Validation of a deep learning computer aided system for CT based lung nodule detection, classification and quantification and growth rate estimation in a routine clinical population

John T. Murchison[1], Gillian Ritchie[1], David Senyszak[2], Jeroen H. Nijwening[3], Edwin J.R. van Beek[1,2]

[1]Department of Radiology, Royal Infirmary of Edinburgh, Edinburgh, UK.

[2]Edinburgh Imaging facility QMRI, University of Edinburgh, Edinburgh, UK.

[3]Aidence, Amsterdam, Netherlands.

Address for correspondence:

Dr. J.H. Nijwening, email: Jeroen.nijwening@aidence.com and Professor John T Murchison, Consultant Radiologist, Royal Infirmary of Edinburgh, Little France Crescent, Edinburgh, UK. Email: john.murchison@luht.scot.nhs.uk

Abstract

Objective

In this study, we evaluated a commercially available computer assisted diagnosis system (CAD). The deep learning algorithm of the CAD was trained with a lung cancer screening cohort and developed for detection, classification, quantification, and growth of actionable pulmonary nodules on chest CT scans. Here, we evaluated the CAD in a retrospective cohort of a routine clinical population.

Materials and Methods

In total, a number of 337 scans of 314 different subjects comprising 470 nodules of 3-30 mm in size were included into the evaluation. Two independent thoracic radiologists alternately reviewed scans with or without CAD assistance to detect, classify, segment, and register pulmonary nodules. One additional, more experienced, radiologist served as an adjudicator. In addition, the cohort was analyzed by the CAD alone. The study cohort was divided into five different groups: 1) 178 CT studies without pulmonary nodules (control), 2) 95 studies with 1-10 pulmonary nodules, 23 studies from the same patients with 3) baseline and 4) follow-up studies, and 5) 18 CT studies with subsolid nodules. A reference standard for nodules was based on majority consensus with a third thoracic radiologist as required. Sensitivity, false positive (FP) rate and Dice inter-reader coefficient were calculated.

Results

After analysis of 470 pulmonary nodules, the sensitivity of CAD as a stand-alone test for detecting nodules was 82.3% with an average FP rate of 1.0 per CT scan. For radiologists without CAD and radiologist with CAD readings, this was 71.9% (95% CI: 66.0%, 77.0%) and 80.3% (95% CI: 75.2%, 85.0%) ($p < 0.01$), with average FP rate of 0.11 and 0.16 per CT scan, respectively. Accuracy and kappa of CAD for classifying solid vs sub-solid nodules was 94.2% and 0.77, respectively. Average inter-reader Dice coefficient for nodule segmentation was 0.83 (95% CI: 0.39, 0.96) and 0.86 (95% CI: 0.51, 0.95) for CAD versus readers. Mean growth percentage discrepancy of readers and CAD alone was 1.30 (95% CI: 1.02, 2.21) and 1.35 (95% CI: 1.01, 4.99), respectively.

Conclusion

The applied CAD significantly increased radiologist's detection of actionable nodules while only minimally increasing the false positive rate. CAD can automatically classify and quantify nodules and calculate nodule growth rate in a cohort of a routine clinical population. Results suggest this software can assist chest radiologists in pulmonary nodule detection and management within their routine clinical practice.

Keywords

machine learning, artificial intelligence, computer assisted diagnosis, lung nodule, Computed Tomography,

Key Points

1. The use of CAD during the readings yielded a nearly 10% higher sensitivity, compared to readings without CAD, while hardly affecting the false positive rate. The CAD system outperformed radiologists reading with and without the CAD system at an average false positive rate of 1.0 per scan.
2. The CAD software yielded a high accuracy of 94.2% and a kappa score of 0.77 for determining the composition (solid, sub-solid) of a pulmonary nodule.
3. The performance of the CAD software for segmenting pulmonary nodules on chest CT is comparable to that of experienced thoracic radiologists. These results show that the tool has the capacity to assist radiologists at pulmonary nodule management, although visual verification of the segmentation is still advised at present.

