Egyptian Journal of Radiology and Nuclear Medicine

RESEARCH

Open Access

Assessment of artificial intelligence-aided computed tomography in lung cancer screening

Noha A. Aboelenin[1*], Ahmed Elserafi[1], Noha Zaki[2], Essam A. Rashed[3] and Mohammad al-Shatouri[1]

Abstract

Background Lung cancer is one of the most common causes of cancer-related deaths in developed and developing countries. Therefore, early detection of lung cancer has a significant impact on lung cancer surveillance. Interpretation of lung CT scans for cancer screening is considered an intensive task for most radiologists, and long experience is required for accurate diagnosis through visual processing. This cross-sectional study introduces automated CAD software (Careline Soft's AVIEW Metric software). This software can detect and classify lung nodules in CT scans. The performance of a deep learning (DL) model embedded in that software will be compared with that of the radiologists. Also, the feasibility of lung cancer screening protocol is evaluated in Suez Canal University Hospital, Ismailia, Egypt, by implementing Lung Imaging Reporting and Data System (Lung-RADS).

Results As for the detection of the pulmonary nodules, the initial review by the CAD system (without validation by the researcher radiologist) has high sensitivity (93.0%) and specificity (95.5%) with overall accuracy of 93.6%. After review of the automatically detected nodules by the researcher radiologist was done, the final CAD has higher sensitivity (98.2%) and comparable specificity (95.5%) for the detection of pulmonary nodules with overall accuracy of 97.4%. As for lung cancer screening (categorization of Lung-RADS 3 and 4 nodules), unrevised initial computer-aided detection has 97.9% specificity and 96.9% for lung cancer screening with overall accuracy of 97.4%. After second look and review of the CAD result by the researcher radiologist, there is total agreement in total number of nodules and categorization of Lung-RADS 3 and 4. This gives an excellent agreement of 88.6% ($\kappa = 0.951$) between the CAD system and reference radiologist in the overall categorization of all lung nodules according to Lung-RADS classification.

Conclusions The application of CAD system demonstrated increased sensitivity and specificity for the detection of lung nodules and total agreement in the detection of suspicious and probably benign nodules (lung cancer screening) and excellent level of agreement in the overall lung nodule categorization (Lung-RADS).

Keywords Artificial intelligence, Lung-RADS, Lung cancer screening, Lung nodules, CT scan

Background

Lung cancer is considered one of the most common causes of cancer-related deaths in the world [1]. Lung cancer represents the most lethal malignancy in Egypt, and it is considered the fourth most common cancer. It is more common in men than women, mainly due to differences in tobacco smoking rates up to the best author's knowledge [2].

Till now, Egypt does not have a national lung cancer screening program and most patients usually present with either locally advanced or metastatic disease. Therefore, early detection of lung cancer has a significant impact on lung cancer surveillance [2].

Interpretation of lung CT scans for cancer screening is considered an intensive task for most radiologists, and the assessment of malignancy risk of pulmonary nodules

*Correspondence:
Noha A. Aboelenin
naboelenin@yahoo.com
[1] Faculty of Medicine, Suez Canal University, Ismailia, Egypt
[2] Faculty of Medicine, Ain Shams University, Cairo, Egypt
[3] Graduate School of Information Science, University of Hyogo, Kobe, Japan

is still challenging as long experience is required for accurate diagnosis through visual processing [3].

In 2014, the American College of Radiology (ACR) published the Lung Imaging Reporting and Data System (Lung-RADS) categories to standardize the CT lung screening reporting and management recommendations and facilitates outcome monitoring [4]. Lung-RADS contains five categories to differentiate risk level of nodules using nodule type, nodule size and growth as classification criteria [4].

Computer-aided detection (CAD) systems help in detecting various diseases in their early stages. It has been reported that only 68% of the lung cancer nodules are correctly diagnosed by one radiologist, and this percentage increases up to 82% with two radiologists. The early detection of lung cancer nodules is very difficult and time-consuming task for most radiologists. Also, screening a lot of scans with care requires plenty of time; meanwhile, it is very much error-prone in the detection of small nodules [5]. Therefore, an automatic diagnosis tool is needed to help radiologists reducing the reading time, detecting the missed nodules and allowing better localization. Nowadays, the latest generation of CAD systems can help in the screening process by categorizing the nodules into benign and malignant. With the recent advances in image analysis, CAD systems outperform expert radiologists in nodule detection and localization. However, the differentiation between benign and malignant nodules is still a challenging issue due to the very close resemblance at early stages [6].

This study introduces an automated system/software (AVIEW Metric software) developed by Coreline Soft, i.e., Coreline soft is a South Korean medical image software company. This software can detect and classify lung nodules in CT scans. The performance of a deep learning (DL) model embedded in that software will be compared with that of the radiologists. Also, the feasibility of lung cancer screening protocol is evaluated in Suez Canal University Hospital, Ismailia, Egypt, by implementing Lung Imaging Reporting and Data System (Lung-RADS).

Methods

A cross-sectional study was conducted at CT unit, Radiology Department, Suez Canal University Hospitals in Ismailia with online remote access to a computer-aided detection system (Coreline Soft's AVIEW Metrics).

A total of 79 CT scans were collected retrospectively through Picture Achieving and Communication System (PACS). All cases aged more than 40 years and had lung nodules (< 3 cm) on their CT scans. We excluded CT scans if not all lung lobes were fully visible in the field of view (e.g., over-sized subjects) or if the images had clear visible motion artifacts.

The CT scans were automatically reviewed using Coreline Soft's AVIEW Metric software, i.e., AVIEW Metric is artificial intelligence software that uses chest CT scan for automatically screening for lung nodules/cancer based on Lung-RADS.

The CAD software provides lung lobe segmentation (Fig. 1), then an initial automated labeled result of the nodules number, size and their Lung-RADS categorization (Figs. 2 and 3). These primary results are then reviewed by the researcher radiologist who can confirm the automatically detected nodules, add other missed nodules (by clicking on the suspected nodules in the scan, then the system can automatically calculate the size and Lung-RADS of the nodules) and remove nodules that are falsely diagnosed by the CAD system. The researcher radiologist also can report any special features of the nodule (such as fat or calcium) which can dramatically change the Lung-RADS categorization (Fig. 4).

Then, a final report containing final nodule count and Lung-RADS classification was given. It is worth noting that the software does not provide a final automated report unless all the automatically detected nodules were reviewed (Fig. 5).

Two radiologists (radiologists A and B), each with approximately 10 years of experience were reviewing the chest CT images, independently. Both radiologists were blind of each other's and the CAD's results. The CT studies were reviewed by using RadiAnt DICOM Viewer (version 4.2.0; Medixant, Oznan, Poland). Window level and window width were typically set at 2600 and 1500 HU, respectively.

The CT scans were reviewed by a third reference radiologist (radiologist C) who had approximately 15 years of experience and acted as the gold standard in our study.

The results of CAD system, radiologists A and B were compared to the reference radiologist.

CT imaging is performed using a 16-slice scanner, Activion 16 model TSX-031A-2012 with standard accessories (Toshiba Medical Systems) installed in Suez Canal University Hospital with regular maintenance and calibration.

CT scan is routinely performed as follows:

- In cranio-caudal direction.
- With breath-holding manner.
- Starting from the apices of the lung to lateral costophrenic sulci.
- Slice thickness = 1 mm.
- 120 kV, and 50–100 mAs.

Patients with breathing difficulties were trained, and the scan was performed after breath hold practice.

