Thoracic Cancer
Open Access

Thoracic Cancer ISSN 1759-7706

ORIGINAL ARTICLE

Evaluating the performance of a deep learning-based computer-aided diagnosis (DL-CAD) system for detecting and characterizing lung nodules: Comparison with the performance of double reading by radiologists

Li Li[1], Zhou Liu[1], Hua Huang[1], Meng Lin[2] & Dehong Luo[1,2]

1 Department of Radiology, National Cancer Center/Cancer Hospital and Shenzhen Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Shenzhen, China

2 Department of Radiology, National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

Keywords

Computer-aided diagnosis (CAD); deep learning based computer-aided diagnosis (DL-CAD); double reading; lung nodule screening; nodule characterization.

Correspondence

Dehong Luo, Department of Radiology, National Cancer Center/Cancer Hospital and Shenzhen Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, 113 Baohe Avenue, Longgang District, Shenzhen, Guangdong 518116, China.
Tel: +86 139 2623 6152
Fax: +86 139 2623 6152
Email: 13926236152@163.com

Li Li and Zhou Liu contributed equally to this work.

Received: 1 October 2018;
Accepted: 13 November 2018.

doi: 10.1111/1759-7714.12931

Thoracic Cancer **10** (2019) 183–192

Abstract

Background: The study was conducted to evaluate the performance of a state-of-the-art commercial deep learning-based computer-aided diagnosis (DL-CAD) system for detecting and characterizing pulmonary nodules.

Methods: Pulmonary nodules in 346 healthy subjects (male: female = 221:125, mean age 51 years) from a lung cancer screening program conducted from March to November 2017 were screened using a DL-CAD system and double reading independently, and their performance in nodule detection and characterization were evaluated. An expert panel combined the results of the DL-CAD system and double reading as the reference standard.

Results: The DL-CAD system showed a higher detection rate than double reading, regardless of nodule size (86.2% vs. 79.2%; $P < 0.001$): nodules ≥ 5 mm (96.5% vs. 88.0%; $P = 0.008$); nodules < 5 mm (84.3% vs. 77.5%; $P < 0.001$). However, the false positive rate (per computed tomography scan) of the DL-CAD system (1.53, 529/346) was considerably higher than that of double reading (0.13, 44/346; $P < 0.001$). Regarding nodule characterization, the sensitivity and specificity of the DL-CAD system for distinguishing solid nodules > 5 mm (90.3% and 100.0%, respectively) and ground-glass nodules (100.0% and 96.1%, respectively) were close to that of double reading, but dropped to 55.5% and 93%, respectively, when discriminating part solid nodules.

Conclusion: Our DL-CAD system detected significantly more nodules than double reading. In the future, false positive findings should be further reduced and characterization accuracy improved.

Introduction

According to a World Health Organization cancer report, lung cancer remains the most lethal cancer and leading cause of mortality globally, with an estimated 1.8 million new cases and 1.6 million deaths worldwide in 2012,[1] and an overall five-year survival rate of only 10–15%.[2] Early detection matters. Patients diagnosed with clinical stage I cancer are reported to have an overall 10-year survival rate of 88%, or 92% in patients that undergo surgical resection right after diagnosis.[3] Moreover, the largest National Lung Screening Trial (NLST), which included 53 454 subjects, showed that low-dose computed tomography (LDCT) detected 13% more instances of lung cancer and represented a 20% reduction in lung cancer-specific five-year mortality than radiography.[4] Thus, based on this result, annual screening is currently recommended for the elderly, especially for heavy smokers.

With high sensitivity for nodule detection, LDCT lung cancer screening normally produces large numbers of very thin sections, which makes interpretation of a whole lung CT so tedious and time consuming that radiologists are prone to overlook small nodules and make diagnostic errors.[5] A study reported that 20–35% of small lung nodules were missed in screening diagnosis by a single radiologist,[6] thus double reading would improve the detection rate but could still could miss quite a few potentially malignant nodules.[7,8] A computer-aided diagnosis (CAD) system was initially developed as early as the late 1980s aiming to assist radiologists to reduce missed nodules and make their work easier. Until now, a large variety of studies based on LDCT using different CAD systems and data have reported a wide range of sensitivity performance from 38%[9] to 100%[10] and false positive rates from 1.0 per scan[11] to 8.2 per scan.[12] With the increasing improvement of CAD systems, the majority of studies have demonstrated that CAD systems could detect more nodules than radiologists,[11,13–15] even after double reading.[16,17] Moreover, in comparison with most CAD systems based on supervised machine learning algorithms,[18,19] multiple studies have shown that deep learning-based CAD systems (DL-CAD) have superior detection rates and further reduce false positive rates.[20–22] However, CAD systems are far from perfect and thus require further development to be improved.

Another concern is whether a CAD system could accurately characterize different types of nodules. It is well known that a large portion of persistent subsolid nodules, which include part-solid nodules and ground-glass nodules (GGNs), correspond to lung adenocarcinomas of various stages of development, such as atypical adenomatous hyperplasia, in situ adenocarcinoma, or minimally invasive adenocarcinoma and invasive adenocarcinoma,[23,24] which highlights the importance of distinguishing different types of lung nodules automatically, especially in a large lung cancer screening program. If a CAD system could automatically and accurately characterize the different types of nodules, it would greatly reduce the workload of radiologists in lung cancer screening programs. However, occasionally even experienced thoracic radiologists have difficulty distinguishing solid, part solid, and GGNs among nodules measuring < 5 mm. Moreover, several large studies have shown that 95% of nodules measuring ≤ 10 mm are benign.[25–27] For very small nodules (< 5 mm with a corresponding volume of 65.4 mm^3), the chance of malignancy is negligible (<1%).[26,28,29] Because of these reasons, we evaluated the performance of a commercial DL-CAD system for characterizing nodules no smaller than 5 mm and for detecting nodules regardless of size.