Abbreviations

CAD- Computer Assisted Detection
CT Computed Tomography
CI – Confidence Interval
FP- False Positive
TP –True positive
FN- False negative
NLST- National Lung Screening Trial
CE- Conformité Européenne
CADe Computer Assisted Detection Device
CADx Computer Assisted Diagnostic Device
kVp- Kilovoltage peak
mAs- Milliamp seconds
CTDIvol (page 5)
mGy- Milligray
GE- General Electric
3D- 3 dimensional
DICOM- Digital Imaging and Communications in Medicine
MPR- Multiplanar reconstruction
MIP- Maximum Intensity Projection
FROC- Free Response Receiver Operating Characteristic
VDT- Volume Doubling Time

Introduction

Lung nodule detection and management is one of the most frequent challenges in chest computed tomography (CT), not just in the context of lung cancer screening, but also in the staging of other malignancies in routine clinical practice. Lung cancer remains the third most prevalent cancer worldwide [1] , is both rising in incidence [2], and maintains high mortality rates with around 1.7 million global deaths annually [3]. Several recent studies demonstrated the benefits of lung cancer screening on early detection and improved outcomes [4-6]. The advent of lung cancer screening results in the need to detect smaller nodules, and therefore, the importance of fast and accurate detection is even more pronounced [7].

Lung cancer is ideally diagnosed by histopathological confirmation. However, the diagnostic process usually begins with chest CT where pulmonary nodules are identified incidentally. Pulmonary nodules are very common and mostly benign, however they should be considered as early stage cancers. The biggest challenges for pulmonary nodule detection on CT are acceptable sensitivity levels and reading times. Many failures in lung cancer diagnosis are due to detection errors rather than interpretation [8,9]. Several studies showed that the performance of (sub-specialist) radiologists for detecting pulmonary nodules is suboptimal [10,11].

Pulmonary nodule guidelines recommend different cut-off levels for nodule size and/or volume and volume doubling time as metrics to assess nodule size and growth [12-17]. There is increasing consensus that semi-automated volume assessment gives the most robust assessment for lung nodules and growth during follow up [7,16,17]. Another important parameter to consider is pulmonary nodule composition (solid vs sub-solid), as their malignant potential is different [18].

The above-mentioned challenges lead to many hospitals currently unable to assess nodules in a timely and accurate manner. Software aided detection and classification of lung nodules should improve the radiologist's diagnostic arsenal and throughput time and additionally could facilitate the roll-out of CT lung cancer screening [19]. Therefore, there has been an increasing focus on developing deep learning based computer assisted detection systems to facilitate more rapid reporting [20-29]. A few of these systems have reached availability for use in clinical practice. The study described here was performed to validate one such system, which was trained on a screening cohort, in a retrospective clinical population cohort of Scottish patients undergoing routine chest CT investigations.

Materials and Methods

This reader study was performed on a historical cohort of routine chest CT scans obtained at a single academic hospital, which were fully anonymized and transferred onto a stand-alone server. A waiver of informed consent was obtained from the South East Scotland Research Ethics Service.

Subject selection

A total of 349 chest CT examinations from 324 subjects, obtained at least 5 years prior to this study, were selected for retrospective analysis. Exclusion criteria were slice thickness >3mm and the presence of diffuse pulmonary disease in the radiology report and/or the CT images, with widespread abnormalities such as interstitial lung disease. Twelve studies were excluded at the time of reading due to rejections from the readers or corrupt studies that could not be processed by the CAD. In total, 337 CT scans from 314 subjects (173 women, 164 men) with a total of 470 pulmonary nodules (≥3mm and ≤30mm) were included (Table 1). From these CT scans, five groups were created. Group 1 consisted of 178 CT scans reported as being free from pulmonary nodules and group 2 consisted of 95 CT scans which were reported to have between 1 and 10 pulmonary nodules. Group 3 (23 CT scans) were patients undergoing follow up of a pulmonary nodule, group 4 consisted of these follow-up CT scans of group 3. Finally, group 5 was selected to enrich the study group and consisted of 18 CT scans with part-solid and/or ground-glass nodule(s). The mean age of all the subjects was 63 ± 7 years (range 32-88 years) and the mean age of the subjects in groups 1 and 2 was years 63 ± 5 (range 50-74). In addition, inclusion criteria for the subjects of the first two groups mimicked a lung cancer screening population: age 50-74 years, current smokers, a smoking history and/or radiological evidence of pulmonary emphysema.

