

RESEARCH ARTICLE

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

John T. Murchison^{1*}, Gillian Ritchie¹, David Senyszak², Jeroen H. Nijwening^{3*}, Gerben van Veenendaal³, Joris Wakkie³, Edwin J. R. van Beek^{1,2}

1 Department of Radiology, Royal Infirmary of Edinburgh, Edinburgh, United Kingdom, **2** Edinburgh Imaging facility QMRI, University of Edinburgh, Edinburgh, United Kingdom, **3** Aidence, Amsterdam, The Netherlands

* john.murchison@luht.scot.nhs.uk (JTM); jeroen.nijwening@aidence.com (JHN)



OPEN ACCESS

Citation: Murchison JT, Ritchie G, Senyszak D, Nijwening JH, van Veenendaal G, Wakkie J, et al. (2022) Validation of a deep learning computer aided system for CT based lung nodule detection, classification, and growth rate estimation in a routine clinical population. *PLoS ONE* 17(5): e0266799. <https://doi.org/10.1371/journal.pone.0266799>

Editor: Chang Min Park, Seoul National University Hospital, Seoul National University College of Medicine, REPUBLIC OF KOREA

Received: March 8, 2021

Accepted: March 28, 2022

Published: May 5, 2022

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0266799>

Copyright: © 2022 Murchison et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

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 with reported 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. A third, 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 reported pulmonary nodules, 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 the 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 readings for radiologists without CAD and radiologist with CAD, were 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

Data Availability Statement: Data cannot be shared publicly because of confidential patient information. Anonymous data is stored on a stand-alone server at the Edinburgh Imaging facility QMRI, University of Edinburgh, Edinburgh, UK. <http://www.ed.ac.uk/edinburgh-imaging> To access the data, please contact the Caldicott Guardian's Office: Caldicott Office NHS Lothian Waverley Gate 2-4 Waterloo Place Edinburgh EH1 3EG Phone +44-131-4655452 Caldicott.guardian@nhslothian.scot.nhs.uk.

Funding: This study was funded by NHS England via the SBRI Phase 1 grant for "Early Detection and Diagnosis of Cancer" which was granted to Aidence (Amsterdam, the Netherlands). Aidence provided support with this grant in the form of salaries for authors [JTM, GR, EJRVB], but did not have any additional role in the study design, data collection and most of the analysis, decision to publish, or preparation of the manuscript. The specific roles of these authors are articulated in the 'author contributions' section. EJRVB is a member of the Medical Advisory Board of Aidence and received support in the form of salary for the work performed in this study. JTM has no affiliation with Aidence and received support in the form of salary for the work performed in this study. GR has no affiliation with Aidence and received support in the form of salary for the work performed in this study. DS has no affiliation with Aidence and received support in the form of salary for the work performed in this study. JHN is a full time, paid employee of Aidence at the time of submission of this manuscript. GvV is a full time, paid employee of Aidence at the time of submission of this manuscript.

Competing interests: JTM, GR, and DS declare no competing interests. EJRVB declares ownership of QCTIS Ltd and serves on the medical advisory boards of Aidence BV and Imbio LLC. EJRVB received a restricted research grant from Siemens Healthineers and speaker fees from AstraZeneca and Roche Diagnostics. EJRVB's non-financial competing interest is serving as an expert witness on medicolegal advice on imaging based cases. JHN and GvV are full time, paid employees of Aidence and have aided in part of the raw data analysis (GvV) or re-submitting this manuscript to PLOS ONE (JHN). These interests do not alter our adherence to PLOS ONE policies on sharing data and materials.

Abbreviations: 3D, 3 dimensional; CAD, Computer Assisted Detection; CADe, Computer Assisted Detection Device; CADx, Computer Assisted Diagnostic Device; CE, Conformité Européenne; CI, Confidence Interval; CT, Computed Tomography;

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 yet also minimally increasing the false positive rate. The CAD can automatically classify and quantify nodules and calculate nodule growth rate in a cohort of a routine clinical population. Results suggest this Deep Learning software has the potential to assist chest radiologists in the tasks of pulmonary nodule detection and management within their routine clinical practice.