The aim of our study was to test the performance of a state-of-the-art commercial DL-CAD system for lung nodule detection and characterization using LDCT images from our lung cancer screening program.

Methods

Study population

Based on NLST inclusion criteria,[4] LDCT scans were performed at our institution from March to November 2017 on healthy subjects aged > 50 years or younger subjects with a heavy smoking history, who were willing to participate in an annual check-up. After excluding subjects with significant morphological changes in lung CT images other than nodules, and those with quality-impaired CT images interfering interpretation because of insufficient inhalation, we enrolled 367 subjects with LDCT scans. The reason for our exclusion criteria was because our DL-CAD system was trained on high-quality CT images with nodules. The Institutional Review Board of our hospital approved this study and written informed consent was obtained from each subject.

Computed tomography protocol for image acquisition

All patients were scanned using a 64-row multi-detector CT (Optima CT660, GE Healthcare, Atlanta, GA, USA). None of the subjects were administered any intravenous contrast media. Low-dose radiation settings were used, with a tube voltage of 120 kVp and a tube current of 20 mAs. Other parameters were matrix size 512 × 512 pixels and collimation 64 × 0.6 mm. Images were reconstructed using a bone recon type at a slice thickness setting of 1.25 mm at a 0.625 mm reconstruction increment. CT was performed at the end of maximal inspiration in a single breath-hold and covered the apex of the lung to the diaphragm.

Image analysis

Double reading by radiologists

Two thoracic radiologists with > 5 and 10 years experience blinded to the results of the DL-CAD system independently interpreted each CT image. The final result of double reading was determined by combining the results of both radiologists; any disagreement was resolved through discussion. The following information from each CT scan was recorded: (i) the presence and total number of nodules; (ii) the location of each nodule (indicated by the corresponding numbering of sections and the shortest distance from the nodule to the chest wall in the same axial section); (iii) the size of each nodule (the largest diameter of the nodule in the axial section with its largest area, and the size of part-solid nodules corresponds to the size of solid component and ground-glass components as a whole); and (iv) first impression when characterizing nodules ≥ 5 mm (solid, subsolid including part-solid, and GGNs). Both solid nodules and GGNs are defined as focal nodular areas of increased attenuation, but the difference between them is whether pulmonary vessels and bronchial structures inside are visible. Part solid nodules are defined as having both ground glass and solid components coexisting in a nodule area. As there is no direct way to evaluate the accuracy of DL-CAD system nodule measurement, all of the detected nodules were also measured manually.

Evaluation by a deep learning-based computer-aided diagnosis (DL-CAD) system

A commercial CAD system (σ-Discover/Lung, 12 Sigma Technologies Co. Ltd., Beijing, China) based on deep convolutional neural networks (DL-CAD) was used to process the LDCT images to identify and characterize lung nodules (nodule by nodule). The training data used to build this DL-CAD system included public databases, such as the Lung Image Database Consortium Image Collection from Cancer Imaging Archive (LIDC/IDRI)[30] and the National Cancer Institute NLST.[4] The DL-CAD system is designed to detect nodules ≥ 3 mm and can calculate three dimensional (3D) quantitative measurements, such as the largest 3D diameter (the largest diameter in any plane of nodules), average 3D diameter (the diameter of a sphere equivalent to the volume of a nodule), 3D mass, and 3D volume. The DL-CAD system also characterizes nodule types, such as solid, part-solid, and GGNs, and predicts the likelihood of malignancy for each detected nodule.

Evaluation by the DL-CAD system + an expert panel

As reported by van Riel *et al*., there is only moderate inter-observer and intra-observer agreement in terms of nodule measurement and characterization.[31] Therefore, when choosing the gold standard, a chest radiologist expert panel (with > 30 years experience interpreting chest images) interpreted the double reading results nodule by nodule and combined the results with the DL-CAD system. The final results including the number of nodules, manual measurement, and characterization of each nodule ≥ 5 mm were taken as the gold standard and any disagreement was resolved through discussion. The positive findings were divided into three groups: true nodules (< 5 mm and ≥ 5 mm); benign lesions (fissure thickening, pleural plaque and thickening, fibrosis, pulmonary plaque, bronchiectasis, etc.); and non-lesions, including artifacts and normal anatomy (lung vasculature, lung hilum, cartilage of rib, azygous vein, mediastinal lymph node, superior vena cava, pericardial fat pad, thoracic bone hyperplasia, wall of bronchus, diaphragm, subclavian artery, etc.).

Data analysis

SPSS version 23 (IBM Corp., Armonk, NY, USA) was used to perform all data analysis. Based on the gold standard (expert panel + the DL-CAD system), the detection rates of true nodules were calculated for double reading and the DL-CAD system independently and compared using a chi-square test, with a P value < 0.05 indicating statistically significant difference. The false positive rate (the mean of the number of false positive findings of each CT scan) of double reading and the DL-CAD system were compared using a Student's *t*-test. The positive predictive value of the DL-CAD system and double reading was defined as the ratios between true positive findings and all findings detected by the DL-CAD system and double reading, respectively.