CT protocol

A Toshiba Aquilion was used for most (330) studies; intravenous contrast was used in 22 CT scans. The mean tube peak potential energies used was 120 kVp, (range: 120-140 kVp), the average tube current was 243 mAs (range: 80-491 mAs) and the average CTDIvol was 14.0 mGy (range: 2.9-29.7). Data was reconstructed at a mean slice thickness of 1.0 mm (range 1.0-2.5mm). All CT scans were reconstructed using filtered back-projection, as these studies predated the routine application of novel reconstruction methods, such as iterative reconstruction.

Nodule definition

The Fleischner Society's definition for pulmonary nodules was broadly used during this study although "pulmonary nodule" was not firmly defined since the notion of nodule may not represent a single entity capable of verbal definition [13]. The size range was 3-30 mm with "actionable nodules" regarded as having a largest axial diameter between ≥5mm (or a volume of ≥80mm3) and ≤30mm.

CAD software

Veye Chest version 2.0 (Aidence B.V., Amsterdam, the Netherlands), which is commercially available in Europe since September 2018, was evaluated in this study. The software is primarily based on

Deep Learning technology, runs automatically and comprises of CADe and CADx functionality and growth rate calculation. The software has a detection threshold based on nodule likelihood values (range 0.0 to 1.0). For this study the threshold was set to 0.1 (high sensitivity and relatively high false positive rates).

Image annotation

A panel consisting of three thoracic radiologists (≥ 9 years' experience; JM, GR and EB, expert readers 1, 2 and 3, respectively) received training on the annotation tasks and annotation tool with written instructions available throughout. The study was performed at the University of Edinburgh between February – May 2018.

During an initial "blinded" phase, readers 1 and 2 independently performed a free search on all CT scans on a radiology reporting workstation. The detection results of CAD were made available at random in half the scans. Each CT scan was reviewed twice, once by one reader with the CAD results (CAD aided) and once by the other reader without the use of CAD (CAD unaided). Readers had to identify all lesions they considered to be a pulmonary nodule without clear benign morphological characteristics (calcification, typical perifissural lymph node). Any nodules requiring follow-up according to lung cancer screening criteria were classified as "actionable nodules". The Reader would mark an actionable pulmonary nodule manually on unaided scans or classify a CAD prompt on an aided scan as either true positive (TP) or false positive (FP). Any actionable nodules identified on aided scans, which had not been detected by CAD were also recorded. Readers registered all actionable nodules present on CT scans from groups 3 and 4. Finally, the readers classified all FP CAD prompts into three different groups: micro-nodules (largest axial diameter <3mm), masses (largest axial diameter >30mm), benign nodules (benign calcification pattern or clear benign perifissural appearance) and non-nodules (pleural plaque, scar tissue, atelectasis, fibrosis, fissure thickening, pleural fluid, pleural thickening, intrapulmonary vessels, consolidations, outside of lung tissue, or other (free format)).

After completing all the readings on the workstations the readers reviewed their own previously identified nodules on a tablet (iPad Pro). The reader was asked to determine the composition (solid or sub-solid) and segment each nodule on every slice. The results from readers 1 and 2 were evaluated for the presence of any discrepancies. Discrepancies were defined as a difference between the results in terms of: location (3D Dice coefficient of 0); composition; segmentation (3D Dice coefficient < -1 standard deviation of the mean) and nodule registration. The Dice coefficient is a spatial overlap index and a reproducibility validation metric with a range of 0.0 (no overlap) to 1.0 (perfect overlap) [30].

Reader 3 subsequently adjudicated all discrepancies without the results of CAD using the same materials used in the blinded phase. Reader 3 created an independent reading for each nodule that had a discrepancy for at least one characteristic.

Reference standard

The reference standard for actionable nodules consisted of lesions from groups 1 and 2 which were marked as a pulmonary nodule by the majority of the panel and met the size criteria. The location of an actionable nodule was defined by averaging the centre of mass of all reader's segmentations. Subsequently, the radius and volume were derived from the segmentations. The reference standard for composition was determined by majority consensus of lesions from groups 1-3 and 5. There was no consensus requirement for the reference standard of segmentation. All segmentations of a nodule from groups 1-3 and 5 were retained. Finally, the reference standard for nodule registration was created using studies from groups 3 and 4, subsequently growth rate was determined as the relative volume difference between nodules visible on a study from group 3 and on its follow-up study from group 4.

