Original Article

Application value of a computer-aided diagnosis and management system for the detection of lung nodules

Jingwen Chen[1#], Rong Cao[1#], Shengyin Jiao[2], Yunpeng Dong[2], Zilong Wang[2], Hua Zhu[1], Qian Luo[1], Lei Zhang[1], Han Wang[1], Xiaorui Yin[1]

[1]Department of Radiology, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China; [2]Department of R&D, VoxelCloud, Shanghai, China

Contributions: (I) Conception and design: X Yin, H Wang; (II) Administrative support: J Chen; (III) Provision of study materials or patients: L Zhang, H Zhu, Q Luo; (IV) Collection and assembly of data: J Chen, R Cao; (V) Data analysis and interpretation: S Jiao, Y Dong, Z Wang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Xiaorui Yin, MD, PhD. Department of Radiology, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, No. 100 Haining Rd., Shanghai 200080, China. Email: yinxiaorui2022@163.com.

Background: Computer-aided diagnosis (CAD) systems can help reduce radiologists' workload. This study assessed the value of a CAD system for the detection of lung nodules on chest computed tomography (CT) images.

Methods: The study retrospectively analyzed the CT images of patients who underwent routine health checkups between August 2019 and November 2019 at 3 hospitals in China. All images were first assessed by 2 radiologists manually in a blinded manner, which was followed by assessment with the CAD system. The location and classification of the lung nodules were determined. The final diagnosis was made by a panel of experts, including 2 associate chief radiologists and 1 chief radiologist at the radiology department. The sensitivity for nodule detection and false-positive nodules per case were calculated.

Results: A total of 1,002 CT images were included in the study, and the process was completed for 999 images. The sensitivity of the CAD system and manual detection was 90.19% and 49.88% (P<0.001), respectively. Similar sensitivity was observed between manual detection and the CAD system in lung nodules >15 mm (P=0.08). The false-positive nodules per case for the CAD system were 0.30±0.84 and those for manual detection were 0.24±0.68 (P=0.12). The sensitivity of the CAD system was higher than that of the radiologists, but the increase in the false-positive rate was only slight.

Conclusions: In addition to reducing the workload for medical professionals, a CAD system developed using a deep-learning model was highly effective and accurate in detecting lung nodules and did not demonstrate a meaningfully higher the false-positive rate.

Keywords: Artificial intelligence (AI); computer-aided diagnosis (CAD); convolutional neural network; pulmonary nodule; nodule detection

Submitted Nov 22, 2022. Accepted for publication Aug 22, 2023. Published online Sep 18, 2023.
doi: 10.21037/qims-22-1297
View this article at: https://dx.doi.org/10.21037/qims-22-1297

© Quantitative Imaging in Medicine and Surgery. All rights reserved.

Introduction

Lung cancer is one of the most common cancers (1,2) and is currently diagnosed through biopsies and imaging techniques such as lung computed tomography (CT) scans (3). Early lung cancer manifests in imaging as small lung nodules and may be either solid or part-solid in attenuation (4). Furthermore, previous studies indicate that 10–31% of small lung nodules may be missed on CT images (5-8). Accurate and early lung cancer screening is essential for improving treatment outcomes and survival (9).

The development of artificial intelligence (AI) using a convolutional neural network has tremendously improved the classification, detection, and segmentation of images (10-12). One of the greatest advances in radiology is the detection of lesions on CT scans using deep-learning models, such as computer-aided diagnosis (CAD) with automated algorithms (10-14). In order to improve the detection rate of small nodules and reduce the challenges faced by radiologists due to the increasing workload, researchers have examined detecting lung nodules using a CAD-based approach (15).

Recently, a study reported the use of a CAD system before manual diagnosis to detect lung cancer using low-dose CT (LDCT) screening (16). Another study on 1,386 Canadian smokers who were randomly assigned to receive LDCT screening by either a CAD system or a radiologist reported that the CAD-based LDCT screening saved radiologists' reading time, particularly in patients with low-risk or no risk of lung nodules (17). In addition, several studies reported that nodule-detection algorithms using CAD could effectively obtain better detection and classification results (18-20). Nevertheless, further clinical evidence, such as that of higher sensitivity and equivalent false-positive rates, is needed to show the concordance between results obtained using AI and regular clinical practice.

The present study investigated the performance of an AI-based CAD system for lung nodule detection as compared with that of routine manual detection. The sponsor of this CAD system submitted the application for clinical use to the National Medical Products Administration (NMPA) in China in 2017 and obtained a class III medical device certification in 2021 based on the results of this study, which was conducted between 2019 and 2020. This study provides high-level evidence to support the application of this CAD system in clinical CT reading and diagnosis, which could serve as a junior reader to assist radiologists during lung nodule detection with LDCT. We present this article in accordance with the STARD reporting checklist (available at https://qims.amegroups.com/article/view/10.21037/qims-22-1297/rc).

Methods

Study design and patients

Chest LDCT images were retrospectively reviewed in this study. LDCT in this study was defined as a chest CT scan with a parameter set as ≤60 mAs at 120 kVp on a 64-detector row or higher CT scanner (21-23). Patients received chest LDCT screening due to personal request or physician's advice in annual health checkups between August 2019 and November 2019. All patients had detected lung nodules in their health checkup report as assessed by the radiology department at 3 hospitals in Shanghai, China: Shanghai General Hospital (Hospital 1), Shanghai East Hospital (Hospital 2), and Shanghai Shuguang Hospital Affiliated to the Shanghai University of Traditional Chinese Medicine (Hospital 3). Patient data were collected from these 3 hospitals in a ratio of 1:1:1. This study was reviewed and approved by the research ethics committees of Shanghai General Hospital (No. 2018-75), Shanghai East Hospital (No. 2017-052), and Shanghai Shuguang Hospital Affiliated to the Shanghai University of Traditional Chinese Medicine (No. 2018-620-49) and was conducted in accordance with the Declaration of Helsinki (as revised in 2013). Informed consent was waived due to the retrospective and anonymized nature of this study. The study was registered with the China Clinical Trials Database (ChiCTR2000029278).

The inclusion criteria were: (I) lung CT images in DICOM format; (II) images covering the whole lung from the pulmonary apex to the diaphragm; (III) images with a thickness ≤2.5 mm and a slice interval ≤2.5 mm; (IV) images from a lung CT screening program; and (V) complete patient information such as age and sex. The exclusion criteria were: (I) lung CT images with no lung nodules according to previous diagnoses; (II) images that could not be imported into an image workstation; (III) severe image artifacts due to metal implants or other reasons; or (IV) images with significant morphological changes other than lung nodules.

Manual detection

Based on a double-reading design, the CT images were reviewed by the junior radiologist first and then by the senior radiologist. The working years of the radiologists

© Quantitative Imaging in Medicine and Surgery. All rights reserved.