To evaluate the accuracy of the DL-CAD system and double reading for characterizing detected nodules, the sensitivity and specificity were calculated. For instance, for nodules ≤ 5 mm, sensitivity in characterizing solid nodules was defined as the proportion of true solid nodules that were correctly diagnosed, while specificity was defined as the percentage of false solid nodules that were correctly discriminated. The Pearson's correlation coefficient was calculated between the largest 3D diameter by the DL-CAD system and size measured manually with > 0.7, 0.5–0.7, and < 0.5 indicating strong, moderate, and weak correlation, respectively.[32]

Results

Population

Based on the inclusion and exclusion criteria, a total of 346 LDCT scans were included from 346 healthy subjects aged > 50 years old or aged < 50 with a heavy smoking habit (mean age 51.01 ± 10.24 years), comprising 221 men (mean age 51.53 ± 9.92 years) and 125 women (mean age 50.10 ± 10.84 years). Twenty-one CT scans were excluded because of either severe morphological changes in the lung or impaired image quality (Fig 1).

Performance of the DL-CAD system for nodule detection

Based on the results of the expert panel and the DL-CAD system, a total of 812 true nodules were detected. Double reading by radiologists detected 643 nodules with a sensitivity of 79.2% (95% confidence interval [CI] 76.4–82.0%), while the DL-CAD system detected 700 nodules with a sensitivity of 86.2% (95% CI 84.1–88.8%; P < 0.001). The DL-CAD system was more sensitive than double reading not only for detecting nodules no smaller than 5 mm (96.5%, 95% CI 93.4–99.5% vs. 88.0%, 95% CI 82.6–93.4%;

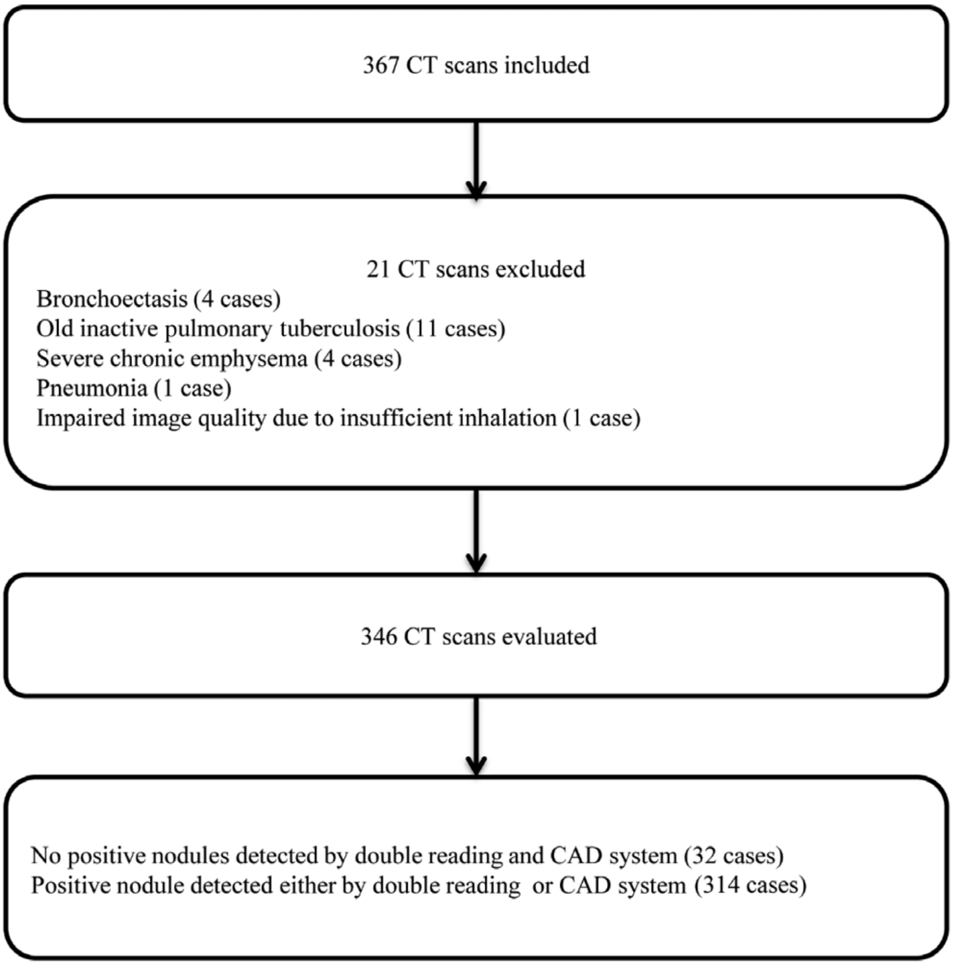
367 CT scans included
21 CT scans excluded
Bronchoectasis (4 cases)
Old inactive pulmonary tuberculosis (11 cases)
Severe chronic emphysema (4 cases)
Pneumonia (1 case)
Impaired image quality due to insufficient inhalation (1 case)
346 CT scans evaluated
No positive nodules detected by double reading and CAD system (32 cases)
Positive nodule detected either by double reading or CAD system (314 cases)

Figure 1 Flowchart showing inclusion and exclusion process. CAD, computer-aided diagnosis; CT, computed tomography.

P = 0.008), but also for detecting nodules < 5 mm (84.3%, 95% CI 81.5–87.0% vs. 77.5%, 95% CI 74.3–80.7%; *P* < 0.001). The false positive rate (per CT scan) of the DL-CAD system (1.53, 529/346, range: 0–11, median 1) was considerably higher than that of double reading (0.13, 44/346, range: 0–3, median 0) (*P* < 0.001). The top five causes of false positive nodules detected by the DL-CAD system were: normal pulmonary vasculature (247/529), pleural plaque and thickening (125/529), hilum (39/529), fibrosis (32/529), and artifacts (25/529) (Table 1). In contrast, double reading was much less likely to misdiagnose non-nodules as nodules, except for occasionally mistreating pulmonary vasculature (36/44) and artifacts (5/44) as nodules. The positive predictive values for double reading and the DL-CAD system were 93.6% (95% CI 91.8–95.4%) and 57.0% (95% CI 54.2–60.0%), respectively (*P* < 0.001) (Table 2). As for missed small nodules (< 5 mm), the average size of nodules missed by the DL-CAD system based on manual measurement was significantly smaller than nodules missed by double reading (0.18 ± 0.09 cm vs. 0.28 ± 0.08 cm; *P* < 0.001).