Data analysis

Findings from a reader or from CAD were scored as either TP, if the centre of the detection was within the volume of actionable nodules in the reference standard, or otherwise as FP. Findings from a reader or CAD in the centre of the detection that was within the volume of a micro-nodule or a mass or a nodule detected by only a single reader were neither scored TP or FP. The absence of a prompt from CAD in the centre of an actionable nodule in the reference standard was considered FN. Sensitivity for detecting actionable nodules and the average number of FP detections per CT scan for AIDED readings, UNAIDED readings and CAD alone was calculated using the reference standard for actionable nodules.

The sensitivity, specificity, positive predictive value and negative predictive value, accuracy and kappa score for determining the composition (solid or sub-solid) by CAD alone was calculated using the reference standard for composition.

The segmentation accuracy of readers was calculated as the Dice coefficient between each reader's segmentation and averaged (inter-reader dice coefficient). The segmentation accuracy of CAD alone was calculated as the Dice coefficient between each CAD segmentation and each individual reader segmentation and averaged. In addition, the inter-reader mean diametric and volumetric discrepancy was calculated using the largest axial diameter and volume from each segmentation of each reader's segmentation and compared to those from the other readers, this was also calculated for CAD alone compared to the other readers.

For sequential scans (groups 3 and 4), nodule registration from CAD was scored as either TP, if the detected registration was included in the nodule registration reference standard, or otherwise as FP. The mean discrepancy between growth percentages determined by readers and CAD alone was calculated.

Statistical Analysis

One-tailed Welch's t-test was used to accept the hypothesis that the mean sensitivity of AIDED is higher than the mean sensitivity of the UNAIDED readings ($p < 0.05$), with the use of bootstrapping

over scans with 2000 iterations. One-tailed Welch's t-test was used to accept the hypothesis that the mean CAD Dice score is higher than the mean inter-reader Dice score ($p < 0.05$).

Results

Groups 1 and 2 consisted of 273 CT scans with 269 actionable nodules. The radiologists with CAD readings showed a sensitivity of 93.5% and average FP rate of 3.0. The sensitivity for detecting actionable nodules of radiologists without CAD on scans from groups 1 and 2 was: 71.9% (95% CI: 66.0%, 77.0%) and 80.3% (95% CI: 75.2%, 85.0%) ($p < 0.01$), respectively. The average FP rate of radiologists alone and radiologists with CAD readers was: 0.11 and 0.16, respectively. The maximum obtainable sensitivity of CAD alone was 95.9% at an average FP rate of 10.9. The sensitivity of CAD alone was equivalent to radiologists without and radiologists with CAD readings at an average FP rate of 0.62 and 0.88, respectively (Figure 1). Details regarding the number of CT scans and nodules per group are described in Table 1.

The composition of nodules within groups 1, 2, 3, and 5 totaled 325 solid nodules and 57 sub-solid nodules. The sensitivity, specificity, positive predictive value and negative predictive value of CAD for determining the composition of solid nodules in groups 1, 2, 3, and 5 was 98.8%, 68.4%, 90.7% and 94.7%, and was 68.4%, 98.8%, 94.7% and 90.7% for sub-solid nodules, respectively. The accuracy and kappa of CAD for determining the composition (solid or sub-solid) of a pulmonary nodule was 94.2% and 0.77.

The CAD software successfully segmented 95% of pulmonary nodules from groups 1-3 and 5. The average inter-reader Dice coefficient was 0.83 (95% CI: 0.39, 0.96) versus 0.86 (95% CI: 0.51, 0.95) for CAD alone ($p<0.01$). The mean largest axial diameter of all nodules was 7.68 ± 3.50 mm (range: 3.42 - 28.45 mm) and the mean volume was 198 ± 333 mm3 (range: 21 - 2797 mm3. The inter-reader geometric mean diameter discrepancy was 1.15 (95% CI: 1.00, 1.58) versus 1.17 (95% CI: 1.01, 1.69) for CAD alone. The inter-reader geometric mean volumetric discrepancy was 1.39 (95% CI: 1.01, 3.19) versus 1.38 (95% CI: 1.01, 3.38) for CAD alone.

The total number of nodules in group 3 and 4 was 68 and 42, respectively. The total number of nodule-pairs in groups 3 and 4 was 23 and all nodules were successfully identified by CAD. The mean growth percentage discrepancy of readers and CAD alone was 1.30 (95% CI: 1.02, 2.21) and 1.35 (95% CI: 1.01, 4.99), respectively, which was not statistically significant.