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, is both rising in incidence [1], and maintains high mortality rates with around 1.8 million global deaths annually. Several recent studies demonstrated the benefits of lung cancer screening on early detection and improved outcomes [2–4]. 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 [5].

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 diagnoses are due to detection errors rather than interpretation [6, 7]. Several studies showed that the performance of (sub-specialist) radiologists for detecting pulmonary nodules is suboptimal with reported sensitivities around 80% [8, 9].

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 [10–15]. There is increasing consensus that semi-automated volume assessment gives the most robust assessment for lung nodule growth during follow up [5, 14, 15]. Another important parameter to consider is pulmonary nodule composition (solid vs sub-solid), as sub-solid nodules are more likely to be malignant [16].

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 [17]. Therefore, there has been an increasing focus on developing deep learning based computer assisted detection systems to facilitate more rapid reporting [18–28]. 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 originally trained on a lung cancer screening cohort, in a retrospective clinical population cohort of Scottish patients undergoing routine chest CT investigations.

CTDIvol, Volume CT Dose Index; DICOM, Digital Imaging and Communications in Medicine; FP, False Positive; FN, False negative; FROC, Free Response Receiver Operating Characteristic; GE, General Electric; MIP, Maximum Intensity Projection; MPR, Multiplanar reconstruction; kVp, Kilovoltage peak; mAs, Milliamp seconds; mGy, Milligray; NLST, National Lung Screening Trial; TP, True positive; VDT, Volume Doubling Time.

Materials and methods

Subject selection

CT studies from a routine clinical population, in a single academic hospital, between January 2008 and December 2009 (9 years before start of this retrospective study), were searched for the following inclusion criteria: age 50–74 years, current smokers, a smoking history and/or radiological evidence of pulmonary emphysema. CT studies excluded from the analysis had slice thickness $>3\text{mm}$, or the presence of diffuse pulmonary disease in the radiology report, and/or the CT images, with widespread abnormalities such as interstitial lung disease.

In total, 337 fully anonymized chest CT examinations from 314 subjects (173 women, 164 men) with reported nodule size of $\geq 3\text{mm}$ and $\leq 30\text{mm}$ were included and transferred onto a stand-alone server. A waiver of informed consent was obtained from the South East Scotland Research Ethics Service.

From these CT scans, five groups were created. Group 1: 178 CT scans, initially reported as being free from pulmonary nodules. Group 2: 95 CT scans, reported to have between 1 and 10 pulmonary nodules. Group 3: 23 CT scans from patients undergoing follow-up of a pulmonary nodule. Group 4, consisted of the 23 follow-up scans of group 3. Finally, group 5 consisted of 18 scans to enrich the study group with part-solid and/or ground-glass nodule(s).

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. Other CT scanners used were: Toshiba Aquilion-CX: 2 scans, Toshiba Aquilion ONE: 1 scan, GE Medical Systems LightSpeed 16: 2 scans, GE Medical Systems LightSpeed: 2 scans.

Nodule definition

The Fleischner Society's definition for pulmonary nodules was broadly used during this study [12]. The size range was 3–30 mm with “actionable nodules” regarded as having a largest axial diameter between $\geq 5\text{mm}$ (or a volume of $\geq 80\text{mm}^3$) and $\leq 30\text{mm}$ as recommended by the British Thoracic Society guidelines [10].

CAD software

Veye Chest version 2.0 (now known as Veye Lung Nodules, developed by Aidence B.V., Amsterdam, the Netherlands), which is CE marked and certified as a Class IIb medical device, was evaluated in this study, see (S1 Fig). The software is primarily based on Deep Learning technology, which was trained on 45k+ chest CT-scans (slice thickness $\leq 3\text{mm}$ without contrast fluid) and 40k+ annotations by radiologists. The software 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 which means that the threshold is set to a high sensitivity and consequently a relatively high false positive rate.