Performance of the DL-CAD system for nodule measurement

We singled out the 700 nodules detected by the DL-CAD system with the largest 3D diameters and then manually measured the size of each nodule, which was defined as the largest diameter of the nodule in the axial section with its largest area. There was a strong correlation between the size measured manually (mean 0.4 ± 0.2 cm) and the largest 3D diameter measured by the DL-CAD system (mean 0.6 ± 0.2 cm) (r = 0.88), as shown in Figure 2.

Performance of the DL-CAD system in characterizing nodules

In order to evaluate the accuracy of the DL-CAD system in characterizing each nodule, we only focused on those true nodules no smaller than 5 mm (based on the size measured manually) determined by the gold standard. As a result, we analyzed 142 nodules ≥ 5 mm, among which 5 (5/142) and 17 (17/142) were missed by the DL-CAD system and

Table 1 Distribution of detected lesions by the DL-CAD system and double reading based on lesion types

Lesion characteristics	Double reading	DL-CAD system	Gold standard (expert panel + DL-CAD system)
Total number of lesions	687	1229	812
Benign lesions			
Pleural plaque & thickening	1	125	—
Parenchymal plaque	0	9	—
Fissure thickening	1	21	—
Fibrosis	1	32	—
Bronchiectasis	0	2	—
Non-lesions (normal anatomy)			
Pulmonary vasculature	36	247	—
Hilum	0	39	—
Rib cartilage	0	9	—
Azygous vein	0	8	—
Superior vena cava	0	1	—
Fat pad of pericardium	0	2	—
Thoracic bone hyperplasia	0	3	—
Wall of bronchus	0	2	—
Diaphragm	0	2	—
Left subclavian artery	0	1	—
Mediastinal lymph nodes	0	1	—
Artifacts	5	25	—
True nodules			
≥ 5 mm	125	137	142
< 5 mm	518	563	668

DL-CAD, deep learning-based computer-aided diagnosis.

Table 2 Performance of the DL-CAD system in nodule detection

Variables	Double reading	DL-CAD system	*P*
Sensitivity			
All nodules	79.2%(95% CI 76.4–82.0)	86.2% (95% CI 84.1–88.8)	< 0.001*
Nodules ≥ 5 mm	88.0% (95% CI 82.6–93.4)	96.5% (95% CI 93.4–99.5)	0.008*
Nodules < 5 mm	77.5% (95% CI 74.3–80.7)	84.3% (95% CI 81.5–87.0)	< 0.001*
False positive/examination	0.13 (44/346)	1.53 (529/346)	< 0.001*
Positive predictive value	93.6% (95% CI 91.8–95.4)	57.0% (95% CI 54.2–60.0)	< 0.001*

*Indicates statistical significance. CI, confidence interval; DL-CAD, deep learning-based computer-aided diagnosis.

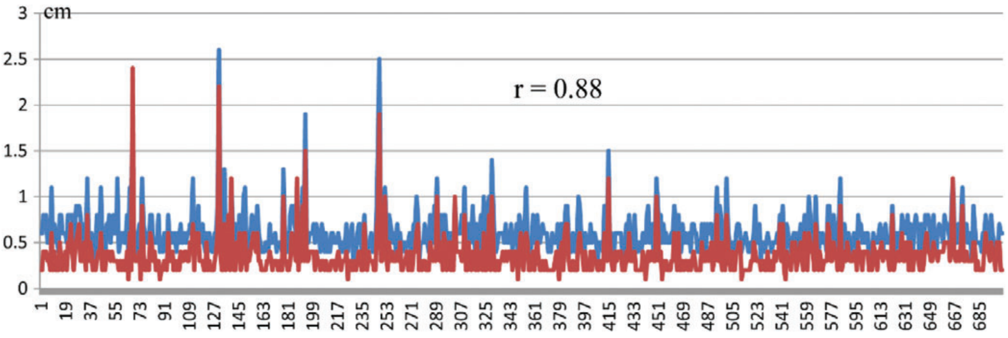

Figure 2 Correlation between the largest three-dimensional diameters by the deep learning-based computer-aided diagnosis (DL-CAD) system and manual measurement. (——) CAD and (——) manually.

double reading, respectively ($P = 0.008$). The DL-CAD system misinterpreted nine solid nodules as part-solid nodules (Fig 3), but no subsolid nodules as solid nodules, giving a sensitivity of 90.3% and specificity of 100.0%. The DL-CAD system misinterpreted four part solid nodules as GGNs (Fig 4), but no GGNs as non-GGNs, giving a

