries L (1999) SEER Cancer statistics review, 1973–1996. National Cancer Institution, Bethesda
2. Henschke CI, McCauley DI, Yankelevitz DF, Naidich DP, McGuinness G, Miettinen OS, Libby D, Pasmantier M, Koizumi J, Altorki N, Smith JP (2001) Early lung cancer action project: a summary of the findings on baseline screening. Oncologist 6:147–152
3. Kaneko M, Eguchi K, Ohmatsu H, Kakinuma R, Naruke T, Suemasu K, Moriyama N (1996) Peripheral lung cancer: screening and detection with low-dose spiral-CT versus radiography. Radiology 201:798–802
4. Sone S, Li F, Yang ZG, Takashima S, Maruyama Y, Hasegawa M, Wang JC, Kawakami S, Honda T (2000) Characteristics of small lung cancers invisible on conventional chest radiography and detected by population based screening using spiral CT. Br J Radiol 73:137–145
5. Yankelevitz DF, Gupta R, Zhao B, Henschke CI (1999) Small pulmonary nodules: evaluation with repeat CT—preliminary experience. Radiology 212:561–566
6. Yankelevitz DF, Reeves AP, Kostis WJ, Zhao B, Henschke CI (2000) Small pulmonary nodules: volumetrically determined growth rates based on CT evaluation. Radiology 217:251–256
7. Ko JP, Betke M (2001) Chest CT: automated nodule detection and assessment of change over time—preliminary experience. Radiology 218:267–273

8. Zhao B, Gamsu G, Ginsber MS, Jiang L, Schwartz LH (2003) Automatic detection of small lung nodules on CT utilizing a local density maximum algorithm. J Appl Clin Med Phys 4:248–260
9. Naidich DP, Rusinek H, McGuinness G, Leitman B, McCauley DI, Henschke CI (1993) Variables affecting pulmonary nodule detection with computed tomography: evaluation with three-dimensional computer simulation. J Thorac Imaging 8:291–299
10. Henschke CI, Yankelevitz DF, Libby D, Kimmel M (2002) CT screening for lung cancer: the first 10 years. Cancer J 8:S47–S54
11. Swensen SJ, Jett JR, Sloan JA, Midthun DE, Hartman TE, Sykes AM, Aughenbaugh GL, Zink FE, Hillman SL, Noetzel GR, Marks RS, Clayton AC, Pairolero PC (2002) Screening for lung cancer with low-dose spiral computed tomography. Am J Respir Crit Care Med 165:508–513
12. Diederich S, Wormanns D, Semik M, Thomas M, Lenzen H, Roos N, Heindel W (2002) Screening for early lung cancer with low-dose spiral CT: prevalence in 817 asymptomatic smokers. Radiology 222:773–781
13. Swensen SJ, Jett JR, Hartman TE, Midthun DE, Sloan JA, Sykes AM, Aughenbaugh GL, Clemens MA (2003) Lung cancer screening with CT: Mayo Clinic experience. Radiology 226:756–761
14. Friese SA, Rieber A, Fleiter T, Brambs HJ, Claussen CD (1994) Pulmonary nodules in spiral volumetric and single slice computed tomography. Eur J Radiol 18:48–51

15. Fan L, Novak CL, Naidich DP, Wei G, Quian J, Ko JP (2002) Improving optimal radiologic interpretation of low-dose multi-slice lung CT studies using ICAD. Radiology 225:475 (RSNA abstract)
16. Press WH, Flamery BP, Teukolsky SA, Vetterling WTne (1992) Numerical recipes, the art of scientific computing. Cambridge University Press, Cambridge
17. Hanley JA, McNeil BJ (1982) The meaning and use of area under a receiver operating characteristic (ROC) curve. Radiology 143:29–36
18. DeLong ER, DeLong DM, Clarke-Pearson DL (1988) Comparing the areas under two or more correlated receiver operating curves: a nonparametric approach. Biometrics 44:837–845
19. Markus H, Bland JM, Rose G, Sitzer M, Siebler M (1996) How good is intercenter agreement in the identification of embolic signals in carotid artery disease? Stroke 27:1249–1252
20. Fiebich M, Wietholt C, Renger BC, Armato S III, Hoffmann K, Wormanns D, Diederich S (1999) Automatic detection of pulmonary nodules in low-dose screening thoracic CT examinations. Proc SPIE 3661:1434–1439
21. van Ginneken B, ter Haar Romeny BM, Viergever MA (2001) Computer-aided diagnosis in chest radiography: a survey. IEEE Trans Med Imaging 20:1228–1241
22. Masutani Y, Macmahon H, Doi K (2002) Computerized detection of pulmonary embolism in spiral CT angiography based on volumetric image analysis. IEEE Trans Med Imaging 21:1517–1523