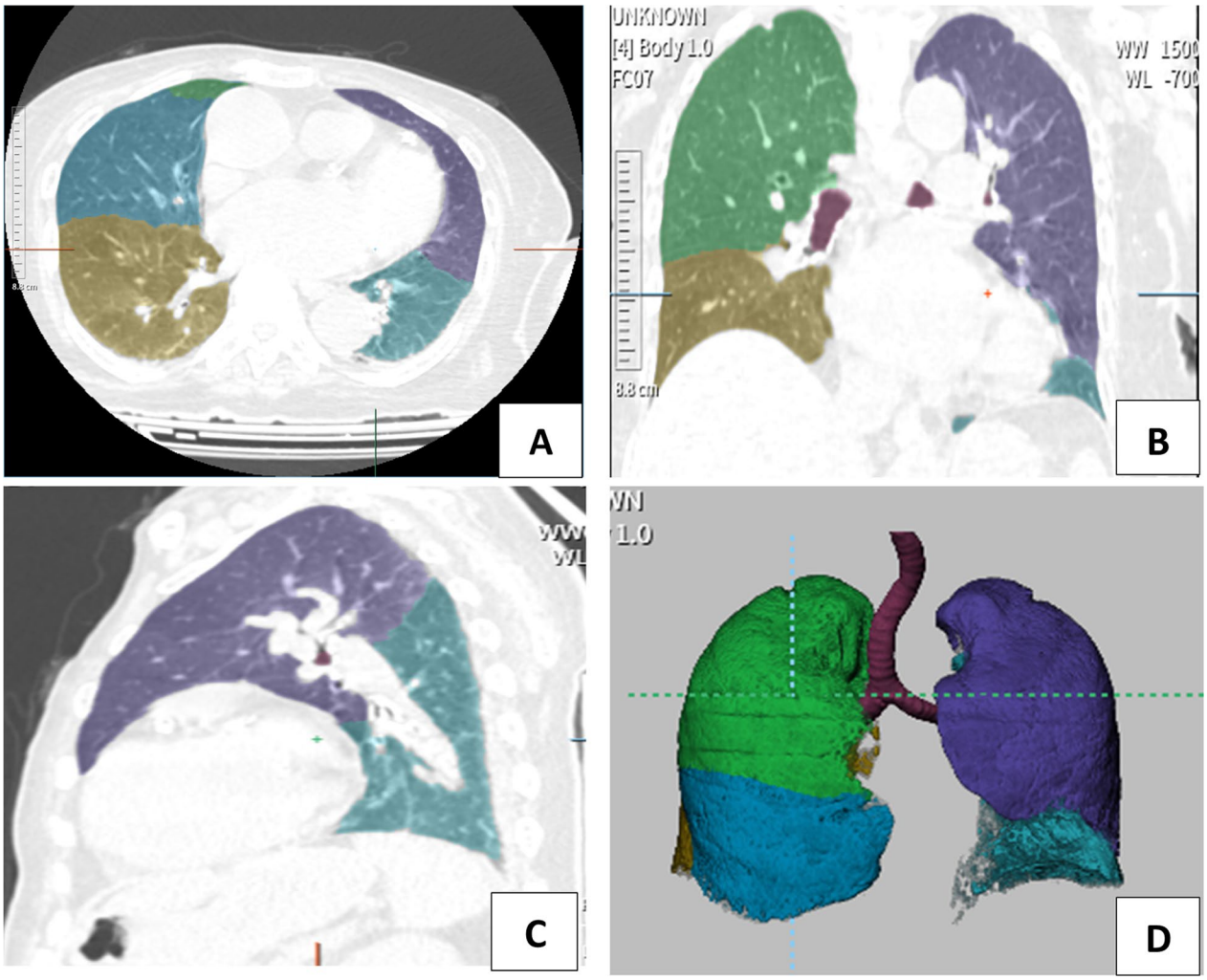


Fig. 1 Automatic lung segmentation in 2D axial (**A**), coronal (**B**) and sagittal planes (**C**) with 3D reconstruction (**D**)

Post-processing and image analysis

First, the researcher attended weekly online sessions of hand-on training arranged by Coreline Soft for 4 weeks, i.e., Coreline soft is a Korean medical image software company that developed AVIEW Metric-Lung which provides various quantitative analysis reports, such as LAA/ LAA size analysis, air-trapping analysis, airway measurement, lung vessel analysis and automatic lung nodule detection and characterization, through easy-to-use automatic pre-processing steps.

Then, the software was tested on few numbers of cases as pilot study.

Image analysis was done using AVIEW software by the following steps:

(A) Precise nodule detection and analysis:

Provide detection of solid, part-solid and ground glass nodules and calculate each nodule's volume and diameter (fully automatic nodule detection by CAD)

(B) Automatic calculation of Lung-RADS score:
Standardize lung cancer screening reports using integrated Lung-RADS and provide PDF reports including nodule analysis results and the Lung-RADS score.

Statistical analysis

Data were collected, coded and then entered as a spread sheet using Microsoft Excel 2010 for Windows, of the Microsoft Office bundle, 2010 of Microsoft Corporation, USA. Data were analyzed using IBM Statistical Package for Social Sciences software (SPSS), 21st

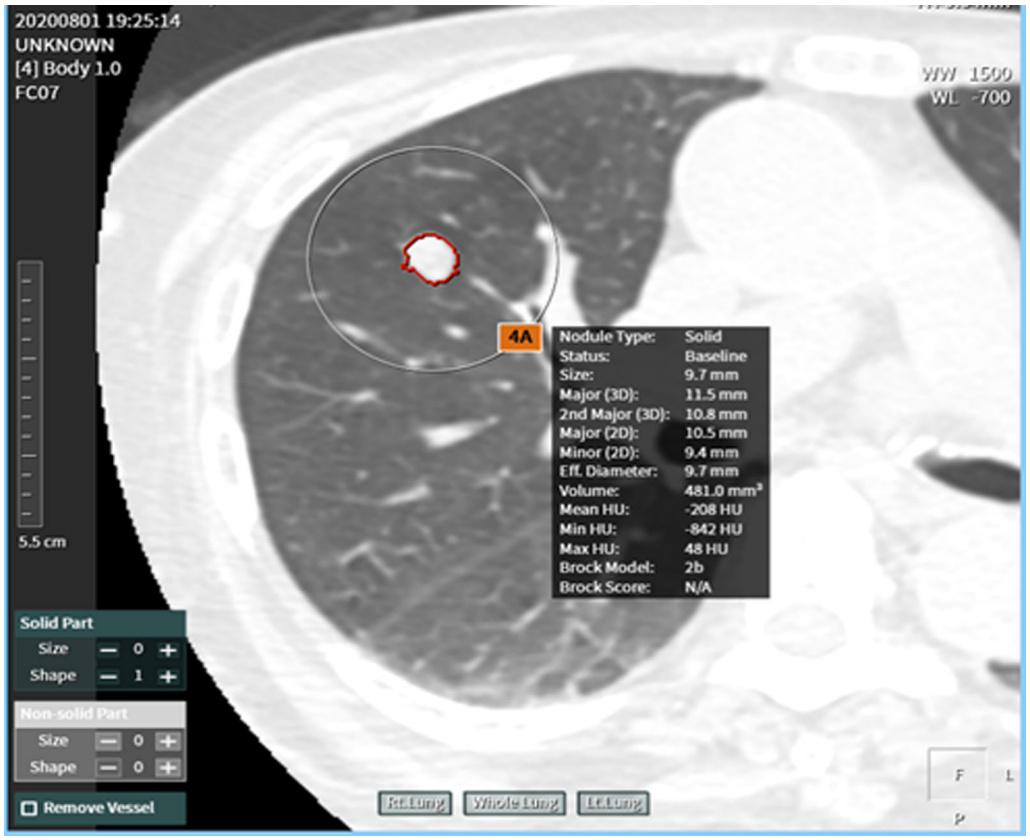


Fig. 2 Axial CT scan showing solid pulmonary nodule, automatically detected, measured, and classified (Lung-RADS 4A)

edition, IBM, USA. The results of CAD of lung nodules were compared with two expert radiologists.