Image annotation

A panel consisting of three thoracic radiologists (≥ 9 years' experience; JTM, GR and EJRVB, 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.

Two datasets were created from the 337 CT scans: one set with CAD results and one set without CAD results. Reader 1 reviewed all the CT scans, but half of the CT scans with the CAD results (CAD aided) and the other half without CAD results (CAD unaided). For reader 2 this was vice versa. Hence, 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” [10]. 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 four different groups: micro-nodules (largest axial diameter $< 3\text{mm}$), masses (largest axial diameter $> 30\text{mm}$), benign nodules (benign calcification pattern or clear benign perifissural appearance) and non-nodules (1088 non-nodules in total. More specific: atelectasis: 283; scar tissue: 229; fibrosis: 157; vessels: 126; non-lung: 81; other: 81; pleural: 80; fissure: 25; pleural plaque: 14; consolidations: 12 pleural plaque).

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) [29].

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 of having a largest axial diameter between $\geq 5\text{mm}$ (or a volume of $\geq 80\text{mm}^3$) and $\leq 30\text{mm}$. The majority consisted of consensus between reader 1 and 2 or, in the case of no consensus, the adjudication of reader 3. The location of an actionable nodule was defined by averaging the center of mass of all reader's segmentations. Subsequently, the radius and volume were derived from these segmentations. The reference standard for composition was determined by majority consensus of lesions from groups 1–3 and 5. Finally, 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 center 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 center 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 center 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 see [Table 1](#). Remarkably, nodules were identified in group 1, highlighting the importance of concurrent reading. 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

Table 1. Distribution of study subjects and nodule size by group.

Group	Number of subjects	Number of CT scans	Number of nodules with largest axial diameter ≥ 3 and < 5 mm	Number of nodules with largest axial diameter or mean volume ≥ 5 mm / ≥ 80 mm ³ and < 30 mm
1	178	178	19	71
2	95	95	34	198
3	23	23	0	68
4		23	6	36
5	18	18	2	36
TOTAL	314	337	61	409

<https://doi.org/10.1371/journal.pone.0266799.t001>

radiologists without and radiologists with CAD readings at an average FP rate of 0.62 and 0.88, respectively (Fig 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.