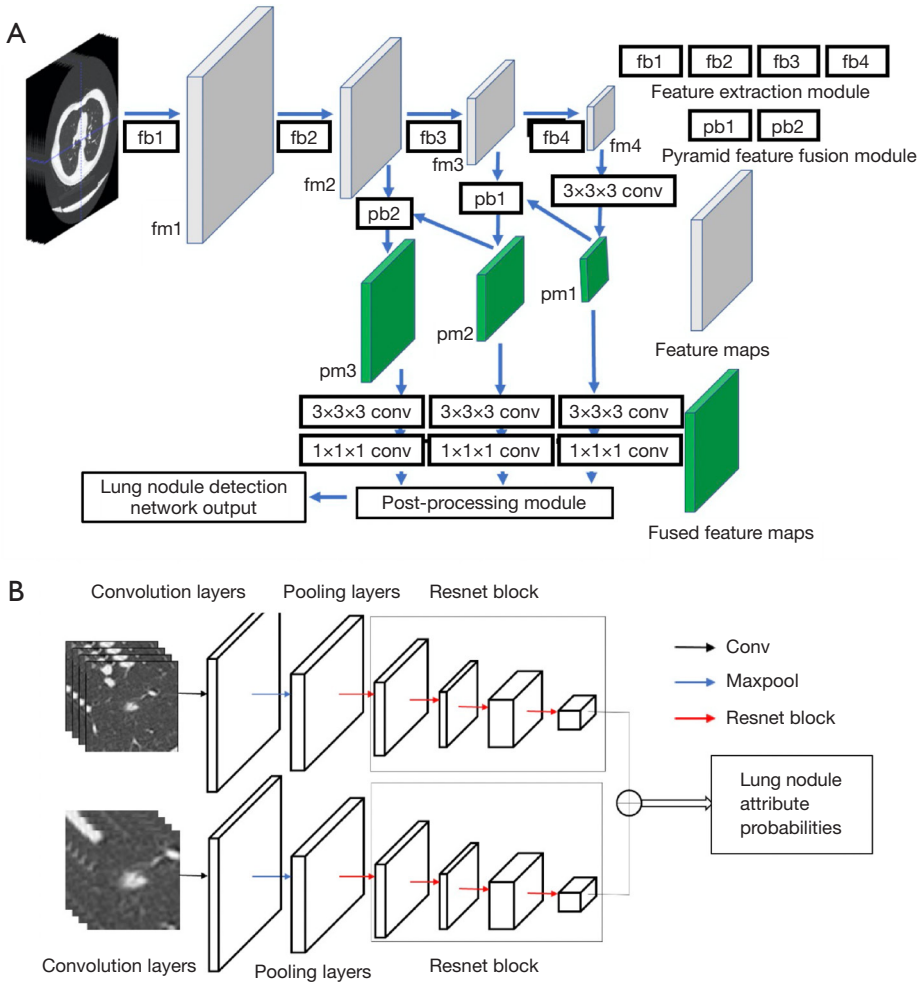

Figure 1 Flowcharts of the deep-learning algorithm in the computer-aided diagnosis system. (A) The nodule detection flowchart. (B) The nodule classification network flowchart. fb, feature block; fm, feature map; pb, pyramid block; conv, convolution.

ranged from 2 to 34 years. They were required to number the nodules and record each nodule's location, type, and size. The details of manual detection are described in Appendix 1.

CAD system and the development of the deep-learning algorithm

After manual detection, all images were input into VoxelCloud Thorax software (version 1.0, VoxelCloud, Los Angeles, CA, USA), and a CAD system was built using a state-of-the-art neural network algorithm. It was designed in 2017 for application in NMPA market authorization. Changes to the system were not permitted after the application was submitted. The nodules' location, type, and size were assessed using the CAD system, which was primed for approval in the Chinese market through this study, functioning as a medical device clinical trial. The radiologists used the CAD system by following the manufacturer's instructions and worked independently of the radiologists who manually assessed the images. A state-of-the-art feature pyramid network (24) was applied to detect lung nodules, and a dual-channel 3D residual neural network was applied to perform nodule classification (*Figure 1*).

Gold standard

A panel comprising 4 associate chief radiologists (A, B, C, and D) with more than 10 years of experience and 1 chief radiologist with more than 15 years of experience

© Quantitative Imaging in Medicine and Surgery. All rights reserved.

specializing in thoracic radiology was established to generate the gold standard for nodule detection in imageology according to the China National Guideline of Classification, Diagnosis, and Treatment for Lung Nodules (2016 version). Associate chief radiologists A and B were blind to which method found the nodules (manual or CAD detection) and independently evaluated each annotated nodule based on the results of the previous radiologists and CAD system. If their evaluations did not match, the chief radiologist evaluated the images for a final diagnosis. Associate chief radiologists A and B evaluated 512 images. Associate chief radiologists C and D did the same as A and B, and they evaluated 490 images.

Follow-up of nodules in a subset of patients

In the Picture Archiving and Communication System (PACS) of Shanghai General Hospital, 133 patients were followed up using imaging numbers, including patients with solid nodules, ground-glass nodules, and mixed ground-glass nodules. Ultimately, 21 patients with significant changes in lung nodules were analyzed in this study. The last follow-up date was in August 2022, and these patients were followed up for at least 3 years.

Statistical analysis

The trial's primary endpoint was a superiority test in which the CAD system showed at least a 10% higher sensitivity than that of manual detection. The equivalence test was that the false-positive nodules per case of the CAD system would be equivalent to those of manual detection. The sample size was calculated using PASS 15.0 (NCSS, Kaysville, UT, USA). The sample size for analyzing the sensitivity of nodule detection was calculated with the superiority test to assess the difference between the 2 proportion models. The sample size for analyzing the mean false-positive nodules per case was calculated by testing 2 mean model with a power of 90%. Based on these 2 sample size calculations, we selected the largest of the results, which was 902 patients. Considering the potential loss of patients, we expanded the sample size to 1,002 patients.

The full analysis set (FAS) included the patients who received at least 1 detection. The per-protocol set (PPS) included the patients who followed the protocol and completed the detection. The statistical analyses were performed in both FAS and PPS.

All statistical analyses were conducted using SAS software (version 9.4; SAS Institute Inc., Cary, NC, USA). Detailed statistical methods are described in Appendix 1.

Subgroup analysis was performed according to the diameter of nodules (<8 *vs.* ≥8 mm and <15 *vs.* ≥15 mm) based on the China National Guideline of Classification, Diagnosis, and Treatment for Lung Nodules (2016 version).

Results

Characteristics of patients and scans

All the chest CT scans included from health checkups in 3 hospitals were conducted in accordance with the LDCT standards. In this study, the CT images of 1,002 patients were reviewed (*Figure 2*). The mean age of all patients was 41.41±12.14 years, and 50.30% were males (*Table 1*). Three patients were not included in the PPS because the gold standard showed no lung nodules. CT scans were obtained using scanners from well-known manufacturers. The mean number of slices, slice thickness, and slice distance were 372.52±81.66, 1.09±0.19, and 0.92±0.13 mm, respectively (Table S1).

According to the gold standard, 5,638 nodules were detected, and the mean diameter was 6.17±1.85 mm [median, 6.10; 25th–75th percentile (Q1–Q3), 5.02–7.05]. The mean diameter of solid, part-solid, and ground-glass nodules was 6.06±1.78 mm (median, 6.06; Q1–Q3, 4.92–6.90; n=4,271), 6.56±2.07 mm (median, 6.45; Q1–Q3, 5.35–7.55; n=584), and 6.48±1.96 mm (median, 6.32; Q1–Q3, 5.26–7.55; n=783), respectively.

Sensitivity and true-positive rate of nodule detection

For each scan, 5.39±5.12 nodules were detected, and 5.09±4.83 true-positive nodules were validated in the CAD system. In the manual detection, 3.05±2.62 nodules and 2.81±2.42 true-positive nodules were detected. In total, 5,085, 2,812, and 5,638 nodules were detected with the CAD system, manual detection, and the reference standard, respectively. The sensitivity of the CAD system and manual detection was 90.19% (95% CI: 89.39–90.96%) and 49.88% (95% CI: 48.56–51.19%) (P<0.001), respectively. The difference in sensitivity between the 2 detection methods was 40.32% (95% CI: 38.80–41.83%), with a lower limit of >10% (*Table 2*). We further evaluated the sensitivity of manual detection in different centers (Table S2). These results seemed to positively correlate with the working years and professional title of the radiologist, which indicated the

© Quantitative Imaging in Medicine and Surgery. All rights reserved.