The predictive values were calculated by obtaining positive predictive values (PPV), negative predictive values (NPV), sensitivity, specificity and total accuracy of CAD system.

Continuous data were expressed as mean ± standard deviation and categorical data as percentage. Data were presented as tables and graphs, and t-test was used to compare between two groups' quantitative data expressed as mean and standard deviation. For comparisons in between more than two groups, analysis of variance (ANOVA) was used.

Chi-squared or Fisher's exact tests were used to compare between the qualitative data expressed as number and percentage, wherever compatible. Correlation (Spearman and Pearson) was used to identify relations between data.

Any other kind of test was performed when appropriate. The results were considered statistically significant at a *p*-value of less than or equal 0.05.

Results

This cross-sectional study included 79 adults (aged more than 40 years) who have lung nodules (< 3 cm) on their CT scans, acquired in the CT unit of Suez Canal University Hospitals. Patients had mean age of (50.9 ± 8.5) years ranging from 40 to 82 years, and 53.2% of them were females. Patients had mean weight of (78.9 ± 8.5) kg, mean length of (1.69 ± 0.12) meter and mean BMI of (27.6 ± 3.4) kg/m^2. Most of patients were non-smokers (57%), 11.4% were ex-smokers and 31.6% were current smokers.

As shown in Table 1, total nodules had median number of 1 ranging from 0 to 15 nodules. They were detected in 57 (72.2%) patients: 29 (36.7%) out of them were ≤ 6 mm and 28 (35.5%) > 6 mm.

Regarding solid nodules, they had median number of 1 ranging from 0 to 13 nodules as they are present in 37 (46.9%) patients: 19 (24.1%) out of them are ≤ 6 mm and 18 (22.8%) > 6 mm.

Regarding subsolid nodules, they had median number of 1 ranging from 0 to 2 nodules as they are present in 25

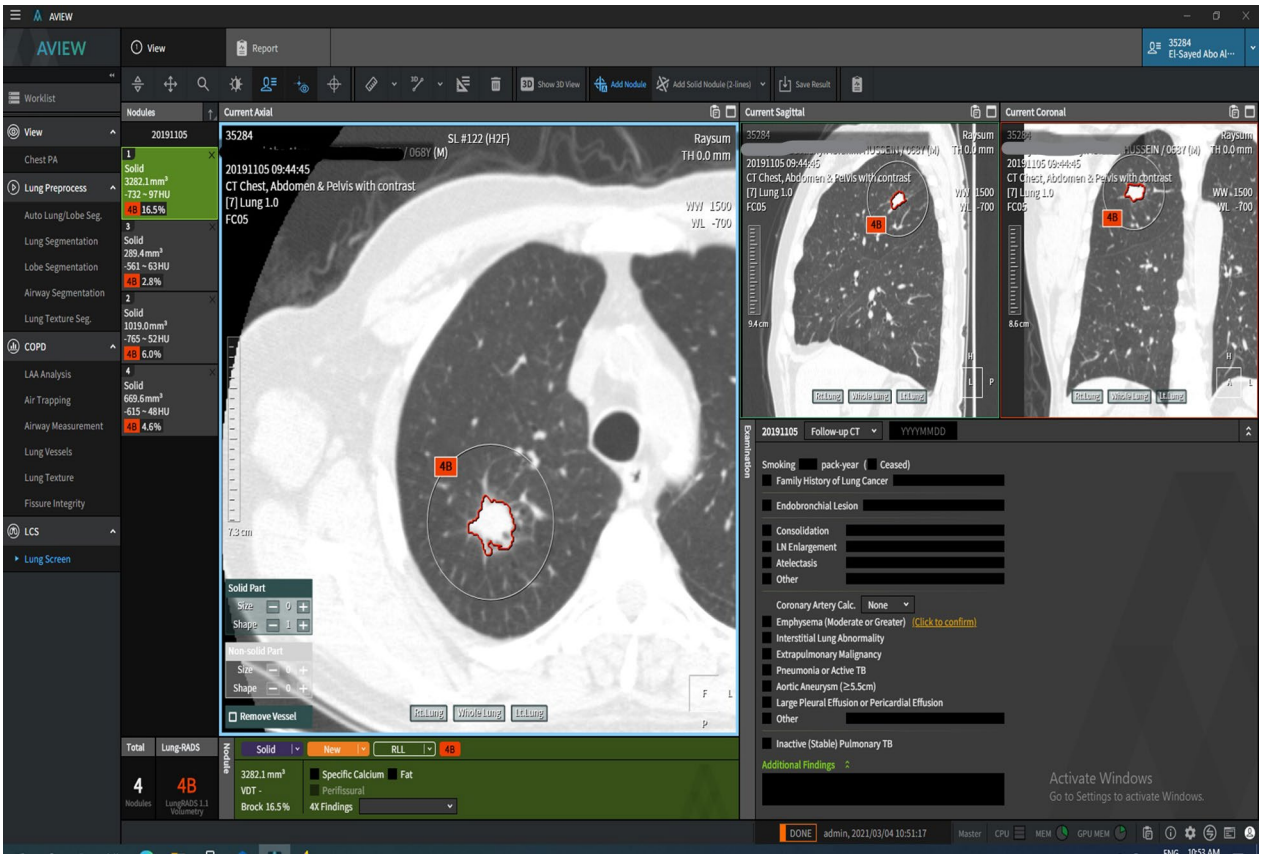


Fig. 3 Four pulmonary nodules were automatically detected by the CAD system and confirmed by the researcher radiologist

(31.6%) patients: 14 (17.7%) out of them are ≤ 6 mm and 11 (13.9%) > 6 mm.

Regarding the calcified nodules, they had median number of 1 ranging from 0 to 1 nodule, as it is present in 1 (1.3%) patient, with size ≤ 6 mm.

Regarding the characteristics of the largest nodule, total nodules had mean size of (15.4 ± 16.8) ranging from 0 to 60, 16 (20.3%) in the left lobe of the lung and 41 (51.9%) in the right lobe of the lung.

Regarding the shape, 1 (1.3%) had irregular shape, 9 (11.4%) had oval shape, 37 (46.8%) had rounded shape and 10 (12.6%) had speculated shape.

As for the Lung-RADS classification, 35 (44.3%) cases were categorized as Lung-RADS 1, 13 (16.5%) Lung-RADS 2, 8 (10.1%) Lung-RADS 3, 5 (6.3%) Lung-RADS 4a, 14 (17.7%) Lung-RADS 4b and 4 (5.1%) Lung-RADS 4x.

As shown in Table 2, there was significant difference between radiologists A, B and CAD system in the detection of the nodules. The difference is due to the use of the CAD system.

Although there were differences between three methods regarding the detection, measuring the size and counting the number of solid nodules, but these differences were statistically insignificant as $p > 0.05$.