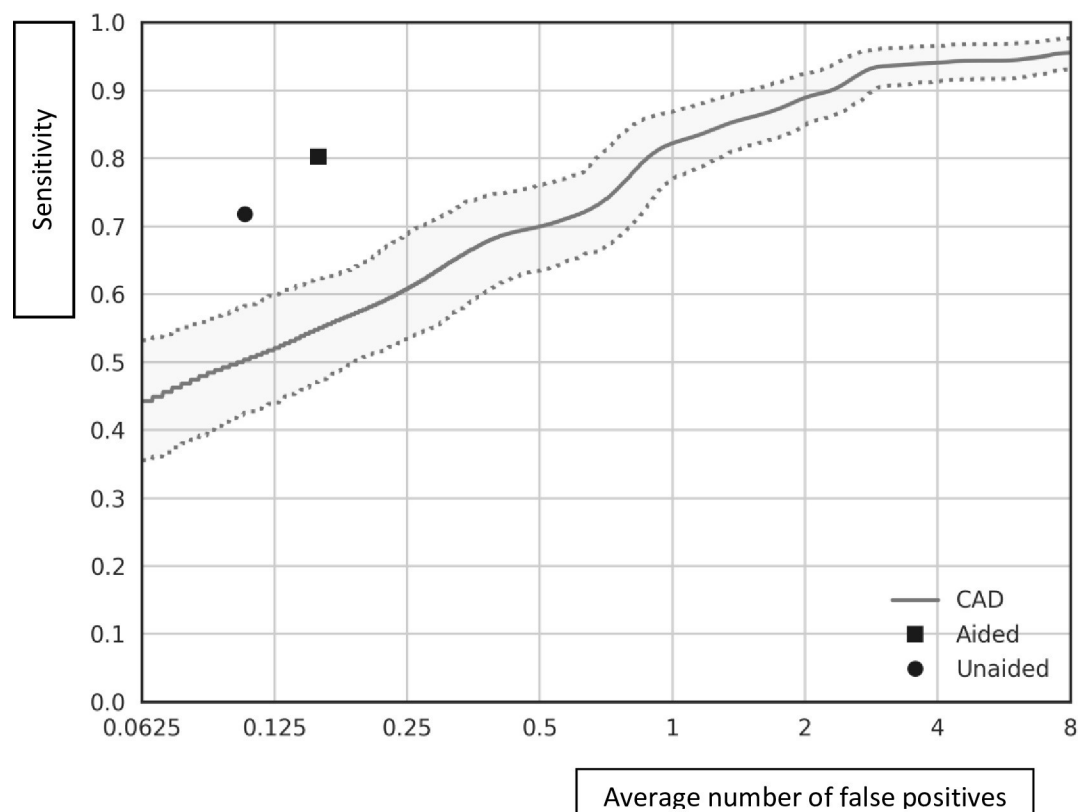


Fig 1. Free-response ROC (FROC) curve. This curve shows the standalone performance of CAD for detecting actionable nodules based on scans from groups 1 and 2. The vertical axis represent the sensitivity and the horizontal axis the average number of false positives per scan. The dashed lines show the upper and lower boundary of 95% confidence interval, bootstrapping of scans with 2000 samples. The circle represents the UNAIDED performance (sensitivity: 71.9% average FP rate 0.11 per scan) and square the AIDED performance (sensitivity: 80.3% average FP rate 0.16 per scan) for detecting actionable nodules.

<https://doi.org/10.1371/journal.pone.0266799.g001>

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

The study described here 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 standalone performance of the CAD, when set to the threshold of 0.1 applied in this study, correlates to an average sensitivity of 95% and an average number of 7 false positives per study based on this dataset.

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 [18, 20]. However, there are also conflicting results. A more recent study [21] 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 [22]. 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 [23].

This is the first study using this CAD software to look at a routine cohort of smokers who underwent chest CT for non-screening purposes. The software tested here was initially validated on a lung cancer screening population [2, 17] and the results of our study are of similar sensitivity and accuracy to that initial cohort, (87% at 1 FP/scan) [2] confirming broader use is feasible.

In this study, 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.

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

In addition, the CAD software yielded a perfect score for a limited number of nodule pairs and analyzing its volumes; 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 [25]. Similarly, a study of 134 participants in the NLST also demonstrated a decrease in variability of detection and volumetry with the use of software [26].

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 [27]. 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 [28, 29]. 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. Recently, the software described here was independently evaluated in a large teaching hospital [30]. This study found a sensitivity of 88% and the mean FP rate was 1.04 FPs per scan.

In conclusion, the use of the CAD significantly increased radiologist's detection of actionable nodules yet also increasing the false positive rate. The Deep Learning model for nodule detection was trained on data from a lung cancer screening cohort. In addition, this study appears to show that it is also effective in a general, "real life" clinical setting where it improves the sensitivity of detection of actionable nodules by thoracic radiologists. This CAD system is able to automatically classify, quantify, and calculate the growth rate of pulmonary nodules. These results suggest that Deep Learning software has the potential to assist radiologists in the tasks of pulmonary nodule detection and management on routine chest CT.

Supporting information

S1 Fig. Screenshot of Veye Chest.
(TIF)

Author Contributions

Conceptualization: John T. Murchison, Edwin J. R. van Beek.

Data curation: Gerben van Veenendaal.

Formal analysis: Gillian Ritchie, David Senyszak.

Investigation: John T. Murchison, Gillian Ritchie, David Senyszak, Edwin J. R. van Beek.

Methodology: John T. Murchison, Gillian Ritchie, David Senyszak, Joris Wakkie, Edwin J. R. van Beek.