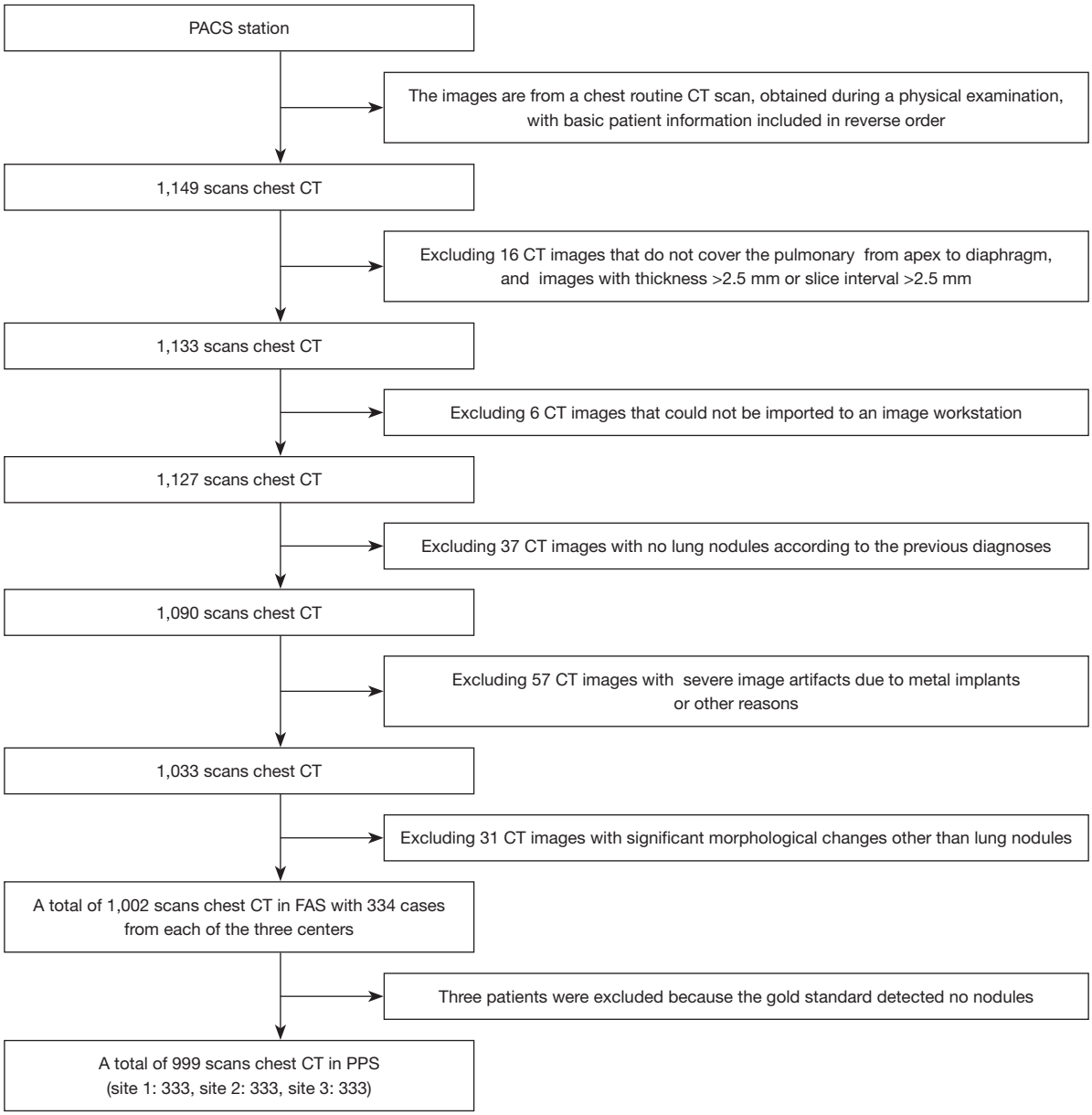

Figure 2 Study flowchart. PACS, picture archiving and communication system; CT, computed tomography; FAS, full analysis set; PPS, per-protocol set.

real scenario of clinical imaging diagnosis in each center.

As shown in *Table 2*, compared with 0.24±0.68 false-positive nodules per case in manual detection, false-positive nodules per case slightly increased to 0.30±0.84 in the CAD system. However, the difference was only 0.06±0.75 and not significant (Q1–Q3, 0.00–0.00; P=0.12). Therefore, the 2 primary endpoints of the clinical trial were met. In addition, sex and age (with 45 years being used to classify patients into the middle-aged and old-aged groups) were used to investigate the sensitivity and false-positive nodules per case between the CAD and manual assessments (Table S3). The results suggested that the false-positive nodules per case of the CAD system were similar to those of the manual detection in the female (0.32±0.89 *vs.* 0.30±0.83, respectively) or middle-aged (0.21±0.57 *vs.* 0.20±0.52, respectively) groups, while the sensitivity of CAD detection

© Quantitative Imaging in Medicine and Surgery. All rights reserved.

Table 1 Characteristics of the patients in FAS

Characteristics of the patients	Hospital 1	Hospital 2	Hospital 3	Total
Age (years)				
N (missing)	334 (0)	334 (0)	334 (0)	1,002 (0)
Mean (SD)	38.84 (10.83)	43.06 (13.79)	42.33 (11.21)	41.41 (12.14)
Min–Max	20.00–68.00	14.00–82.00	18.00–80.00	14.00–82.00
Median	36.00	41.00	41.00	40.00
Q1–Q3	30.00–49.00	31.00–54.00	34.00–51.00	31.00–51.00
Sex, n (%)				
Male	178 (53.29)	168 (50.30)	158 (47.31)	504 (50.30)
Female	156 (46.71)	166 (49.70)	176 (52.69)	498 (49.70)
Manufacturer, n (%)				
GE HealthCare	22 (6.59)	0 (0.00)	0 (0.00)	22 (2.20)
Philips	0 (0.00)	18 (5.39)	1 (0.30)	19 (1.90)
Siemens	171 (51.20)	0 (0.00)	61 (18.26)	232 (23.15)
Canon	0 (0.00)	316 (94.61)	0 (0.00)	316 (31.54)
United Imaging Healthcare	141 (42.22)	0 (0.00)	272 (81.44)	413 (41.22)
Model, n (%)				
Aquilion ONE	0 (0.00)	24 (7.19)	0 (0.00)	24 (2.40)
Aquilion PRIME	0 (0.00)	292 (87.43)	0 (0.00)	292 (29.14)
Brilliance 64	0 (0.00)	18 (5.39)	0 (0.00)	18 (1.80)
Light Speed VCT	22 (6.59)	0 (0.00)	0 (0.00)	22 (2.20)
SOMATOM Definition Flash	142 (42.51)	0 (0.00)	0 (0.00)	142 (14.17)
SOMATOM Force	29 (8.68)	0 (0.00)	61 (18.26)	90 (8.98)
iCT 256	0 (0.00)	0 (0.00)	1 (0.30)	1 (0.10)
uCT 760	141 (42.22)	0 (0.00)	272 (81.44)	413 (41.22)

FAS, full analysis set; SD, standard deviation; Q1–Q3, 25th–75th percentile.

Table 2 Sensitivity and true positives of nodule detection in FAS

Sensitivity and true positives of nodule detection	CAD system	Manual detection	Difference	P value
Gold standard, n	5,638	5,638	–	–
Detection nodules, n	5,085	2,812	–	–
Sensitivity (95% CI) (%)	90.19 (89.39, 90.96)	49.88 (48.56, 51.19)	40.32 (38.80, 41.83)	<0.001
False-positive nodules per case				
N (missing)	1,002 (0)	1,002 (0)	–	–
Mean (SD)	0.30 (0.84)	0.24 (0.68)	0.06 (0.75)	0.12
Min–Max	0.00–11.00	0.00–11.00	–12	–
Median	0.00	0.00	0.00	–
Q1–Q3	0.00–0.00	0.00–0.00	0.00–0.00	–

FAS, full analysis set; CAD, computer-aided diagnosis; CI, confidence interval; SD, standard deviation; Q1–Q3, 25th–75th percentile.