Discussion

Computer assisted detection and diagnosis software, including convolutional neural networks and machine learning approaches have shown promising results in aiding radiologists to identify

incidental pulmonary nodules. A study using the LIDC database as a comparison tested 108 CT scans and demonstrated high sensitivity and specificity [21]. However, there are also conflicting results. A more recent study [22] demonstrated moderately high sensitivity of 84% and a corresponding positive predictive value of 67% when tested in 100 patients with 106 biopsied lung nodules at a slice thickness of 3 mm. Another commercial system was clearly suboptimal when tested on 50 pure ground glass and 50 part solid nodules [23].

The most comprehensive deep learning system to date used 11,625 chest CT scans for model training and validation and subsequently used 1,129 chest CT studies for testing of the model with a sensitivity between 74%-86% at FP rates of 1-8, respectively [24].

This study shows improved sensitivity of experienced thoracic radiologists using aided detection from 71.9% to 90.3% with a minor increase in FP rate. The maximum stand-alone CAD sensitivity was 95.9% at an average FP rate of 10.9, which would be unworkable in clinical practice. A more acceptable average FP rate would be between 1 and 2 with corresponding sensitivity range (82.3% - 89.0%), outperforming thoracic radiologists with and without using CAD.

The AIDED readings outperformed UNAIDED readings, yielding a sensitivity of 93.5% at an average FP rate of 3.0. However, 36 CAD detected nodules confirmed by the majority of the panel were scored as FP by one reader. A possible explanation could be that due to the high number the readers develop a tendency to call CAD prompts FP. Another explanation could be a structural difference in pulmonary nodule definition between the readers. Even allowing for this, the number of TP nodules detected by CAD was higher than without CAD.

The Deep Learning model for nodule detection was trained on data from a lung cancer screening cohort but this study shows that it is effective in a general, “real life” clinical setting where it improves the sensitivity of detection of actionable nodules by thoracic radiologists.

The CAD software yielded a high accuracy of 94.2% and a kappa score of 0.77 for determining the composition (solid, sub-solid) of a pulmonary nodule. The segmentation accuracy of CAD was similar to that of thoracic radiologists, CAD dice 0.86 and inter-reader dice 0.83 ($p < 0.01$).

This is the first study using Veye Chest to look at a routine cohort of smokers who underwent chest CT for non-screening purposes. Veye Chest was initially validated on a lung cancer screening population [4,19] and the results of our study are of similar sensitivity and accuracy to that initial cohort, (87% at 1 FP/scan) confirming broader use is feasible. These findings compare favorably with several other software tools [22, 24].

In addition, the CAD software yielded a perfect score for a limited number of nodule pairs; sensitivity 100.0% without FP pairs, but further validation will be required. The mean growth percentage

discrepancy of readers was 1.30 compared to 1.35 for CAD alone. However, due to a single incorrect segmentation of the CAD software, the upper end of its confidence interval (95% CI,1.01-4.99) is twice as high compared to that of readers (95% CI,1.02-2.21), illustrating that visual verification is still required. Nevertheless, this compares favorably with results from a software comparison sub-study of the NELSON study in 50 subjects [26]. Similarly, a study of 134 participants in the NLST also demonstrated a decrease in variability of detection and volumetry with the use of software [27].

This study has several limitations. First, the data was obtained from a single site and the vast majority of CT scans were acquired by a single CT scanner vendor. A recent study demonstrated decreased diagnostic performance of machine learning-based radiomics models in 26 patients with subsolid adenocarcinoma nodules when iterative reconstruction was applied [28]. Therefore, care must be taken to validate any software tool on actual datasets. Although differences between scanner manufacturers and CT imaging protocols may alter the interpretation of lung parenchymal features, it is unlikely to significantly affect the presence or absence of actionable pulmonary nodules. Indeed, all vendors have taken part in various CT lung screening trials and have shown similar results. Second, the readings were performed under artificial conditions and therefore the performance of the CAD software and the radiologists may be different in a real-world setting. This is considered of potential importance, as artificial conditions and use in selected datasets tend to lead to excellent results of lung CT CAD systems [29, 30]. Further prospective clinical validation is therefore required, and this also highlights the need for seamless workflow integration of this software for it to become standard practice. Lastly, the sensitivity of the readers without and with CAD versus CAD alone was calculated using the reference standard established by the same readers and CAD. The only addition to this was the third reader, who effectively assured consensus on the final classification and morphologic features of lung nodules. One could consider performing the same test in multiple readers, but this would be time consuming and unlikely lead to significantly different results.