While there were statistically significant differences between three methods regarding the detection, counting the number and measuring the size of subsolid nodules ($p = 0.005$, 0.002 and 0.039, respectively).

Regarding calcified nodules, three methods showed total agreement regarding the detection, counting the number and measuring the size of calcified nodules ($p = 1.00$).

There was significant difference between radiologists A, B and CAD system regarding the size and site of the largest nodule. The difference is due to CAD system ($p = 0.001$ and < 0.001, respectively) (Table 3).

Although there were differences between the three methods regarding the shape and Lung-RADS categorization of the largest nodule, the differences were statistically insignificant as $p > 0.05$.

The final CAD system and the reference radiologist had total agreement regarding the detection of total, solid, subsolid and calcified nodules as the differences were statistically insignificant as $p > 0.05$ (Table 4).

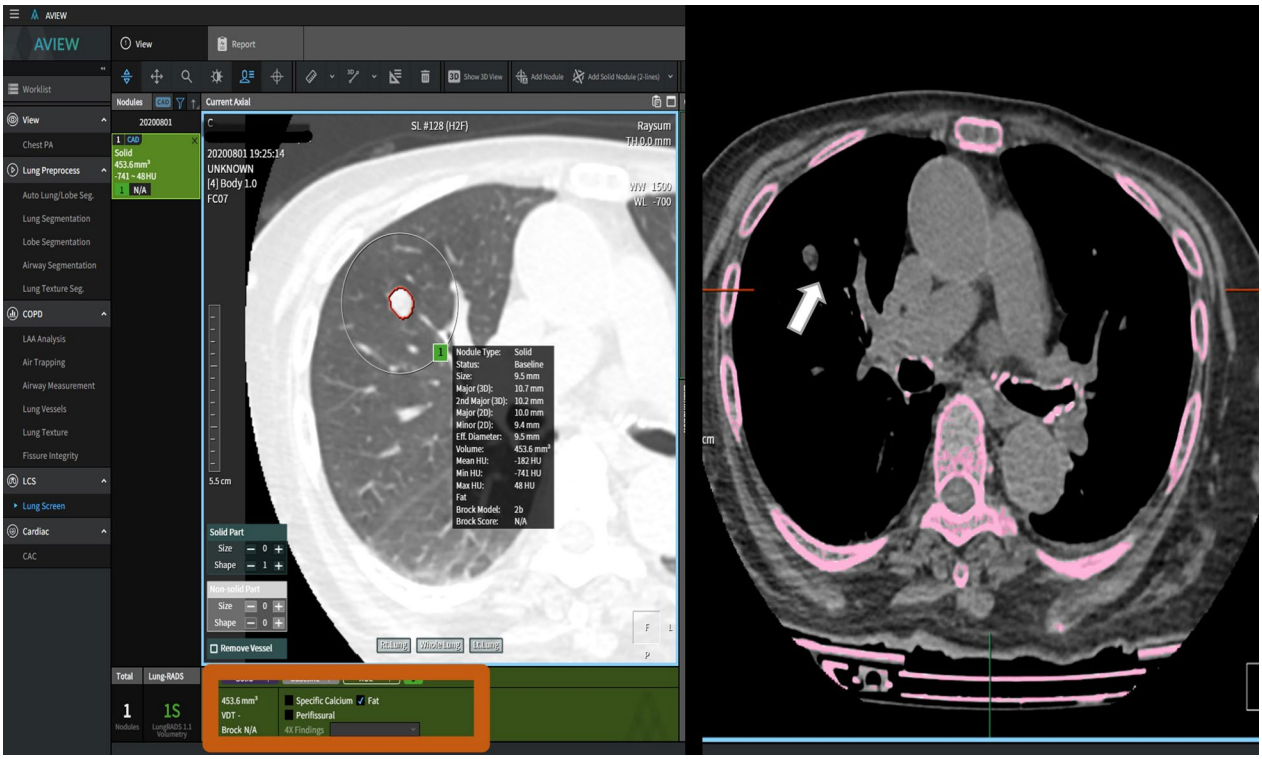


Fig. 4 Final CAD system categorization of the solid nodule into Lung-RADS 2 after adding the fat as special feature of the nodule (Right: lung window, Left: mediastinum window)

There was significant difference between CAD system and the reference radiologist regarding size of the largest nodules ($p = 0.018$). Although there were differences in Lung-RADS categorization of the largest nodules, the differences were statistically insignificant as ($p = 0.592$) (Table 5).

Radiologist A had the lowest prediction of the presence of total nodules (70.2%), solid nodules (92.5%) and subsolid nodules (40%), while CAD system had the highest prediction of the presence of total nodules (97.4%), solid nodules (78.8%) and subsolid nodules (100%). All had 100% prediction of calcified nodule (Fig. 6).

Unrevised primary CAD system has high sensitivity (93.0%) and specificity (95.5%) for the detection of pulmonary nodules, 98.1% positive predictive value and 84.0% negative predictive value. The overall accuracy of the CAD system is 93.6%. Only 7% of cases were false negative and 4.5% were false positive.

Revised final CAD system has high sensitivity (98.2%) and specificity (95.5%) for the detection of pulmonary nodules, 98.1% positive predictive value and 98.2% negative predictive value. The overall accuracy of the CAD system is 97.4%. Only 1.8% of cases were false negative and 4.5% were false positive.

We assumed that Lung-RADS categories 1 and 2 are benign nodules having the same follow-up recommendations (− ve test result) and categories 3 and 4 are malignant nodules requiring further diagnostic work up (+ve test result).

Unrevised primary CAD system has 97.9% specificity and 96.9% for lung cancer screening. This discrepancy was seen in two cases. The first case was primarily diagnosed as negative case despite the presence of large malignant mass about 4 cm.

The second case was primarily diagnosed as malignant nodule based on the size (15 mm) despite the presence of fat component.

The overall accuracy of the initial unrevised CAD system in lung cancer screening (malignant nodule detection) is 97.4%. Only 3.2% of cases were false negative and 2.1% of the cases were false positive.

There is excellent 88.6% agreement (kappa 0.951) between the CAD system and reference radiologist in the categorization of lung nodules according to Lung-RADS classification.

There is good 96.2% (kappa 0.657) interobserver agreement between the radiologist A and B readings without CAD system in the categorization of lung nodules according to Lung-RADS classification.

Exam Date	2019/01/03	Slice Thickness	1.00 mm	Model	TOSHIBA/Activion16	Kernel	FC05	CTDIvol	12.40 mGy

Clinical Info.

Smoking History	- pack-year	Weight	- kg	Height	- cm	Family History of Lung Cancer	-

Lung-RADS 4B

4B - Very Suspicious

Findings for which additional diagnostic testing and/or tissue sampling is recommended.
Chest CT with or without contrast, PET/CT and/or tissue sampling depending on the probability of malignancy and comorbidities. PET/CT may be used when there is a >=8mm solid component.