© Quantitative Imaging in Medicine and Surgery. All rights reserved.

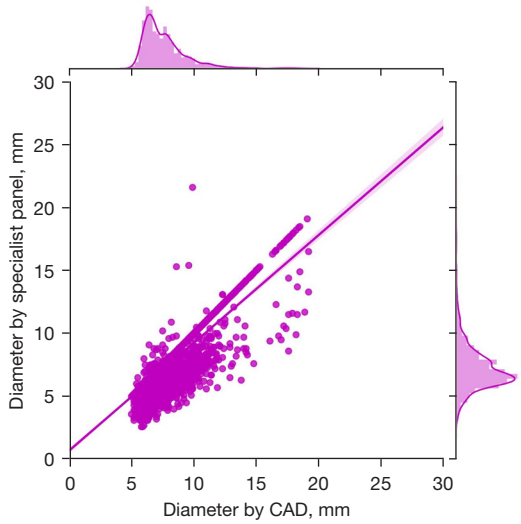

Figure 3 Correlation of diameters between the CAD system and the reference standard. CAD, computer-aided diagnosis.

was better than that of the manual assessment in the male (90.98% *vs.* 46.42%, respectively) and old-aged (90.79% *vs.* 48.20%, respectively) groups (Table S3).

The mean diameter of the nodules measured with the CAD system was 7.67±1.83 mm, which was similar to the mean diameter of 7.25±1.86 mm of the reference standard. *Figure 3* shows the diameter histogram and correlation between diameters in the CAD system and the reference standard (r=0.85; P<0.001). The results in the PPS were the same as those in the FAS.

Figure 4 displays the images of the CAD system and manual detection false-negative nodules, which included solid nodules and ground-glass nodules. Solid nodule 1 and ground glass nodule were missed by the readers but detected by the CAD system. Solid nodule 2 was detected by the readers but missed by the CAD system.

Sensitivity and true-positive rate of nodule detection of nodule diameter

In the FAS, for the nodules with a diameter <8 mm, the sensitivity of the CAD system and manual detection was 89.04% and 52.53%, respectively. There was a significant difference between the 2 methods (P<0.001). For the ≥8 and <15 mm nodules, the sensitivity of the CAD system and manual detection was 98.19% and 31.85%, respectively, representing a significant difference between the 2 detection approaches (P<0.001). For the nodules with a diameter of ≥15 mm, there were no significant differences in the sensitivity between the CAD system and manual detection (P=0.08) (*Table 3*). The results of the PPS were the same as those of the FAS.

Sensitivity and true-positive rate of nodule detection according to different CT manufacturers

The results in the FAS showed that the sensitivity of the CAD system was 88.70% (95% CI: 87.30–89.99%), 92.04% (95% CI: 90.79–93.18%), 89.28% (95% CI: 87.34–91.01%), 95.24% (95% CI: 89.92–98.23%), and 89.38% (95% CI: 82.18–94.39%) for the United-Imaging Health, Canon, Siemens, GE Healthcare, and Philips manufacturers, respectively, all of which were higher than that of manual detection (all P values <0.001) (Table S4).

Detection accuracy according to different nodule types

Table 4 shows that the concordance of solid nodules was high, while the concordance of the part-solid and ground-glass nodules was low. *Table 4* also shows that 3,981 nodules were correctly classified in total, which indicated an accuracy of 78.04% (95% CI: 76.91–79.18%).

The safety evaluation of the medical device clinical trial indicated that the satisfaction rate of both importing CT images and manipulating the software was 100.0% (95% CI: 99.63–100.00%).

Follow-up changes in different types of detected nodules

A total of 21 patients with significant changes in lung nodules were reviewed, including 30 solid, 16 part-solid, and four ground glass nodules. After a follow-up of at least 3 years, 3 nodules (6%) became larger, and 11 nodules (22%) became smaller [including 6 (12%) that disappeared]. Meanwhile, 3 nodules (6%) had a higher density, and 3 nodules (6%) displayed a lower density (Table S5).

Discussion

We found that the detection sensitivity of the CAD system was higher than that of the radiologists, with no marked increase in the false-positive rate. CAD nodule detection techniques have experienced rapid development since the development of neural network–based systems (25,26). Zhang *et al.* reported a study using a deep-learning algorithm to detect and classify lung nodules in

© Quantitative Imaging in Medicine and Surgery. All rights reserved.

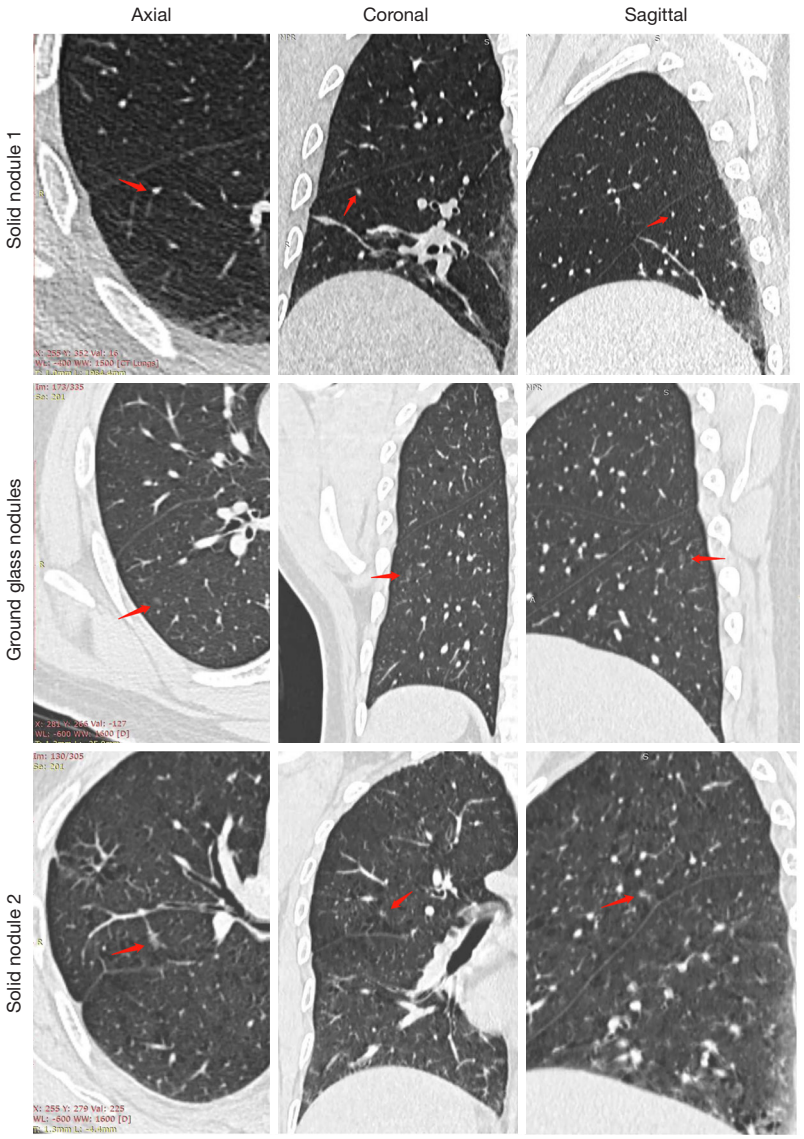

Figure 4 Images of false-negative nodules of the CAD system and manual detection. Solid nodule 1 and ground-glass nodule were missed by manual detection but detected by the CAD system (red arrows). Solid nodule 2 was detected by manual detection but missed by the CAD system (red arrows). CAD, computer-aided diagnosis.