23. Uppaluri R, Hoffman EA, Sonka M, Hunninghake GW, McLennan G (1999) Interstitial lung disease: a quantitative study using the adaptive multiple feature method. Am J Respir Crit Care Med 159:519–525

24. Wormanns D, Kohl G, Klotz E, Marheine A, Beyer F, Heindel W, Diederich S (2004) Volumetric measurements of pulmonary nodules at multi-row detector CT: in vivo reproducibility. Eur Radiol 14:86–92

25. Giger ML, Bae KT, MacMahon H (1994) Computerized detection of pulmonary nodules in computed tomography images. Invest Radiol 29:459–465

26. Armato SG III, Giger ML, MacMahon H (2001) Automated detection of lung nodules in CT scans: preliminary results. Med Phys 28:1552–1561

27. Fan L, Novak CL, Qian J, Kohl G, Naidich DP (2001) Automatic detection of lung nodules from multi-slice low-dose CT images. Proc SPIE 4322:1828–1835

28. Satoh H, Ukai Y, Niki N, Eguchi K, Mori K, Ohmatsu H, Kakinuma R, Kaneko M, Moriyama N (1999) Computer aided diagnosis system for lung cancer based on retrospective helical CT images. Proc SPIE 3661:1324–1335

29. Okumura T, Miwa T, Kako J, Yamamoto S, Matsumoto M, Tateno Y, Linuma T, Matsumoto T (1998) Image processing for computer-aided diagnosis of lung cancer screening system by CT (LSCT). Proc SPIE 3338:1314–1322

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

31. Lou S, Chang C, Lin K, Chen T (1999) Object-based deformation technique for 3-D CT lung nodule detection. Proc SPIE 3661:1544–1552

32. Brown MS, McNitt-Gray MF, Goldin JG, Suh RD, Sayre JW, Aberle DR (2001) Patient-specific models for lung nodule detection and surveillance in CT images. IEEE Trans Med Imaging 20:1242–1250

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

34. Wormanns D, Fiebich M, Saidi M, Diederich S, Heindel W (2002) Automatic detection of pulmonary nodules at spiral CT: clinical application of a computer-aided diagnosis system. Eur Radiol 12:1052–1057

35. Croisille P, Souto M, Cova M, Wood S, Afework Y, Kuhlman J, Zerhouni E (1995) Pulmonary nodules: improved detection with vascular segmentation and extraction with spiral CT. Radiology 197:397–401

36. Armato SG III, Giger ML, Moran CJ, Blackburn JT, Doi K, MacMahon H (1999) Computerized detection of pulmonary nodules on CT scans. Radiographics 19:1303–1311

37. Armato SG III, Li F, Giger ML, MacMahon H, Sone S, Doi K (2002) Lung cancer: performance of automated lung nodule detection applied to cancers missed in a CT screening program. Radiology 225:685–692

38. Rubin G, Naidich D, Sherbondy A, Lyo J, Napel S (2003) Inadequacy of lung nodule reference standard based upon standard methods of expert consensus review using cine-paging of transverse thin-section MDCT lung scans. Radiology 229:441–442 (RSNA abstract)

39. Marten K, Seyfarth T, Auer F, Wiener E, Grillhösl A, Obenauer S, Rummeny EJ, Engelke C (2004) 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 14:1930–1938

40. Herzog P, Seyfarth T, Shen H, Schoenberg S, Reiser M (2003) Performance of a computer-aided diagnosis tool for the detection of pulmonary nodules at multidetector-row CT data. Radiology 229:292 (RSNA abstract)

41. Miller D, Burns R, Schneider A, O'Shaughnessy K, Wood S, Castellino R (2003) A multi-reader, multi-case ROC study demonstrating improved performance detecting lung nodules using computer-aided detection (CAD) in multi-slice CT (MSCT) of the chest. Radiology 229:291–292 (RSNA abstract)