Total Nodules: 12

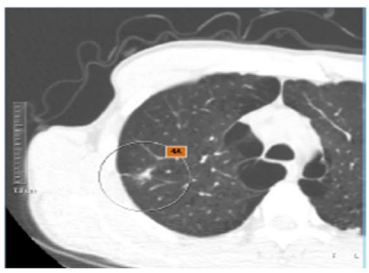
Nodule #1

	Baseline CT
Location	RUL
Image Number	#178
Type	Solid
Status	Baseline
Diameter	27.2 mm
Effective Diameter	27.2 mm
Volume	10533.1 mm^3
Mass	N.A.
4X Findings	
Lung-RADS	4B - Very Suspicious

Nodule #3

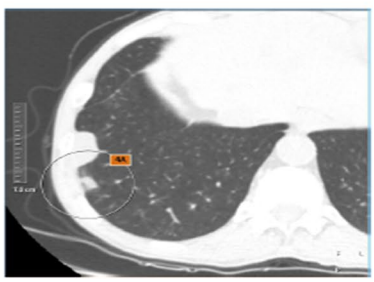
	Baseline CT
Location	RUL
Image Number	#160
Type	Part Solid
Status	Baseline
Diameter	17.8 mm (7.1 mm)
Effective Diameter	17.8 mm
Volume	2948.2 mm^3 (186.1 mm^3)
Mass	1.30 g
4X Findings	
Lung-RADS	4A - Suspicious

Nodule #4



	Baseline CT
Location	RUL
Image Number	#116
Type	Part Solid
Status	Baseline
Diameter	12.9 mm (6.1 mm)
Effective Diameter	12.9 mm
Volume	1111.3 mm^3 (116.7 mm^3)
Mass	0.53 g
4X Findings	
Lung-RADS	4A - Suspicious

Nodule #5



	Baseline CT
Location	RLL
Image Number	#332
Type	Solid
Status	Baseline
Diameter	11.0 mm
Effective Diameter	11.0 mm
Volume	690.4 mm^3
Mass	N.A.
4X Findings	
Lung-RADS	4A - Suspicious

Fig. 5 Automated final lung cancer screening report as generated by the CAD system

Table 1 Frequency of nodules size and type as described by the reference radiologist ($N=79$)

Variables	Reference radiologist
No. of total nodules	
Median (range)	1 (0–15)
No. of total nodules (n, %)	
Absent	22 (27.8%)
Present	57 (72.2%)
Size of total nodules	
≤ 6 mm	29 (36.7%)
> 6 mm	28 (35.5%)
Solid nodules	
No. of solid nodules	
Median (range)	1 (0–13)
No. of solid nodules (n, %)	
Absent	42 (53.1%)
Present	37 (46.9%)
Size of solid nodules	
≤ 6 mm	19 (24.1%)
> 6 mm	18 (22.8%)
Subsolid nodules	
No. of subsolid nodules	
Median (range)	1 (0–2)
No. of subsolid nodules (n, %)	
Absent	54 (68.4%)
Present	25 (31.6%)
Size of subsolid nodules	
≤ 6 mm	14 (17.7%)
> 6 mm	11 (13.9%)
Calcified nodules	
No. of calcified nodules	
Median (range)	1 (0–1)
No. of calcified nodules (n, %)	
Absent	78 (98.7%)
Present	1 (1.3%)
Size of calcified nodules	
≤ 6 mm	1 (1.3%)
> 6 mm	0 (0%)
Characteristics of largest nodule	
Size	
Mean ± SD	15.4 ± 16.8
Median (range)	8 (1–60)
Site	
Left side	16 (20.3%)
Upper lobe	12 (15.2%)
Lower lobe	4 (5.1%)
Right side	41 (51.9%)
Upper lobe	23 (29.1%)
Middle lobe	3 (3.8%)
Lower lobe	15 (19%)
Shape	
Irregular	1 (1.3%)
Oval	9 (11.4%)
Rounded	37 (46.8%)
Speculated	10 (12.6%)
Special feature	
Calcium	1 (1.3%)
Fat content	3 (3.8%)
Lung-RADS	
1	35 (44.3%)
2	13 (16.5%)
3	8 (10.1%)
4a	5 (6.3%)
4b	14 (17.7%)
4x	4 (5.1%)

Discussion

The application of CAD to CT screening is introduced in the terms of error reduction, time saving and work efficiency. Up to our knowledge, only few studies have reported the accuracy, sensitivity and specificity of CAD and level of agreement in Lung-RADS-based assessments [7].

There are several commercial CAD systems that provide diagnosis assistance. This study introduces an evaluation study of using automated system/software (AVIEW Metric software) developed by Coreline Soft.

A total number of 253 nodules were detected in our study with 190 nodules that were primarily detected by the CAD system. Then, 14 nodules were excluded and 77 nodules were added by the researcher radiologist.

There was significant difference between radiologist A, radiologist B and CAD system in detection of pulmonary nodules, with significant improvement in the detection of the pulmonary nodules after using the CAD method. This improvement is observed regardless the size and consistency of the nodule. Park et al. also indicate that CAD may provide readers with additional numbers of nodules to assess. CAD should be able to characterize nodules beyond their size, in which case it may be able to suggest more clinically relevant nodules [7].

In our study, there is good 96.2% ($\kappa=0.657$) interobserver agreement between the radiologist A and B readings without CAD system in the categorization of lung nodules according to Lung-RADS classification. This percentage is higher that the percentage demonstrated in park et al. which showed moderate agreement (Fleiss kappa, 0.60 [95% CI 0.57, 0.63]) [7]. These could be because the latter study used five observers instead of the two observers in our study. Our interobserver agreement was also higher than that described in Jacobs et al.'s 2021

Table 2 Frequency of nodules type and size as described by radiologist A, radiologist B and CAD system ($N=79$)

Variables	Radiologist A	Radiologist B	CAD system	*p* value
No. of total nodules				
Median (range)	1 (0–10)	1 (0–13)	1 (0–100)	0.217[1]
Total nodules				
Absent	39 (49.4%)	35 (44.3%)	22 (27.8%)	< 0.001*[b]
Present	40 (50.6%)	44 (55.7%)	57 (72.2%)	
Size of total nodules				
≤ 6 mm	19 (24.1%)	23 (29.1%)	29 (36.7%)	
> 6 mm	21 (26.5%)	21 (26.6%)	28 (35.5%)	0.108[b]
Solid nodules				
No. of solid nodules				
Median (range)	1 (0–10)	0 (0–13)	1 (0–91)	0.323[a]
Solid nodules				
Absent	39 (49.4%)	42 (44.3%)	32 (40.5%)	0.105[b]
Present	40 (50.6%)	37 (55.7%)	47 (59.5%)	
Size of solid nodules				
≤ 6 mm	18 (22.8%)	20 (30.3%)	23 (30.4%)	0.332[a]
> 6 mm	22 (27.8%)	17 (25.4%)	24 (29.1%)	
Subsolid nodules				
No. of subsolid nodules				
Median (range)	0 (0–1)	0 (0–2)	0 (0–9)	0.005*[a]
Subsolid nodules (n, %)				
Absent	69 (87.4%)	62 (78.5%)	54 (68.4%)	0.002*[b]
Present	10 (12.6%)	17 (21.5%)	25 (31.6%)	
Size of subsolid nodules				
≤ 6 mm	10 (12.6%)	11 (13.9%)	14 (17.7%)	
> 6 mm	0 (0%)	6 (7.6%)	11 (13.9%)	0.039*[b]
Calcified nodules				
No. of calcified nodules				
Median (range)	0 (0–1)	0 (0–1)	0 (0–1)	1.00[a]
Calcified nodules (n, %)				
Absent	78 (98.7%)	78 (98.7%)	78 (98.7%)	1.00[c]
Present	1 (1.3%)	1 (1.3%)	1 (1.3%)	
Size of calcified nodules				
≤ 6 mm	1 (1.3%)	1 (1.3%)	1 (1.3%)	1.00[c]
> 6 mm	0 (0%)	0 (0%)	0 (0%)	

*Statistically significant as $p < 0.05$

[a] Kruskal–Wallis test

[b] Chi-square test

[c] Fisher exact test

study. Their interobserver agreement was moderate for the standard viewer with Fleiss kappa values of 0.58 (95% CI 0.55, 0.60).