50 CT scans. They found that the sensitivity and specificity of the model were 96.0% and 88.0%, compared with 81.3% and 77.9% for radiologists, respectively. Nevertheless, they did not have strict inclusion and exclusion criteria, and the radiologists' years of experience were not considered (27). Li *et al.* reported a 99.1% sensitivity of a well-trained model compared with a 43.0% sensitivity of experienced radiologists in a dataset with 200 randomly selected CT scans, but their model detected 4.9 false-positive nodules per CT scan (28). Ardila *et al.* demonstrated that the application of their model decreased the false-positive rate by 11% compared with only a 5% decrease in the false-positive rate by radiologists if previous CT scans were not available. When previous CT scans were provided, the detection performance of their model was equal to that of the radiologists (29). Compared with these studies on state-of-the-art models, our study included a considerably larger number of individuals from multiple centers, which was representative of the population that could potentially benefit from such a system. In our study, we found that the

© Quantitative Imaging in Medicine and Surgery. All rights reserved.

Table 3 Sensitivity and true positives of nodule detection for different nodule diameters in FAS

Different nodule diameters	CAD detection	Manual detection	Difference	P value
<8 mm				
Gold standard, n	4,908	4,908	–	–
Detection nodules, n	4,370	2,578	–	–
Sensitivity (95% CI) (%)	89.04 (88.13, 89.90)	52.53 (51.12, 53.93)	36.51 (34.86, 38.16)	<0.001
False-positive				0.01
N (missing)	1,002 (0)	1,002 (0)		
Mean (SD)	0.1627 (0.5217)	0.2196 (0.6613)	−0.0569 (0.6777)	
Min–Max	0.00–4.00	0.00–11.00	−13	
Median	0.00	0.00	0.00	
Q1–Q3	0.00–0.00	0.00–0.00	0.00–0.00	
≥8 and <15 mm				
Gold standard, n	719	719	–	–
Detection nodules, n	706	229	–	–
Sensitivity (95% CI) (%)	98.19 (96.93, 99.03)	31.85 (28.46, 35.39)	66.34 (62.80, 69.88)	<0.001
False-positive				<0.001
N (missing)	1,002 (0)	1,002 (0)		
Mean (SD)	0.1158 (0.4694)	0.0040 (0.0631)	0.1118 (0.4746)	
Min–Max	0.00–7.00	0.00–1.00	−8	
Median	0.00	0.00	0.00	
Q1–Q3	0.00–0.00	0.00–0.00	0.00–0.00	
≥15 mm				
Gold standard, n	11	11	–	–
Detection nodules, n	9	5	–	–
Sensitivity (95% CI) (%)	81.82 (48.22, 97.72)	45.45 (16.75, 76.62)	36.36 (0.86, 73.58)	0.08
False-positive				0.08
N (missing)	1,002 (0)	1,002 (0)		
Mean (SD)	0.0030 (0.0547)	0.0000 (0.0000)	0.0030 (0.0547)	
Min–Max	0.00–1.00	0.00–0.00	0.00–1.00	
Median	0.00	0.00	0.00	
Q1–Q3	0.00–0.00	0.00–0.00	0.00–0.00	

FAS, full analysis set; CAD, computer-aided detection; CI, confidence interval; SD, standard deviation; Q1–Q3, 25th–75th percentile.

sensitivity of the CAD system was 90.19% compared with 49.88% for radiologists. The lower sensitivity of manual detection might have been caused by the fewer years of experience. The multicenter design of this study could balance the differences among readers in different centers and reduce the influence of different years of experience on the accuracy of manual detection.

Our model did not remarkably increase the false-positive rate, which suggests that it has a more accurate performance in clinical practice compared with other models in the

© Quantitative Imaging in Medicine and Surgery. All rights reserved.

Table 4 Accuracy of nodule type detection of CAD in the FAS and PPS

Nodule type	FAS		PPS	
	True (n=3,981)	False (n=1,120)	True (n=3,981)	False (n=1,120)
Solid nodules, n (%)	3,317 (83.32)	132 (11.79)	3,317 (83.32)	132 (11.79)
Part-solid nodules, n (%)	87 (2.19)	373 (33.30)	87 (2.19)	373 (33.30)
Ground-glass nodules, n (%)	577 (14.49)	615 (54.91)	577 (14.49)	615 (54.91)

CAD, computer-aided detection; FAS, full analysis set; PPS, per-protocol set.

field. The CAD system was not only superior to the gold standard in terms of sensitivity but also had a similar false-positive rate to the gold standard, with both measures reaching the target endpoint. The safety and effectiveness of this CAD system were unique among the U.S. Food and Drug Administration (FDA)- or NMPA-approved medical devices for the same intended use. The device could read lung CT independently in the clinical validation phase with false-positive consistency for nodular levels. Due to the high false-positive rate of previous CAD-based screening systems (30), radiologists need to review and reduce the false-positive rate before CAD-based screening can be used in clinical practice. The superiority of our CAD system is that it can work independently in lung screening. Compared with system examined in the study by Li *et al.*, which included a single device and a small sample size (31), our CAD system showed high accuracy and stability when applied to clinical patients in the real world. In our study, the false-positive nodules per case of CAD system were similar to those of manual detection in female and middle-aged patients, while the sensitivity of the CAD system was better than that of manual detection in male or old-aged patients. In clinical practice, it has been found that compared with female and middle-aged patients, male and older adult patients have a higher rate of nonnodular lesions on chest CT images, which can easily affect the detection of nodules.

Based on our results, compared with the CAD system, radiologists are more likely to miss small nodules in routine work. Screening using the CAD system followed by radiologist-based diagnosis according to CT images might be a better method in clinical practice; it may reduce the time spent by radiologists finding suspicious nodules, especially small benign ones, and therefore reduce the possibility of missed detection. Although a given CAD system may the potential to play a key role in early detection and treatment planning, the results generated by the CAD system still need to be reviewed by a senior radiologist. The detection cannot be performed by the CAD system alone. Diseases should be diagnosed and treated based on the cooperation of multidisciplinary teams, including radiologists and clinicians. The contribution of a CAD system includes automated lesion detection, characterization, and segmentation. Clinical tasks such as the prediction of outcomes and treatment response still need to be led by physicians.

Nevertheless, for the further applications of deep-learning models in lung cancer screening, accurately detecting nodules based on imaging findings alone is not enough. It is better to compare pathology findings of the unresectable nodules with those of CAD system results in the healthy population. More analysis of the nodules should be applied in the future. In addition, some have investigators explored the ability of models to predict the characteristics of nodules. For example, researches have trained models to predict the histopathological subtypes of ground-glass nodules, such as adenocarcinoma in situ, minimally invasive adenocarcinoma, and invasive adenocarcinoma (32-34). Some studies applied their model to predict the malignancy of detected nodules (35-37). Ohno *et al.* reported that applying a convolutional neural network-based CAD system improved the area under the curve of nodule volume measurement from 0.67 to 0.94 (38). More recently, Lin *et al.* reported a well-trained deep-learning model for differentiating between granuloma nodules and solid lung cancer (39). Moreover, some investigators even explored using deep-learning models on CT scans to predict the mutations of genes in the tumor, including those of *EGFR* and *TP53* (40,41). During the development of AI models, our goal was to provide more useful information for lung nodule detection.