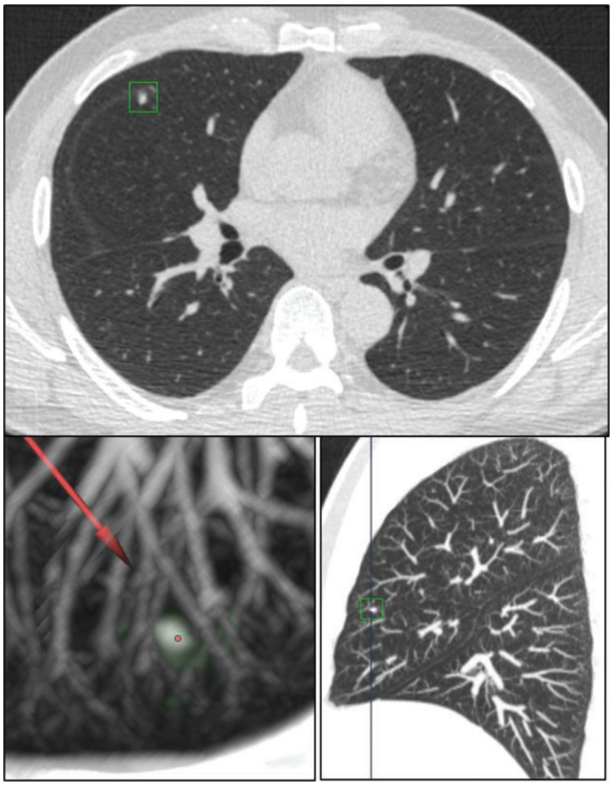

Figure 3 The deep learning-based computer-aided diagnosis (DL-CAD) system misinterpreted a fissure-attached solid nodule as a part-solid nodule.

sensitivity of 100.0% and specificity of 96.1%. However, for part solid nodules, the sensitivity dropped to 55.5% and specificity to 93.0%, as the DL-CAD system misinterpreted nine solid nodules as part-solid nodules (Fig 3) and four part-solid nodules as GGNs (Fig 4). In contrast, double reading only misinterpreted one solid nodule as a part-solid nodule (Fig 5) and one part-solid nodule as a solid nodule (Fig 6), which yielded sensitivity and specificity as high as 98.7% and 97.7% for solid nodules, 90.9% and 99.1% for part solid nodules, and 100.0% each for GGNs, respectively (Tables 3–4).

Discussion

In this study, we evaluated the performance of a commercial DL-CAD system for detecting and characterizing different types of lung nodules. Consistent with the results of most previous studies, our study showed that the DL-CAD system could detect more nodules than double reading, regardless of nodule size.[11,13–15,17,33] Recently, a report based on the International Early Lung Cancer Action Program (I-ELCAP) study showed that in 75% of confirmed cancer patients, the corresponding small nodules could be detected in previous CT scans.[34] Thus, the first priority of lung cancer screening is to ensure detection of all small nodules and record them as the baseline for future surveillance. In this regard, we found that the DL-CAD system showed a great advantage in detecting nodules, especially small nodules (< 5 mm), compared to double reading. Furthermore, a large number of studies have suggested that by complementing each other, the combination of the DL-CAD system and analysis by radiologists could detect more nodules than radiologists alone.[33,35–38] Therefore, the DL-CAD system could act as a second pair of eyes to double check for any possible missed nodules, which could significantly reduce the number of missed nodules and potentially make an impact on lung cancer screening.

Also consistent with previous findings by conventional CAD, our study showed that the sensitivity in nodule detection increases with increases in nodule size when evaluated by both radiologists and the DL-CAD system.[10,39–41]

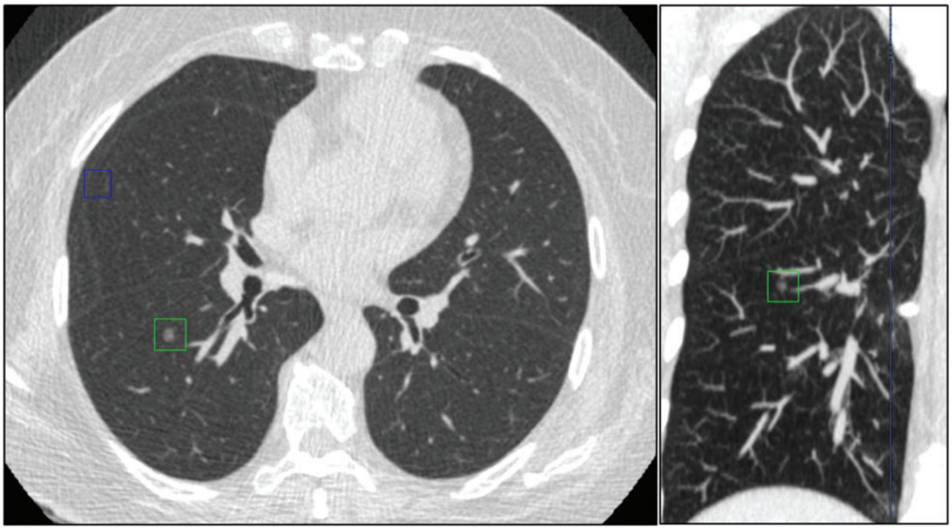

Figure 4 The deep learning-based computer-aided diagnosis DL-CAD system mistakenly interpreted a part solid nodule as a ground-glass nodule (GGN).