In conclusion, the use of Veye Chest significantly increased radiologist's detection of actionable nodules while only minimally increasing the false positive rate. This CAD system is able to automatically classify, quantify and calculate the growth rate of pulmonary nodules. These results suggest that Veye Chest has the potential to assist radiologists in the tasks of pulmonary nodule detection and management on routine chest CT.

References

1. International Agency for Research on Cancer (IARC). GLOBOCAN 2018. https://www.uicc.org/news/new-global-cancer-data-globocan-2018 (accessed 12/07/2020).
2. Jemal A, Bray F, Center MM, Ferlay J, Ward E and Forman D, Global cancer statistics, CA Cancer J Clin 61 (2011) 69–90. https://doi.org/10.3322/caac.20107.
3. International Agency for Research on Cancer (IARC). GLOBOCAN Cancer Fact Sheets: Lung Cancer. International Agency for Research on Cancer (IARC); 2017, http://gco.iarc.fr/today/data/factsheets/cancers/15-Lung-fact-sheet.pdf (accessed 12/07/2020).
4. Church TR, Black WC, Aberle D, et al. , National Lung Screening Trial, New Engl J Med 368 (2013) 1980-1991. https://doi.org/10.1056/NEJMoa1209120.

5. Becker N, Motsch E, Trotter A, et al., Lung cancer mortality reduction by LDCT screening-Results from the randomized German LUSI trial, Int J Cancer 146 (2020) 1503-1513. https://doi.org/10.1002/ijc.32486.
6. de Koning HJ, van der Aalst CM, de Jong PA, et al., Reduced Lung-Cancer Mortality With Volume CT Screening in a Randomized Trial, N Engl J Med 382 (2020) 503-513. https://doi.org/10.1056/NEJMoa1911793.
7. Oudkerk M, Deveraj A, Vliegenthart R, et al., European position statement on lung cancer screening, Lancet Oncol 18 (2017) e754-766. https://doi.org/10.1016/S1470-2045(17)30861-6.
8. Kakinuma R, Ohmatsu H, Kaneko M et al., Detection failures in spiral CT screening for lung cancer: analysis of CT findings, Radiology 212 (1999) 61–66. https://doi.org/10.1148/radiology.212.1.r99jn1461.
9. White CS, Romney BM, Mason AC, Austin JH, Miller BH, Protopapas Z, Primary carcinoma of the lung overlooked at CT: analysis of findings in 14 patients. Radiology 199 (1996)109–115. https://doi.org/ 10.1148/radiology.199.1.8633131.
10. Kakinuma R, Ashizawa K, Kobayashi T et al. , Comparison of sensitivity of lung nodule detection between radiologists and technologists on low-dose CT lung cancer screening images. Br J Radiol 85 (2012) e603–608. https://doi.org/ 10.1259/bjr/75768386.
11. Nair A, Gartland N, Barton B et al., Comparing the performance of trained radiographers against experienced radiologists in the UK lung cancer screening (UKLS) trial, Br J Radiol 89 (2016) 20160301 https://doi.org/10.1259/bjr.20160301.
12. Callister MEJ, Baldwin DR, Akram AR, et al., British Thoracic Society guidelines for the investigation and management of pulmonary nodules, Thorax 70 (2015) suppl 2:ii1-ii54. https://doi.org/10.1136/thoraxjnl-2015-207168.
13. MacMahon H, Naidich DP, Goo JM, et al., Guidelines for management of incidental pulmonary nodules detected on CT imaging: From the Fleischner Society 2017, Radiology 284 (2017) 228-243. https://doi.org/ 10.1148/radiol.2017161659.
14. McKee BJ, Regis SM, McKee AB, Flacke S, Wald C, Performance of ACR lung-RADS in a clinical CT lung screening program, J Am Coll Radiol 12 (2015) 273-276. https://doi.org/10.1016/j.jacr.2014.08.004.
15. Bankier AA, MacMahon H, Goo JM, Rubin GD, Schaefer-Prokop CM, Naidich DP, Recommendations for measuring pulmonary nodules at CT: A statement from the Fleischner Society, Radiology 285 (2017) 584-600. https://doi.org/10.1148/radiol.2017162894.
16. Deveraj A, van Ginneken B, Nair A, Baldwin D, Use of volumetry for lung nodule management: theory and practice, Radiology 284 (2017) 630-644. https://doi.org/10.1148/radiol.2017151022.
17. Heuvelmans MA, Walter JE, Vliegenthart R, et al., Disagreement of diameter and volume measurements for pulmonary nodule size estimation in CT lung cancer screening, Thorax 73 (2018) 779-781. https://doi.org/ 10.1136/thoraxjnl-2017-210770.
18. Naidich DP, Bankier AA, MacMahon H, et al., Recommendations for the management of subsolid pulmonary nodules detected at CT: A statement from the Fleischner Society. Radiology 266 (2013) 304-317. https://doi.org/10.1148/radiol.12120628.
19. Armato SG, McLennan G, McNitt-Gray MF, et al., Lung image database consortium: developing a resource for the medical imaging research community, Radiology 232 (2004) 739–748. https://doi.org/ 10.1148/radiol.2323032035.
20. Roos JE, Paik D, Olsen D et al., Computer-aided detection (CAD) of lung nodules in CT scans: radiologist performance and reading time with incremental CAD assistance, Eur Radiol 20 (2010) 549–557 https://doi.org/10.1007/s00330-009-1596-y.