Nevertheless, our study has some limitations. The CAD system did not accurately classify the subtypes of nodules, particularly part-solid nodules, with only 87 of the 460 nodules being correctly classified. One possible explanation for this could be that the subtypes

© Quantitative Imaging in Medicine and Surgery. All rights reserved.

of nodules are not evenly distributed naturally. Part-solid nodules account for only a small fraction of all nodules. Therefore, we believe that the model might not have been trained with a sufficient number of part-solid nodules during development. In the future, we will include and sample more part-solid nodules in the training set to solve this problem. Another problem with the CAD system was that most of the automated measurements of nodule diameters were larger than the corresponding ones measured manually. This might be because the contours of nodules annotated for training the model were always slightly larger than the actual size of the nodules. More investigations are needed to develop a proper postprocessing method that presents a contour size similar to the actual size of the nodules. In addition, the CT images with no lung nodules were excluded from this study. Further large-scale studies will be performed in the future to comprehensively evaluate the false-positive rate of CAD systems, including both patients with detected lung nodules and healthy individuals. We will also consider increasing the number of radiologists and include radiologists of different nationalities in our future studies.

Conclusions

This study found that a CAD software system had high sensitivity for detecting lung nodules and did not markedly increase the false-positive rate compared with manual assessment by experienced radiologists. The accurate and consistent detection performance of our CAD system implies promising application in clinical practice. The clinical significance is that a commercial CAD system demonstrated better effectiveness in serving patients in clinical settings. The positive impact on clinical practice is that a CAD system will be able to assist radiologists in improving accuracy and work efficiency, reducing workload and work intensity. The negative clinical impact is that it is prone to overdiagnosis. A CAD system can detect many nodules missed by radiologists, and most of these nodules are small nodules (<5 mm). These nodules are mostly benign, and even if they are malignant, they will not affect the morbidity and mortality of patients. The common treatment for such patients is an extended watch-and-wait protocol, which may be extremely burdensome psychologically. Therefore, we should combine the results of the CAD system with diagnosis by multidisciplinary teams to evaluate the size, shape, and dynamic changes of lung nodules in the future.

Acknowledgments

Funding: This study was supported in part by the National Natural Science Foundation of China (Nos. 82271969, 81871400 and 81901795), the Program of Shanghai Education Commission (No. 202101070002E00085), the Grant of Shanghai Municipal Science and Technology Commission (Nos. 19411951403 and 19YF1440600), and the Grant of Shanghai Hospital Development Center (No. SHDC22011310-A). The funders had no role in the study design, collection, analysis, and interpretation of data, writing of the report, or the decision to submit the article for publication.

Footnote

Reporting Checklist: The authors have completed the STARD reporting checklist. Available at https://qims.amegroups.com/article/view/10.21037/qims-22-1297/rc

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://qims.amegroups.com/article/view/10.21037/qims-22-1297/coif). SJ, YD and ZW are employees of VoxelCloud Co. Ltd., which is the sponsor of this clinical trial. The other authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013) and was approved by the ethics committees of Shanghai General Hospital (No. 2018-75), Shanghai East Hospital (No. 2017-052), and Shanghai Shuguang Hospital Affiliated to the Shanghai University of Traditional Chinese Medicine (No. 2018-620-49). Individual consent for this retrospective analysis was waived.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: https://creativecommons.org/licenses/by-nc-nd/4.0/.

© Quantitative Imaging in Medicine and Surgery. All rights reserved.

References

1. Siegel RL, Miller KD, Fuchs HE, et al. Cancer Statistics, 2021. CA Cancer J Clin 2021;71:7-33.
2. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin 2021;71:209-49.
3. Maconachie R, Mercer T, Navani N, et al. Lung cancer: diagnosis and management: summary of updated NICE guidance. BMJ 2019;364:l1049.
4. Rampinelli C, Calloni SF, Minotti M, et al. Spectrum of early lung cancer presentation in low-dose screening CT: a pictorial review. Insights Imaging 2016;7:449-59.
5. Messay T, Hardie RC, Rogers SK. A new computationally efficient CAD system for pulmonary nodule detection in CT imagery. Med Image Anal 2010;14:390-406.
6. Del Ciello A, Franchi P, Contegiacomo A, et al. Missed lung cancer: when, where, and why? Diagn Interv Radiol 2017;23:118-26.
7. Chen H, Huang S, Zeng Q, et al. A retrospective study analyzing missed diagnosis of lung metastases at their early stages on computed tomography. J Thorac Dis 2019;11:3360-8.
8. Singh R, Kalra MK, Homayounieh F, et al. Artificial intelligence-based vessel suppression for detection of sub-solid nodules in lung cancer screening computed tomography. Quant Imaging Med Surg 2021;11:1134-43.
9. Balata H, Quaife SL, Craig C, et al. Early Diagnosis and Lung Cancer Screening. Clin Oncol (R Coll Radiol) 2022;34:708-15.
10. Oren O, Gersh BJ, Bhatt DL. Artificial intelligence in medical imaging: switching from radiographic pathological data to clinically meaningful endpoints. Lancet Digit Health 2020;2:e486-8.
11. Hosny A, Parmar C, Quackenbush J, et al. Artificial intelligence in radiology. Nat Rev Cancer 2018;18:500-10.
12. de Margerie-Mellon C, Chassagnon G. Artificial intelligence: A critical review of applications for lung nodule and lung cancer. Diagn Interv Imaging 2023;104:11-7.
13. Lee CS, Baughman DM, Lee AY. Deep learning is effective for the classification of OCT images of normal versus Age-related Macular Degeneration. Ophthalmol Retina 2017;1:322-7.
14. Codella NCF, Nguyen QB, Pankanti S, Gutman DA, Helba B, Halpern AC, Smith JR. Deep learning ensembles for melanoma recognition in dermoscopy images. IBM J Res Dev 2017. doi: 10.1147/JRD.2017.2708299.
15. Pehrson LM, Nielsen MB, Ammitzbøl Lauridsen C. Automatic Pulmonary Nodule Detection Applying Deep Learning or Machine Learning Algorithms to the LIDC-IDRI Database: A Systematic Review. Diagnostics (Basel) 2019;9:29.
16. Yuan R, Mayo J, Streit I, Atkar-Khattra S, Myers R, Yee J, Lam S. Randomized Clinical Trial with Computer Assisted Diagnosis (CAD) Versus Radiologist as First Reader of Lung Screening LDCT. J Thorac Oncol 2019;14:S287-8.
17. Burki TK. The role of AI in diagnosing lung diseases. Lancet Respir Med 2019;7:1015-6.
18. Park S, Park H, Lee SM, et al. Application of computer-aided diagnosis for Lung-RADS categorization in CT screening for lung cancer: effect on inter-reader agreement. Eur Radiol 2022;32:1054-64.
19. Ueda D, Yamamoto A, Shimazaki A, et al. Artificial intelligence-supported lung cancer detection by multi-institutional readers with multi-vendor chest radiographs: a retrospective clinical validation study. BMC Cancer 2021;21:1120.
20. Shan R, Rezaei T. Lung Cancer Diagnosis Based on an ANN Optimized by Improved TEO Algorithm. Comput Intell Neurosci 2021;2021:6078524.
21. Rampinelli C, Origgi D, Bellomi M. Low-dose CT: technique, reading methods and image interpretation. Cancer Imaging 2013;12:548-56.
22. American College of Radiology. Lung Rads [database on the Internet]. Available online: https://www.acr.org/Clinical-Resources/Reporting-and-Data-Systems/Lung-Rads
23. NCCN Guidelines for Patients: Lung Cancer Screening. 2020. Available online: https://www.nccn.org/patients/guidelines/content/PDF/lung_screening-patient.pdf
24. Lin TY, Dollár P, Girshick R, He K, Hariharan B, Belongie S. Feature pyramid networks for object detection. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 21-26 July 2017; Honolulu, HI, USA. IEEE; 2017.
25. Manickavasagam R, Selvan S, Selvan M. CAD system for lung nodule detection using deep learning with CNN. Med Biol Eng Comput 2022;60:221-8.
26. Nam JG, Hwang EJ, Kim J, et al. AI Improves Nodule Detection on Chest Radiographs in a Health Screening Population: A Randomized Controlled Trial. Radiology 2023;307:e221894.
27. Zhang C, Sun X, Dang K, et al. Toward an Expert Level of Lung Cancer Detection and Classification Using