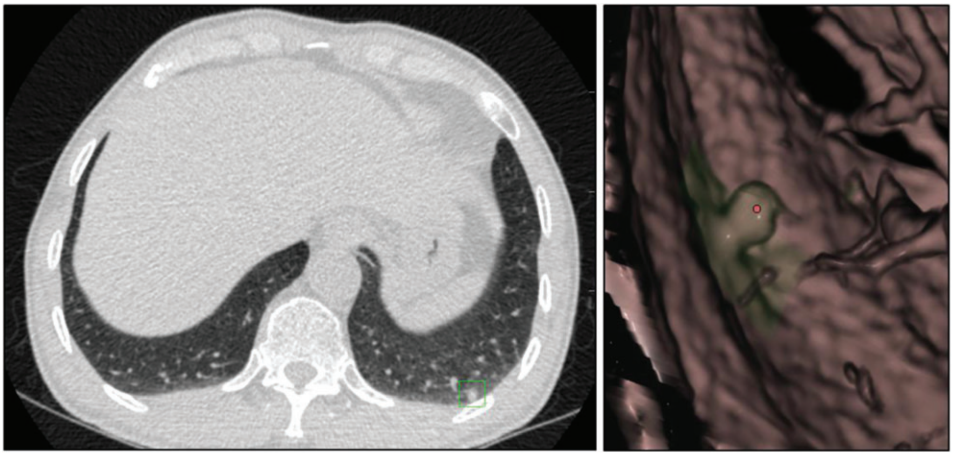

Figure 5 Double reading misinterpreted a solid nodule as a part-solid nodule.

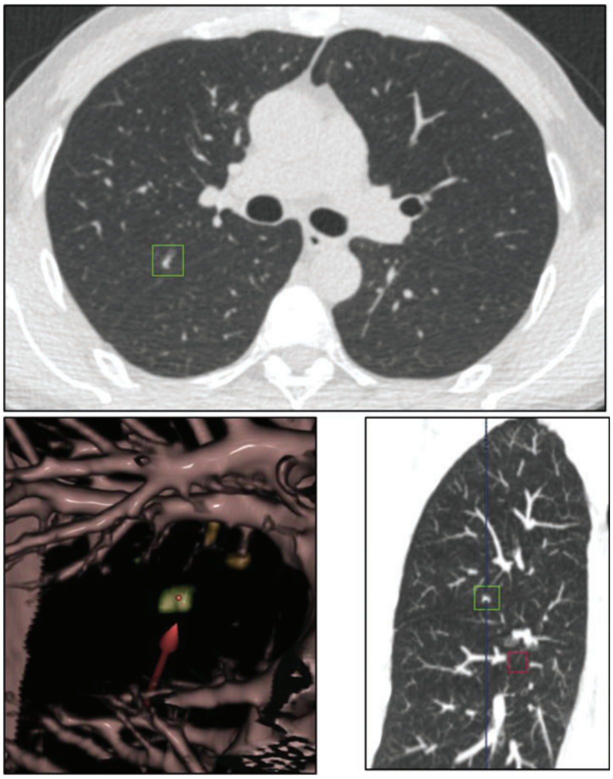

Figure 6 Double reading misdiagnosed a part-solid nodule as a solid nodule.

Interestingly, our study showed that the nodules (< 5 mm) missed by DL-CAD system tended to be smaller than those missed by double reading (0.18 ± 0.09 cm vs. 0.28 ± 0.08 cm; $P < 0.001$). The DL-CAD system we used was initially trained to detect nodules ≥ 3 mm, and was therefore less sensitive in detecting nodules < 3mm, which may explain why 86 (81.9%) of the 105 small nodules (< 5 mm) missed by the DL-CAD system were as small as 1–2 mm. Surprisingly, 56.7% (85/150) of small nodules (< 5 mm) missed by double reading were within 3–4 mm. Double reading in our study might have over detected very small nodules (1–2 mm), which may have led to a slight overestimation of the real detection rate of double reading in clinical practice. Therefore, because most missed nodules by the DL-CAD system were under the sensitivity of the lower size limit, our results suggest that the DL-CAD system may provide a more consistent detection rate.

Currently, the major concern regarding the DL-CAD system is that high sensitivity in nodule detection is always accompanied by high false positive findings. Among 17 CAD studies, the median of false positive rates of the CAD system was 4.1 per scan.[12] For the commercial DL-CAD system we used, based on a convolution neural network algorithm, the false positive rate was 1.53 per CT scan, which is lower than that of most previous studies. Specifically, pulmonary vasculature and artifacts were two of the main causes for false positive findings in both the DL-CAD system and after double reading. In the foreseeable future, with the integration of more data and more sophisticated algorithms into the DL-CAD system, false positive rates could be further reduced to an acceptable level in a clinical setting.

Volumetric measurement of nodules could be tracked over time to assess growth, which is extremely useful to discriminate malignant from benign nodules in the long run. However, the accuracy of volumetric measurement of different CAD systems depends on the segmentation accuracy of different algorithms. We did not use volumetric parameters generated by the DL-CAD system to subgroup nodules in our study, as no standard reference is currently available to evaluate its accuracy. However, we found a strong correlation between the size measured manually and the largest 3D diameter measured by the DL-CAD system (r = 0.88), which indicated the accuracy of measurement by the DL-CAD system. In theory, the CAD system could generate more accurate size parameters than those

Table 3 Performance of the DL-CAD system and double reading for characterizing different types of nodules

Evaluation format	True positive	False positive	False negative	True negative	Sensitivity	Specificity
DL-CAD system						
Solid nodule	84	0	9	44	90.30%	100.00%
Part solid nodule	5	9	4	119	55.50%	93.00%
GGN	35	4	0	98	100.00%	96.10%
Double reading						
Solid nodule	82	1	1	42	98.70%	97.70%
Part solid nodule	10	1	1	113	90.90%	99.10%
GGN	31	0	0	94	100.00%	100.00%

DL-CAD, deep learning-based computer-aided diagnosis; GGN, ground-glass nodule.