Park et al. also suggest automatic measurement as a possible benefit of CAD for improving Lung-RADS agreement [7].

Regarding the characterization of the largest detected pulmonary nodule, there was significant difference between radiologists A, B and the CAD system regarding the size and site of the largest nodules, with the CAD system that provides a more accurate result regarding the size and site owing to the automated segmentation of the lung ($p=0.001$ and <0.001, respectively).

In Park et al.'s study, there was no reference standard such as follow-up results with which to make comparisons and their study focused on inter-reader agreement, so a reference standard was not required [7].

Table 3 Characteristics of the largest nodule as diagnosed by radiologist A, radiologist B and CAD system ($N=79$)

Variables	Radiologist A	Radiologist B	CAD system	*p* value
Size				
Mean ± SD	10.3 ± 16.1	10.1 ± 14.5	11.8 ± 16.6	0.001*[a]
Median (range)	3.5 (1–57)	4 (1–60)	7 (1–63)	
Site				
Left side	11 (13.9%)	11 (13.9%)	16 (20.3%)	< 0.001*[b]
Upper lobe	7 (8.9%)	7 (8.8%)	12 (15.2%)	
Lower lobe	4 (5.1%)	4 (5.1%)	4 (5.1%)	
Right side	29 (36.7%)	33 (41.7%)	41 (51.9%)	
Upper lobe	28 (35.4%)	21 (29.1%)	23 (29.1%)	
Middle lobe	0 (0%)	1 (1.3%)	3 (3.8%)	
Lower lobe	1 (1.3%)	1 (1.3%)	15 (19%)	
Shape				
Irregular	0 (0%)	1 (1.3%)	1 (1.3%)	
Oval	7 (8.8%)	7 (8.9%)	9 (11.4%)	0.108[b]
Rounded	23 (29.1%)	25 (31.6%)	33 (41.8%)	
Speculated	10 (12.6%)	11 (13.9%)	14 (17.7%)	
Special feature				
Calcium	1 (1.3%)	1 (1.3%)	1 (1.3%)	0.323[a]
Fat content	3 (3.8%)	3 (3.8%)	3 (3.8%)	
Lung-RADS				
1	42 (53.2%)	38 (48.1%)	26 (32.9%)	0.105[b]
2	7 (8.8%)	10 (12.6%)	22 (27.8%)	
3	8 (10.1%)	8 (10.1%)	8 (10.1%)	
4a	4 (5.1%)	5 (6.3%)	5 (6.3%)	
4b	14 (17.7%)	14 (17.7%)	14 (17.7%)	
4x	4 (5.1%)	4 (5.1%)	4 (5.1%)	

*Statistically significant as $p < 0.05$

[a] Kruskal–Wallis test

[b] Fisher exact test

As for the detection of the pulmonary nodules, the initial review by the CAD system (unrevised by the researcher radiologist) has high sensitivity (93.0%) and specificity (95.5%), 98.1% positive predictive value and 84.0% negative predictive value with overall accuracy of 93.6%. However, all cases must by revised and reviewed by the radiologist after being automatically reviewed to confirm or exclude false positive nodules which could be due to atelectasis or even add nodules which couldn't be detected, so this CAD tool is only an aiding tool to help the radiologist save time and increase the efficiency of the detection.

The CAD system we were using always gives a warning message that the operator must review all the automatically detected nodules before the final report is generated.

Table 4 Comparison between nodules type and size as described by final CAD system and the reference radiologist ($N=79$)

Variables	Final CAD system	Reference radiologist	*p* value
No. of total nodules			
Median (range)	1 (0–100)	1 (0–15)	0.852[a]
Total nodules			
Absent	22 (27.8%)	22 (27.8%)	1.00[b]
Present	57 (72.2%)	57 (72.2%)	
Size of total nodules			
≤ 6 mm	30 (37.9%)	29 (36.7%)	
> 6 mm	28 (35.5%)	28 (35.5%)	0.954[b]
No. of solid nodules			
Median (range)	1 (0–91)	1 (0–13)	0.773[a]
Solid nodules			
Absent	32 (40.5%)	42 (53.1%)	0.782[b]
Present	47 (59.5%)	37 (46.9%)	
Size of solid nodules			
≤ 6 mm	23 (30.4%)	19 (24.1%)	0.817[a]
> 6 mm	24 (29.1%)	18 (22.8%)	
No. of subsolid nodules			
Median (range)	0 (0–9)	1 (0–2)	0.978[a]
Subsolid nodules (n, %)			
Absent	54 (68.4%)	54 (68.4%)	1.00[b]
Present	25 (31.6%)	25 (31.6%)	
Size of subsolid nodules			
≤ 6 mm	14 (17.7%)	14 (17.7%)	1.00[b]
> 6 mm	11 (13.9%)	11 (13.9%)	
No. of calcified nodules			
Median (range)	0 (0–1)	1 (0–1)	0.993[a]
Calcified nodules (n, %)			
Absent	78 (98.7%)	78 (98.7%)	1.00[c]
Present	1 (1.3%)	1 (1.3%)	
Size of calcified nodules			
≤ 6 mm	1 (1.3%)	1 (1.3%)	1.00[c]
> 6 mm	0 (0%)	0 (0%)	

[a] Mann–Whitney test

[b] Chi–square test

[c] Fisher exact test

Atelectasis and infection were commonly misidentified as nodules likely due to their relative mass-like area of hyper-density with adjacent normal or emphysematous lung parenchyma [8].

The lobular contour of the extra pleural fat and protruding osteophytes from thoracic vertebral bodies, in

Table 5 Characteristics of the largest nodule as described by final CAD system and reference radiologist ($N=79$)

Variables	Final CAD system	Reference radiologist	*p* value
Size			
Mean ± SD	11.8 ± 16.6	15.4 ± 16.8	0.018*[a]
Median (range)	7 (1–63)	8 (1–60)	
Site			
Left side	16 (20.3%)	16 (20.3%)	1.00[b]
Upper lobe	12 (15.2%)	12 (15.2%)	
Lower lobe	4 (5.1%)	4 (5.1%)	
Right side	41 (51.9%)	41 (51.9%)	
Upper lobe	23 (29.1%)	23 (29.1%)	
Middle lobe	3 (3.8%)	3 (3.8%)	
Lower lobe	15 (19%)	15 (19%)	
Shape			
Irregular	1 (1.3%)	1 (1.3%)	
Oval	9 (11.4%)	9 (11.4%)	0.984[2]
Rounded	33 (41.8%)	37 (46.8%)	
Speculated	14 (17.7%)	10 (12.6%)	
Special feature			
Calcium	1 (1.3%)	1 (1.3%)	1.00[2]
Fat content	3 (3.8%)	3 (3.8%)	
Lung-RADS			
1	26 (32.9%)	35 (44.3%)	0.592[2]
2	22 (27.8%)	13 (16.5%)	
3	8 (10.1%)	8 (10.1%)	
4a	5 (6.3%)	5 (6.3%)	
4b	14 (17.7%)	14 (17.7%)	
4x	4 (5.1%)	4 (5.1%)	

*Statistically significant as $p<0.05$

[a] Mann–Whitney test

[b] Fisher exact test

direct contact with the lung parenchyma, likely led to their misidentification as nodules. While future study is necessary to compensate for the presence of these coincident findings, quantification of rates of known false positives may be useful when using the AI software as an adjunct tool for diagnosis.