© Quantitative Imaging in Medicine and Surgery. All rights reserved.

a Deep Convolutional Neural Network. Oncologist 2019;24:1159-65.
28. Li X, Guo F, Zhou Z, Zhang F, Wang Q, Peng Z, Su D, Fan Y, Wang Y. Performance of Deep-learning-based Artificial Intelligence on Detection of Pulmonary Nodules in Chest CT. Zhongguo Fei Ai Za Zhi 2019;22:336-40.
29. Ardila D, Kiraly AP, Bharadwaj S, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nat Med 2019;25:954-61.
30. 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.
31. Li L, Liu Z, Huang H, et al. 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. Thorac Cancer 2019;10:183-92.
32. Gong J, Liu J, Hao W, et al. Computer-aided diagnosis of ground-glass opacity pulmonary nodules using radiomic features analysis. Phys Med Biol 2019;64:135015.
33. Wang X, Li Q, Cai J, et al. Predicting the invasiveness of lung adenocarcinomas appearing as ground-glass nodule on CT scan using multi-task learning and deep radiomics. Transl Lung Cancer Res 2020;9:1397-406.
34. Lv Y, Wei Y, Xu K, et al. 3D deep learning versus the current methods for predicting tumor invasiveness of lung adenocarcinoma based on high-resolution computed tomography images. Front Oncol 2022;12:995870.
35. Baldwin DR, Gustafson J, Pickup L, et al. External validation of a convolutional neural network artificial intelligence tool to predict malignancy in pulmonary nodules. Thorax 2020;75:306-12.
36. Heuvelmans MA, van Ooijen PMA, Ather S, et al. Lung cancer prediction by Deep Learning to identify benign lung nodules. Lung Cancer 2021;154:1-4.
37. Paez R, Kammer MN, Balar A, et al. Longitudinal lung cancer prediction convolutional neural network model improves the classification of indeterminate pulmonary nodules. Sci Rep 2023;13:6157.
38. Ohno Y, Aoyagi K, Yaguchi A, et al. Differentiation of Benign from Malignant Pulmonary Nodules by Using a Convolutional Neural Network to Determine Volume Change at Chest CT. Radiology 2020;296:432-43.
39. Lin X, Jiao H, Pang Z, et al. Lung Cancer and Granuloma Identification Using a Deep Learning Model to Extract 3-Dimensional Radiomics Features in CT Imaging. Clin Lung Cancer 2021;22:e756-66.
40. Zhu Y, Guo YB, Xu D, et al. A computed tomography (CT)-derived radiomics approach for predicting primary co-mutations involving TP53 and epidermal growth factor receptor (EGFR) in patients with advanced lung adenocarcinomas (LUAD). Ann Transl Med 2021;9:545.
41. Jiang M, Yang P, Li J, et al. Computed tomography-based radiomics quantification predicts epidermal growth factor receptor mutation status and efficacy of first-line targeted therapy in lung adenocarcinoma. Front Oncol 2022;12:985284.

Cite this article as: Chen J, Cao R, Jiao S, Dong Y, Wang Z, Zhu H, Luo Q, Zhang L, Wang H, Yin X. Application value of a computer-aided diagnosis and management system for the detection of lung nodules. Quant Imaging Med Surg 2023;13(10):6929-6941. doi: 10.21037/qims-22-1297

© Quantitative Imaging in Medicine and Surgery. All rights reserved.

Appendix 1

Manual detection

For manual detection, each radiologist was given half of the total number of computed tomography (CT) scans to annotate and were required to validate each other's reports independently. After the radiologist found the nodules during the evaluation, the horizontal CT image was scrolled, stopped at the layer with the largest section area of the nodules, and the X-, Y-, and Z-axis coordinates of the center point of the nodules were recorded. Then measuring tools were used to measure the length and diameter of the nodules through the center of the nodules. The nodule type (solid nodule, part-solid nodule, or ground-glass nodule) was then determined. There was no time limit to the interpretation of each CT image, but radiologists were required to read at a speed similar to normal work.

Development of the deep learning algorithm

There were two training data sets, including 635 and 578 patients, respectively. The validation set included 196 patients, and the test set (clinical pretrial data set) included 198 patients. For lung nodule detection, a state-of-the-art feature pyramid network (24) was applied. The voxel intensity ranged from 0 to 1, and the entire CT volume was split into multiple small three-dimensional (3D) patches (size: 128, 128, 128). The 3D patches were fed into the feature pyramid network to output probabilities for various nodule size at different image locations. The lung nodules were detected by removing those nodules with probability below a predefined threshold.

For nodule classification, a dual-channel 3D residual neural network was used. Two patches were extracted from the CT images based on the nodule detection bounding box. The size of the first patch was twice the size of the nodule detection bounding box, while the size of the second patch was fixed to [65, 65, 65]. Such a multi-scale strategy is commonly used for improving the accuracy and robustness of many deep-learning neural networks. The robustness test in the test set included flipping, scaling, rotation and translation, and the results were stable. Both extracted patches were resized to the same size [33, 33, 33] before feeding into the dual-channel 3D residual neural network. The network outputs were the probability scores for the three nodule types (solid nodules, part-solid nodules, and ground-glass nodules).

Statistical analysis

Descriptive data were presented as mean ± standard deviation (SD) if they were normally distributed and median with interquartile range if they were non-normally distributed. The sensitivity for nodule detection and the mean of false-positive nodules per case were calculated (31). Sensitivity was calculated by dividing the sum of true-positive nodules by the sum of nodules in the reference standard. The mean of false-positive nodules per case was derived by dividing the sum of false-positive nodules by the number of CT scans. The correlation index between diameters measured by the computer-aided diagnosis (CAD) system and those measured by the ground truth and the accuracy of nodule classification by CAD system was analyzed. The statistical difference in the mean of false-positive nodules per case was calculated and analyzed using the Wilcoxon rank-sum test (using Pratt's method if there were zero values) (42). In order to evaluate secondary endpoints, Wald 95% confidence interval (CI) was calculated for the accuracy of nodule classification, and the exact Clopper-Pearson 95% CI was calculated for satisfaction degree. All analyses were two-tailed, and the significance level was set at <0.05, if not specified.

References

42. Pratt JW. Remarks on Zeros and Ties in the Wilcoxon Signed Rank Procedures. Journal of the American Statistical Association 1959;54:655-67.

© Quantitative Imaging in Medicine and Surgery. All rights reserved. https://dx.doi.org/10.21037/qims-22-1297

Table S1 Information from computed tomography scans in FAS

Variables	Hospital 1	Hospital 2	Hospital 3	Total
Total slices of images				
N (missing)	334 (0)	334 (0)	334 (0)	1,002 (0)
Mean (SD)	354.60 (111.45)	413.51 (41.19)	349.45 (58.19)	372.52 (81.66)
Min–Max	237.00–979.00	255.00–603.00	247.00–567.00	237.00–979.00
Median	329.50	411.00	331.00	352.50
Q1–Q3	305.00–354.00	393.00–438.00	310.00–361.00	317.00–416.00
Slice thickness (mm)				
N (missing)	334 (0)	334 (0)	334 (0)	1,002 (0)
Mean (SD)	1.15 (0.15)	1.05 (0.23)	1.06 (0.16)	1.09 (0.19)
Min–Max	1.00–1.30	1.00–2.00	1.00–1.50	1.00–2.00
Median	1	1	1	1
Q1–Q3	1.00–1.30	1.00–1.00	1.00–1.00	1.00–1.00
Slice distance (mm)				
N (missing)	334 (0)	334 (0)	334 (0)	1,002 (0)
Mean (SD)	0.99 (0.12)	0.81 (0.05)	0.95 (0.12)	0.92 (0.13)
Min–Max	0.70–1.30	0.80–1.00	0.70–1.00	0.70–1.30
Median	1	0.80	1	1
Q1–Q3	1.00–1.00	0.80–0.80	1.00–1.00	0.80–1.00

FAS, full analysis set; SD, standard deviation; Q1–Q3, 25th–75th percentile.

