

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 \pm 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.

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				

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.

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