Table 4 Nodules characterized based on different standards

Variables	Total number	Solid nodule	Part solid nodule	GGN
Based on gold standard				
DL-CAD system	137	93	9	35
Double reading	125	83	11	31
DL-CAD system + expert panel	142	93	11	38
Based on its system standards				
DL-CAD system	137	84	14	39
Double reading	125	83	11	31

DL-CAD, deep learning based computer-aided diagnosis; GGN, ground-glass nodule.

measured manually, as the subjective variability of different observers is avoided, but only if nodules are accurately segmented.

Although the DL-CAD system detected more nodules in our study, it misinterpreted slightly more nodules ≥ 5 mm than double reading (13 vs. 2 nodules). Indeed, the DL-CAD system showed relatively high accuracy in characterizing solid nodules and GGNs, but not part-solid nodules. One area that has much room to improve is nodule type characterization. Current CAD products cannot accurately distinguish part-solid nodules from either solid nodules or GGNs. As most previous studies have focused on whether the CAD system could detect different types of nodules, and few studies have evaluated whether the CAD system could accurately distinguish different types of nodules, our results need to be confirmed through further investigation.

Our study had some limitations. Firstly, in combining the results of the DL-CAD system + expert panel and determining the characterization of each nodule, subjective evaluation by the radiologists outweighed the influence of the DL-CAD system, which might lead to the superiority of double reading for distinguishing different types of nodules. Because of a lack of objective diagnosis criteria, future studies should determine the Hounsfield unit value range and cutoff value between solid nodules, part solid nodules, and GGNs, in order to reduce the subjective variability of radiologists. Secondly, when the two radiologists were assigned to the double reading, they tended to spend more time and were more alert than in their routine clinical practice, which might have led to a slight overestimation of the real detection rate of double reading in clinical practice. The DL-CAD system we used is a commercial, sophisticated, well-trained model, but details on how this model was built were not provided. Our focus in this study was only to evaluate its current performance in our screening data by external validation. If possible, the performance of the conventional CAD system and the DL-CAD system we used should be compared using the same data. Furthermore, the DL-CAD system automatically provided the malignancy likelihood of each nodule detected. We did not validate its accuracy in this paper, simply because we have not yet accumulated a sufficient number of nodules with confirmed pathology. In our next investigation we intend to directly compare the performance of the conventional CAD system and the DL-CAD system using the same data, and then evaluate the performance of the DL-CAD system for predicting the malignancy of nodules.

In conclusion, our DL-CAD system could sensitively detect more nodules than double reading and showed relatively comparative performance in distinguishing solid nodules and GGNs. In the future, the false positive and mischaracterization rates of the DL-CAD system need to be further reduced.

Disclosure

No authors report any conflict of interest.

References

1 World Health Organization. *Cancer Fact sheet No 297*. 2013. [Cited 23 Nov 2018.] Available from URL: http://www.who.int/news-room/fact-sheets/detail/cancer

2 Stewart B, Wild C. *World Cancer Report 2014*. 2015. International Agency for Research on Cancer, Lyon.

3 International Early Lung Cancer Action Program Investigators, Henschke CI, Yankelevitz DF *et al*. Survival of patients with stage I lung cancer detected on CT screening. *N Engl J Med* 2006; **355**: 1763–71.

4 National Lung Screening Trial Research Team, Aberle DR, Adams AM *et al*. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011; **365**: 395–409.

5 Li F, Sone S, Abe H, MacMahon H, Armato SG III, Doi K. Lung cancers missed at low-dose helical CT screening in a general population: Comparison of clinical, histopathologic, and imaging findings. *Radiology* 2002; **225**: 673–83.

6 Torres EL, Fiorina E, Pennazio F *et al*. Large scale validation of the M5L lung CAD on heterogeneous CT datasets. *Med Phys* 2015; **42**: 1477–89.

7 Baldwin DR, Duffy SW, Wald NJ, Page R, Hansell DM, Field JK. UK Lung Screen (UKLS) nodule management protocol: Modelling of a single screen randomised controlled trial of low-dose CT screening for lung cancer. *Thorax* 2011; **66**: 308–13.

8 van Klaveren RJ, Oudkerk M, Prokop M *et al*. Management of lung nodules detected by volume CT scanning. *N Engl J Med* 2009; **361**: 2221–9.

9 Wormanns D, Fiebich M, Saidi M, Diederich S, Heindel W. Automatic detection of pulmonary nodules at spiral CT: Clinical application of a computer-aided diagnosis system. *Eur Radiol* 2002; **12**: 1052–7.

10 Brown MS, Goldin JG, Suh RD, McNitt-Gray MF, Sayre JW, Aberle DR. Lung micronodules: Automated method for detection at thin-section CT--initial experience. *Radiology* 2003; **226**: 256–62.

11 Armato SG III, Li F, Giger ML, MacMahon H, Sone S, Doi K. Lung cancer: Performance of automated lung nodule detection applied to cancers missed in a CT screening program. *Radiology* 2002; **225**: 685–92.

12 Al Mohammad B, Brennan PC, Mello-Thoms C. A review of lung cancer screening and the role of computer-aided detection. *Clin Radiol* 2017; **72**: 433–42.

13 Yuan R, Vos PM, Cooperberg PL. Computer-aided detection in screening CT for pulmonary nodules. *AJR Am J Roentgenol* 2006; **186**: 1280–7.

14 Lee IJ, Gamsu G, Czum J, Wu N, Johnson R, Chakrapani S. Lung nodule detection on chest CT: Evaluation of a computer-aided detection (CAD) system. *Korean J Radiol* 2005; **6**: 89–93.

15 Jacobs C, van Rikxoort EM, Murphy K, Prokop M, Schaefer-Prokop CM, van Ginneken B. Computer-aided detection of pulmonary nodules: A comparative study using the public LIDC/IDRI database. *Eur Radiol* 2016; **26**: 2139–47.