Table S2 Sensitivity of manual detection by clinical center

Clinical centers	Roles	Working experience (years)	Ground truth, n	Detected nodules, n	Sensitivity	Double-diagnosis
Hospital 1	Radiologist 1	2	1,679	781	46.52%	Half of images were reviewed by radiologist 1 first, and the other half was reviewed by radiologist 2 first
	Radiologist 2	2				
Hospital 2	Radiologist 3	8	2,155	1,066	49.47%	Radiologist 3 read the images first
	Radiologist 4	24				
Hospital 3	Radiologist 5	12	1,804	965	53.49%	Radiologist 5 read the images first
	Radiologist 6	34				

© Quantitative Imaging in Medicine and Surgery. All rights reserved.
https://dx.doi.org/10.21037/qims-22-1297

Table S3 Age- and sex-stratified statistical analysis of primary endpoint in FAS

Variables	CAD system	Manual detection	Difference	P
Male				
N (missing)	504 (0)	504 (0)	–	–
Ground truth, n	3,027	3,027	–	–
Detected nodules, n	2,754	1,405	–	–
Sensitivity (95% CI) (%)	90.98 (89.90, 91.98)	46.42 (44.63, 48.21)	44.57 (42.48, 46.65)	<0.001
False-positive nodules per case, n, mean (SD)	0.28 (0.78)	0.18 (0.48)	0.10 (0.68)	0.03
Female				
N (missing)	498 (0)	498 (0)	–	–
Ground truth, n	2,611	2,611	–	–
Detected nodules, n	2,331	1,407	–	–
Sensitivity (95% CI) (%)	89.28 (88.03, 90.44)	53.89 (51.95, 55.81)	35.39 (33.10, 37.68)	<0.001
False-positive nodules per case, n, mean (SD)	0.32 (0.89)	0.30 (0.83)	0.02 (0.81)	0.91
Age <45 years				
N (missing)	624 (0)	624 (0)	–	–
Ground truth, n	3,022	3,022	–	–
Detected nodules, n	2,710	1,551	–	–
Sensitivity (95% CI) (%)	89.68 (88.54, 90.74)	51.32 (49.52, 53.12)	38.35 (36.23, 40.47)	<0.001
False-positive nodules per case, n, mean (SD)	0.21 (0.57)	0.20 (0.52)	0.01 (0.62)	0.83
Age ≥45 years				
N (missing)	378 (0)	378 (0)	–	–
Ground truth, n	2,616	2,616	–	–
Detected nodules, n	2,375	1,261	–	–
Sensitivity (95% CI) (%)	90.79 (89.61, 91.87)	48.20 (46.27, 50.14)	42.58 (40.33, 44.83)	<0.001
False-positive nodules per case, n, mean (SD)	0.44 (1.13)	0.31 (0.89)	0.13 (0.92)	0.04

FAS, full analysis set; CAD, computer-aided diagnosis; CI, confidence interval; SD, standard deviation; Q1–Q3, 25th–75th percentile.

© Quantitative Imaging in Medicine and Surgery. All rights reserved.
https://dx.doi.org/10.21037/qims-22-1297

Table S4 Sensitivity and true positives of nodule detection using different computed tomography manufacturers in FAS

Manufacturers	CAD system	Manual detection	Difference	P
UIH				
Ground truth, n	2,203	2,203	–	–
Detected nodules, n	1,954	1,131	–	–
Sensitivity (95% CI) (%)	88.70 (87.30, 89.99)	51.33 (49.23, 53.45)	37.36 (34.89, 39.83)	<0.001
False-positive nodules per case, n				0.84
N (missing)	413 (0)	413 (0)	–	
Mean (SD)	0.2688 (0.7022)	0.2615 (0.8031)	0.0073 (0.7156)	
Min–Max	0.00–6.00	0.00–11.00	–9	
Median	0	0	0	
Q1–Q3	0.00–0.00	0.00–0.00	0.00–0.00	
Canon				
Ground truth, n	2,049	2,049	–	–
Detected nodules, n	1,886	1,010	–	–
Sensitivity (95% CI) (%)	92.04 (90.79, 93.18)	49.29 (47.11, 51.48)	42.75 (40.29, 45.21)	<0.001
False-positive nodules per case, n				0.004
N (missing)	316 (0)	316 (0)	–	
Mean (SD)	0.3861 (0.9643)	0.2468 (0.6191)	0.1392 (0.8045)	
Min–Max	0.00–6.00	0.00–6.00	–6	
Median	0	0	0	
Q1–Q3	0.00–0.00	0.00–0.00	0.00–0.00	
Siemens				
Ground truth, n	1,147	1,147	–	–
Detected nodules, n	1,024	560	–	–
Sensitivity (95% CI) (%)	89.28 (87.34, 91.01)	48.82 (45.89, 51.76)	40.45 (37.05, 43.86)	<0.001
False-positive nodules per case, n				0.41
N (missing)	232 (0)	232 (0)	–	
Mean (SD)	0.1638 (0.4914)	0.1853 (0.4791)	–0.0216 (0.5300)	
Min–Max	0.00–3.00	0.00–3.00	–6	
Median	0	0	0	
Q1–Q3	0.00–0.00	0.00–0.00	0.00–0.00	
GE Healthcare				
Ground truth, n	126	126	–	–
Detected nodules, n	120	50	–	–
Sensitivity (95% CI) (%)	95.24 (89.92, 98.23)	39.68 (31.08, 48.78)	35.40 (24.59, 46.20)	<0.001
False-positive nodules per case, n				1.0000
N (missing)	22 (0)	22 (0)	–	
Mean (SD)	0.4545 (0.9117)	0.4091 (0.6661)	0.0455 (0.8985)	
Min–Max	0.00–4.00	0.00–2.00	–4	
Median	0	0	0	
Q1–Q3	0.00–1.00	0.00–1.00	0.00–0.00	
Philips				
Ground truth, n	113	113	–	–
Detected nodules, n	101	61	–	–
Sensitivity (95% CI) (%)	89.38 (82.18, 94.39)	53.98 (44.35, 63.40)	55.56 (46.24, 64.87)	<0.001
False-positive nodules per case, n				0.07
N (missing)	19 (0)	19 (0)	–	
Mean (SD)	0.8947 (2.5797)	0.2105 (0.9177)	0.6842 (1.7337)	
Min–Max	0.00–11.00	0.00–4.00	0.00–7.00	
Median	0	0	0	
Q1–Q3	0.00–0.00	0.00–0.00	0.00–0.00	

FAS, full analysis set; CAD, computer-aided diagnosis; CI, confidence interval; SD, standard deviation; Q1–Q3, 25th–75th percentile.

Table S5 Follow-up of nodules in diameter and density for at least 3 years in 21 patients with 50 nodules

Nodule type	Volume change			Density change		In total
	Increased	Decreased [disappeared]	Unchanged	Increased	Decreased	
Ground glass nodules	0	1 [1]	3	1	0	4
Part-solid nodules	1	4 [1]	11	2	0	16
Solid nodules	2	6 [4]	22	0	3	30

© Quantitative Imaging in Medicine and Surgery. All rights reserved.
https://dx.doi.org/10.21037/qims-22-1297