The CAD software automatically presented the Lung-RADS category after readers determined the nodule selection, nodule type and measurement. Van Riel et al. found that wrong assignment of the Lung-RADS category was not uncommon (6% of all readings) [3].

Park et al. think that supporting function of the CAD system will help radiologists to be more efficient and accurate in Lung-RADS categorization [7].

After second look and revision of the automated detected nodules was done, the revised final computer-aided detection has higher sensitivity (98.2%) and comparable specificity (95.5%) for the detection of pulmonary nodules, comparable 98.1% positive predictive value and higher 98.2% negative predictive value with higher overall accuracy (97.4%).

Our sensitivity is slightly lower than the sensitivity described in Chamberlin et al. 2021 (sensitivity = 100%), while our specificity was higher than theirs (specificity = 70%) [8].

As for lung cancer screening (categorization of Lung-RADS 3 and 4 nodules), unrevised primary computer-aided detection has 97.9% specificity and 96.9% for lung cancer screening. There was discrepancy in two cases, the first one was primarily diagnosed as negative case despite the presence of large malignant mass about 4 cm and that is due to the inability of the CAD system to automatically detect lesions more than 3 cm and the lesion was manually added by the reviewing radiologist and Lung-RADS 4b was given to the case for further histopathology and was confirmed to be malignant mass. The second case was primarily diagnosed as suspicious nodule about 11 mm in size despite the presence of fat component making it a definite benign pulmonary hamartoma (Fig. 4). And that is because the introduced CAD system cannot automatically detect the special features such as calcium and fat which must be manually added/marked under the special features icon by the operator/researcher radiologist. That discrepancy was not present in the other two fat containing lesions as their sizes were ≤ 6 mm, making them benign based on their size.

After second look and review of the CAD result by the researcher radiologist, there is total agreement in total number of nodules and categorization of Lung-RADS 3 and 4. Park et al. think that supporting function of the CAD system will help radiologists to be more efficient and accurate in Lung-RADS categorization [7].

There is difference between the revised final CAD system and reference radiologist regarding the categorization of Lung-RADS 1 and 2, mostly due to size difference. However, this difference is insignificant and does not affect the patient's prognosis as both categories have the same annual follow-up recommendation.

This gives an excellent agreement of 88.6% (kappa 0.951) between the CAD system and reference radiologist in the overall categorization of all lung nodules according to Lung-RADS classification. There is discrepancy in only nine cases out of 79. All of them fall under categories 1 and 2 which does not affect the prognosis of the patient or the accuracy of the CAD system for lung cancer screening.

Our interobserver agreement was higher than that described in Jacobs et al.'s 2021 study. Their interobserver agreement was good for the dedicated viewer, with Fleiss kappa values of 0.66 (95% CI 0.64, 0.68).

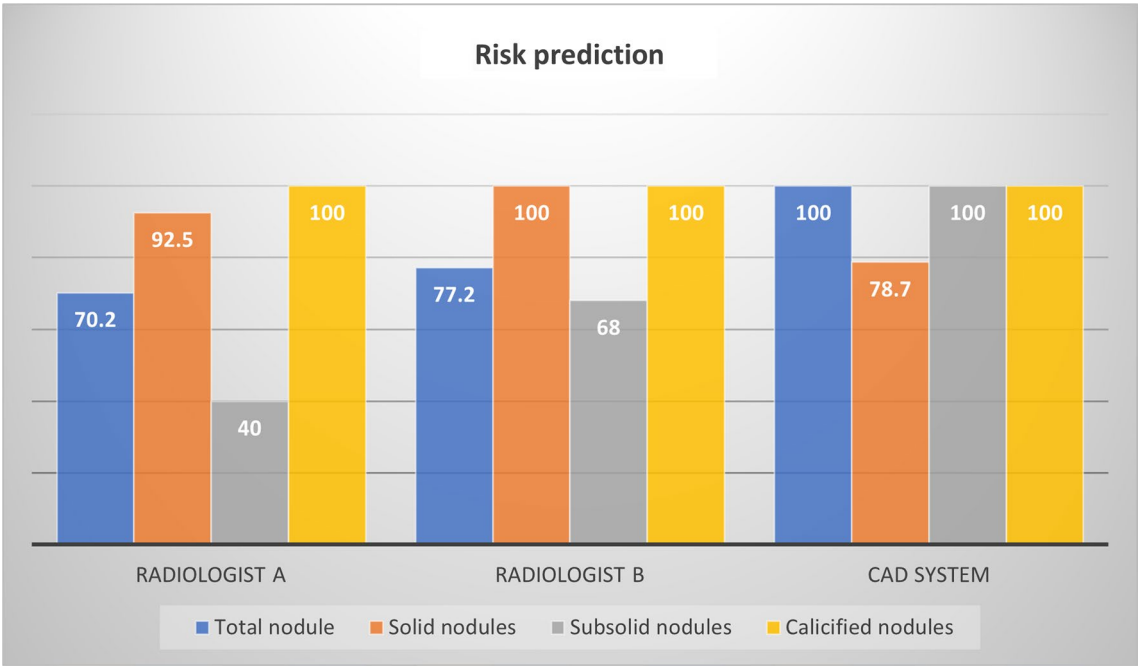


Fig. 6 Accuracy of risk prediction (in %) between radiologist A, radiologist B and CAD system in lesions with different consistencies

Jacobs et al. suggest that automatic nodule segmentation allows a more precise measurement of nodule size, especially when a prior scan is available. This is reflected in the finding that there were 67% (207 vs 68) fewer disagreement pairs that were due to different nodule diameter measurements when the dedicated CT lung screening viewer was used [9]. This study also indicates that most disagreements were related to determining the risk-dominant nodule, a task that was not automated but was performed by the individual observers.

Our hypothesis is the same as Jacobs et al. that there would be a shortened reading time in the CAD system owing to three factors: less time nodule measurement (automatic volumetry vs manual measurement), time saving in the synchronization of the baseline and follow-up scan sections and less time in report generation (automatic recording of the nodule size and Lung-RADS category vs manual reporting of the diameters and Lung-RADS category). However, our study did not include the precise time analysis in our results, so we have no data to test our hypotheses. We recommend that future studies to include the time factor in their analysis.

Nodule growth rate could not be involved in our results, even though it is one of the major factors in the Lung-RADS categorization. Only one case was presented with base line and follow-up studies, and the comparative analysis was automatically done (Fig. 7).

Application of CAD to comparisons of CT scans acquired at different time points will be the aim of a future study [7].

Limitations

This study has several limitations. First, the sample size was relatively small compared to the other related work. This could be due to the absence of national screening system data base that could have helped in more accurate and precise analysis. Second, we did not follow up the patients who tested positive due to short study time and difficulties to trace patient medical records. Only one patient was followed up after three months.

Third, as mentioned before we did not include the time factor in our data analysis despite being significantly decreased.

Conclusions

In conclusion, the application of CAD demonstrated increased sensitivity and specificity for the detection of lung nodules and total agreement in the detection of suspicious and probably benign nodules (lung cancer screening) and excellent level of agreement in the overall lung nodule categorization (Lung-RADS).