16 Christe A, Leidolt L, Huber A *et al*. Lung cancer screening with CT: Evaluation of radiologists and different computer assisted detection software (CAD) as first and second readers for lung nodule detection at different dose levels. *Eur J Radiol* 2013; **82**: e873–8.

17 Zhao Y, de Bock GH, Vliegenthart R *et al*. Performance of computer-aided detection of pulmonary nodules in low-dose CT: Comparison with double reading by nodule volume. *Eur Radiol* 2012; **22**: 2076–84.

18 Filho AOC, Silva AC, de Paiva AC, Nunes RA, Gattass M. 3D shape analysis to reduce false positives for lung nodule detection systems. *Med Biol Eng Comput* 2017; **55**: 1199–213.

19 Mao K, Deng Z. Lung nodule image classification based on local difference pattern and combined classifier. *Comput Math Methods Med* 2016; **2016**: 1091279.

20 da Silva GLF, Valente TLA, Silva AC, de Paiva AC, Gattass M. Convolutional neural network-based PSO for lung nodule false positive reduction on CT images. *Comput Methods Programs Biomed* 2018; **162**: 109–18.

21 Dou Q, Chen H, Yu L, Qin J, Heng PA. Multilevel contextual 3-D CNNs for false positive reduction in pulmonary nodule detection. *IEEE Trans Biomed Eng* 2017; **64**: 1558–67.

22 Li W, Cao P, Zhao D, Wang J. Pulmonary nodule classification with deep convolutional neural networks on computed tomography images. *Comput Math Methods Med* 2016; **2016**: 6215085.

23 Lederlin M, Puderbach M, Muley T *et al*. Correlation of radio- and histomorphological pattern of pulmonary adenocarcinoma. *Eur Respir J* 2013; **41**: 943–51.

24 Lim HJ, Ahn S, Lee KS *et al*. Persistent pure ground-glass opacity lung nodules ≥ 10 mm in diameter at CT scan: Histopathologic comparisons and prognostic implications. *Chest* 2013; **144**: 1291–9.

25 Diederich S, Wormanns D, Semik M *et al*. Screening for early lung cancer with low-dose spiral CT: Prevalence in 817 asymptomatic smokers. *Radiology* 2002; **222**: 773–81.

26 Henschke CI, McCauley DI, Yankelevitz DF *et al*. Early lung cancer action project: A summary of the findings on baseline screening. *Oncologist* 2001; **6**: 147–52.

27 Swensen SJ, Jett JR, Sloan JA *et al*. Screening for lung cancer with low-dose spiral computed tomography. *Am J Respir Crit Care Med* 2002; **165**: 508–13.

28 Henschke CI, Yankelevitz DF, Naidich DP *et al*. CT screening for lung cancer: Suspiciousness of nodules according to size on baseline scans. *Radiology* 2004; **231**: 164–8.

29 Swensen SJ, Jett JR, Hartman TE *et al*. Lung cancer screening with CT: Mayo Clinic experience. *Radiology* 2003; **226**: 756–61.

30 Armato III SG, McLennan G, Bidaut L, et al. Data from LIDC-IDRI. *The Cancer Imaging Archive*. 2015. [Cited 23 Nov 2018.] Available from URL: http://doi.org/10.7937/K9/TCIA.2015.LO9QL9SX

31 van Riel SJ, Sánchez CI, Bankier AA *et al*. Observer variability for classification of pulmonary nodules on low-dose CT images and its effect on nodule management. *Radiology* 2015; **277**: 863–71.

32 Sedgwick P. Pearson's correlation coefficient. *BMJ* 2012; **345**: e4483.

33 Fraioli F, Bertoletti L, Napoli A *et al*. Computer-aided detection (CAD) in lung cancer screening at chest MDCT: ROC analysis of CAD versus radiologist performance. *J Thorac Imaging* 2007; **22**: 241–6.

34 Henschke CI, Yankelevitz DF, Yip R *et al*. Lung cancers diagnosed at annual CT screening: Volume doubling times. *Radiology* 2012; **263**: 578–83.

35 Rubin GD, Lyo JK, Paik DS *et al*. Pulmonary nodules on multi-detector row CT scans: Performance comparison of radiologists and computer-aided detection. *Radiology* 2005; **234**: 274–83.

36 Bogoni L, Ko JP, Alpert J *et al*. Impact of a computer-aided detection (CAD) system integrated into a picture archiving and communication system (PACS) on reader sensitivity and efficiency for the detection of lung nodules in thoracic CT exams. *J Digit Imaging* 2012; **25**: 771–81.

37 Awai K, Murao K, Ozawa A *et al*. Pulmonary nodules at chest CT: Effect of computer-aided diagnosis on radiologists' detection performance. *Radiology* 2004; **230**: 347–52.

38 White CS, Pugatch R, Koonce T, Rust SW, Dharaiya E. Lung nodule CAD software as a second reader: A multicenter study. *Acad Radiol* 2008; **15**: 326–33.

39 Setio AA, Jacobs C, Gelderblom J, van Ginneken B. Automatic detection of large pulmonary solid nodules in thoracic CT images. *Med Phys* 2015; **42**: 5642–53.

40 Marten K, Engelke C, Seyfarth T, Grillhosl A, Obenauer S, Rummeny EJ. Computer-aided detection of pulmonary nodules: Influence of nodule characteristics on detection performance. *Clin Radiol* 2005; **60**: 196–206.

41 Ko JP, Betke M. Chest CT: Automated nodule detection and assessment of change over time--preliminary experience. *Radiology* 2001; **218**: 267–73.