21. Brown MS, Lo P, Goldin JG, et al., Towards clinically usable CAD for lung cancer screening with computed tomography, Eur Radiol 24 (2014) 2719-2728. https://doi.org/10.1007/s00330-014-3329-0.
22. Wagner AK, Hapich A, Pscyhogios MN, Teichgraber U, Malich A, Papageorgiou I, Computer-aided detection of pulmonary nodules in computed tomography using ClearRead CT, J Med Syst 43 (2019) 58. https://doi.org/ 10.1007/s10916-019-1180-1.
23. Benzakoun J, Bommart S, Coest J, et al., Computer-aided diagnosis (CAD) of subsolid nodules: evaluation of a commercial CAD system, Eur J Radiol 85 (2016) 1728-1734. https://doi.org/10.1007/s10916-019-1180-1.
24. Liu K, Li Q, Ma J, et al., Evaluating a fully automated pulmonary nodule detection approach and its impact on radiologist performance, Radiol Artificial Intell 1 (2019) e180084. https://doi.org/ 10.1148/ryai.2019180084
25. Zhao YR, van Ooijen PMA, Dorriums MD, et al., Comparison of three software systems for semi-automatic volumetry of pulmonary nodules on baseline and follow-up CT examinations, Acta Radiol 55 (2014) 691-698. https://doi.org/10.1177/0284185113508177.
26. Jeon KN, Goo JM, Lee CH, et al., Computer-aided nodule detection and volumetry to reduce variability between radiologists in the interpretation of lung nodules at low-dose screening CT, Invest Radiol 47 (2012) 457-461. https://doi.org/ 10.1097/RLI.0b013e318250a5aa.
27. Kim HJ, Park CM, Gwak J, et al., Effect of CT reconstruction algorithm on the diagnostic performance of radiomics models: a task-based approach for pulmonary subsolid nodules, Am J Roentegenol 212 (2018) 505-612. https://doi.org/ 10.2214/AJR.18.20018.
28. Jacobs C, van Rikxoort EM, Murphy K, Prokop M, Schaefer-Prokop CM, van Ginneken B., Computer-aided detection of pulmonary nodules: a comparative study using the public LIDC/IDRI database, Eur Radiol 26 (2016) 2139-2147. https://doi.org/10.1007/s00330-015-4030-7.
29. Setio AAA, Traverso A, de Bel T, et al., Validation, comparison, and combination of algorithms for automated detection of pulmonary nodules in computed tomography imaging: The LUNA16 challenge, Med Image Anal 42 (2017) 1-13. https://doi.org/10.1016/j.media.2017.06.015.
30. Zou KH, Warfield SK, Bharatha A, et al., Statistical validation of image segmentation quality based on a spatial overlap index, Acad Radiol 11 (2004) 178-189. https://doi.org/10.1016/s1076-6332(03)00671-8.