CAD systems with deep learning that is available as online software can help national wide scan campaigns that required minimum efforts/time from expert

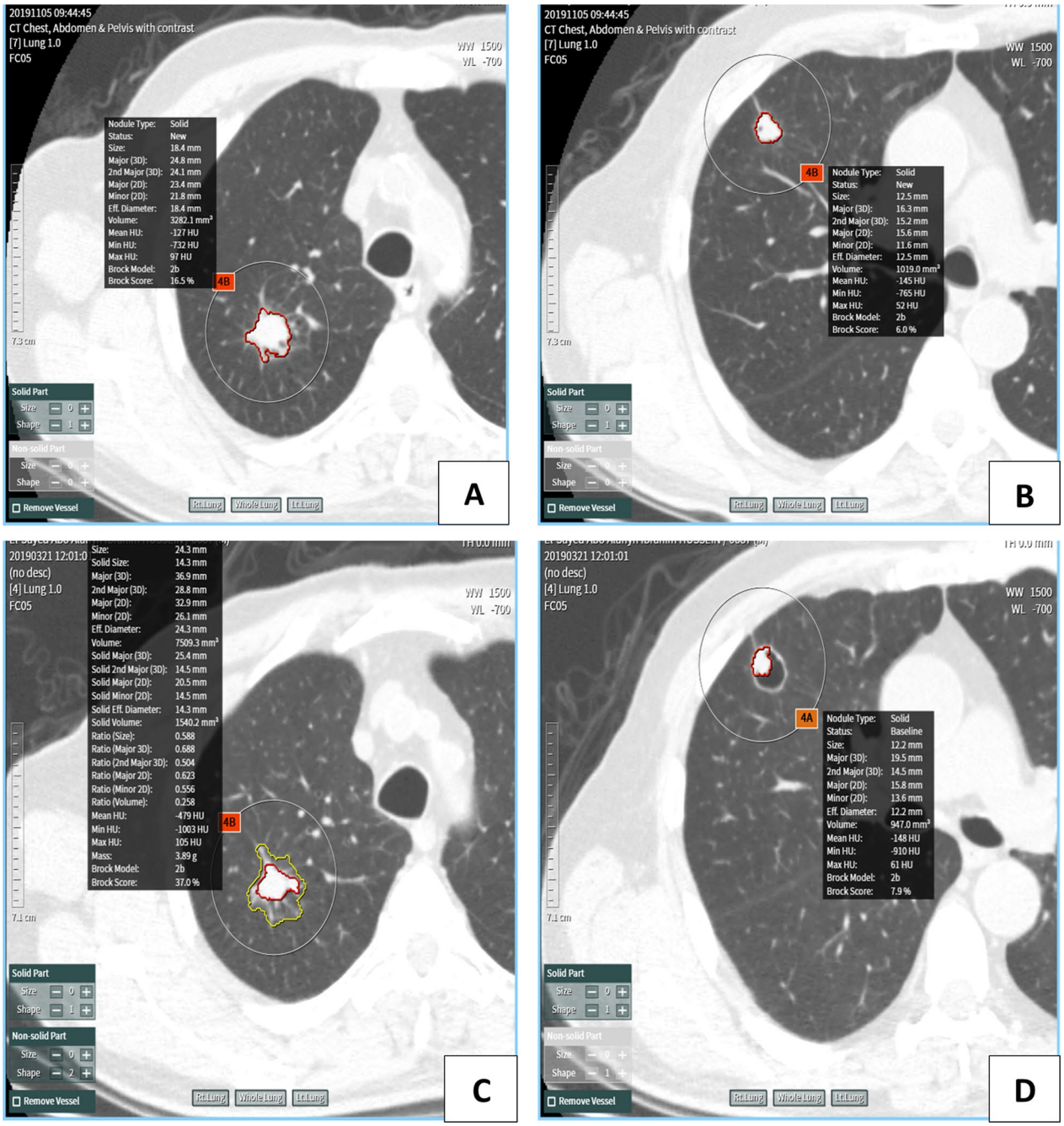


Fig. 7 **A**, **B**: base line axial CT of two pulmonary nodules, **C**, **D**: follow-up scan showed change in morphology

radiologists with diagnosis quality that is equal of above-10-year experienced radiologist.

Deep learning models are still limited to training data, training strategy and optimization criteria. This is clear from the failure of diagnosis, where the nodule was extremely large.

Abbreviations

ACR	American College of Radiology
AI	Artificial intelligence
BMI	Body mass index
CAD	Computer-aided detection
CT	Computed tomography
DL	Deep learning model
LAA	Low attenuation areas

Lung-RADS Lung Imaging Reporting and Data System
NPV Negative predictive value
PACS Picture archiving and communication system
PPV Positive predictive value

Acknowledgements
N/A.

Author contributions
NA was involved in data collection, validating the results obtained by the CAD system and original manuscript draft writing. AE, NZ and MA were involved in manual reporting of the cases and manuscript review and editing. ER and AE were involved in AI section review and editing. All authors have read and approved the manuscript.

Funding
The research was self-funded by the authors.

Availability of data and materials
The datasets used and analyzed in this study are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate
The study protocol was approved by the Research Ethics Committee of Faculty of Medicine at Suez Canal University on 28/6/2020 under the number of 4231.

Consent for publication
The authors grant the publisher the consent for publication of this work. It is a retrospective study, and a waiver of informed consent was obtained and approved by the Research Ethics Committee.

Competing interests
There is no conflict of interest between the researcher and the proposed software developers (we were offered a free trial version of the software, i.e., Coreline Soft's AVIEW metrics in exchange for honest review and research results).

Received: 15 December 2022 Accepted: 1 April 2023
Published online: 21 April 2023

References

1. Dubey AK, Gupta U, Jain S (2016) Epidemiology of lung cancer and approaches for its prediction: a systematic review and analysis. Chin J Cancer 35(1):71
2. Ibrahim A, Khaled H, Mikhail N, Baraka H, Kamel H (2014) Cancer incidence in Egypt: results of the national population-based cancer registry program. J Cancer Epidemiol 2014:437971
3. van Riel SJ, Jacobs C, Scholten ET, Wittenberg R, Winkler Wille MM, de Hoop B et al (2019) Observer variability for Lung-RADS categorisation of lung cancer screening CTs: impact on patient management. Eur Radiol 29(2):924–931
4. Martin MD, Kanne JP, Broderick LS, Kazerooni EA, Meyer CA (2017) Lung-RADS: pushing the limits. Radiographics 37(7):1975–1993
5. Al Mohammad B, Brennan PC, Mello-Thoms C (2017) A review of lung cancer screening and the role of computer-aided detection. Clin Radiol 72(6):433–442
6. Nasrullah N, Sang J, Alam MS, Mateen M, Cai B, Hu H (2019) Automated lung nodule detection and classification using deep learning combined with multiple strategies. Sensors (Basel) 19(17):3722
7. Park S, Park H, Lee SM, Ahn Y, Kim W, Jung K et al (2022) Application of computer-aided diagnosis for lung-RADS categorization in CT screening for lung cancer: effect on inter-reader agreement. Eur Radiol 32(2):1054–1064
8. Chamberlin J, Kocher MR, Waltz J, Snoddy M, Stringer NFC, Stephenson J et al (2021) Automated detection of lung nodules and coronary artery calcium using artificial intelligence on low-dose CT scans for lung cancer screening: accuracy and prognostic value. BMC Med 19(1):55
9. Jacobs C, Schreuder A, van Riel SJ, Scholten ET, Wittenberg R, Wille MMW et al (2021) Assisted versus manual interpretation of low-dose CT scans for lung cancer screening: impact on lung-RADS agreement. Radiol Imaging Cancer 3(5):e200160

Publisher's Note