Supervision: John T. Murchison, Edwin J. R. van Beek.

Writing – review & editing: John T. Murchison, Gillian Ritchie, David Senyszak, Jeroen H. Nijwening, Joris Wakkie, Edwin J. R. van Beek.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin*. 2021 May; 71(3):209–249. <https://doi.org/10.3322/caac.21660> Epub 2021 Feb 4. PMID: 33538338.
2. The National Lung Screening Trial Research. Results of Initial Low-Dose Computed Tomographic Screening for Lung Cancer. <https://doi.org/10.1056/NEJMoa1209120> [Internet]. 2013 May 22 [cited 2021 Sep 15];368(21):1980–91. Available from: <https://www.nejm.org/doi/10.1056/NEJMoa1209120>
3. Becker N, Motsch E, Trotter A, Heussel CP, Dienemann H, Schnabel PA, et al. Lung cancer mortality reduction by LDCT screening—Results from the randomized German LUSI trial. *International Journal of Cancer*. 2020; 146(6). <https://doi.org/10.1002/ijc.32486> PMID: 31162856
4. de Koning HJ, van der Aalst CM, de Jong PA, Scholten ET, Nackaerts K, Heuvelmans MA, et al. Reduced Lung-Cancer Mortality with Volume CT Screening in a Randomized Trial. *New England Journal of Medicine*. 2020; 382(6). <https://doi.org/10.1056/NEJMoa1911793> PMID: 31995683
5. Oudkerk M, Devaraj A, Vliegenthart R, Henzler T, Prosch H, Heussel CP, et al. European position statement on lung cancer screening. Vol. 18, *The Lancet Oncology*. 2017. [https://doi.org/10.1016/S1470-2045\(17\)30861-6](https://doi.org/10.1016/S1470-2045(17)30861-6) PMID: 29208441
6. Kakinuma R, Ohmatsu H, Kaneko M, Eguchi K, Naruke T, Nagai K, et al. Detection failures in spiral CT screening for lung cancer: Analysis of CT findings. *Radiology*. 1999;212(1).
7. White CS, Romney BM, Mason AC, Austin JHM, Miller BH, Protopapas Z. Primary carcinoma of the lung overlooked at CT: Analysis of findings in 14 patients. *Radiology*. 1996; 199(1). <https://doi.org/10.1148/radiology.199.1.8633131> PMID: 8633131
8. Kakinuma R, Ashizawa K, Kobayashi T, Fukushima A, Hayashi H, Kondo T, et al. Comparison of sensitivity of lung nodule detection between radiologists and technologists on low-dose CT lung cancer screening images. *British Journal of Radiology*. 2012; 85(1017). <https://doi.org/10.1259/bjr.75768386> PMID: 22919013
9. Nair A, Gartland N, Barton B, Jones D, Clements L, Screation NJ, et al. Comparing the performance of trained radiographers against experienced radiologists in the UK lung cancer screening (UKLS) trial. *British Journal of Radiology*. 2016; 89(1066). <https://doi.org/10.1259/bjr.20160301> PMID: 27461068
10. Baldwin DR, Callister MEJ. The British Thoracic Society guidelines on the investigation and management of pulmonary nodules. *Thorax*. 2015; 70(8).
11. McKee BJ, Regis SM, McKee AB, Flacke S, Wald C. Performance of ACR lung-RADS in a clinical CT lung screening program. *Journal of the American College of Radiology*. 2015; 12(3).
12. MacMahon H, Naidich DP, Goo JM, Lee KS, Leung ANC, Mayo JR, et al. Guidelines for management of incidental pulmonary nodules detected on CT images: From the Fleischner Society 2017. *Radiology*. 2017 Jul 1; 284(1):228–43. <https://doi.org/10.1148/radiol.2017161659> PMID: 28240562
13. 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*. 2017 Nov 1; 285(2):584–600. <https://doi.org/10.1148/radiol.2017162894> PMID: 28650738
14. Devaraj A, van Ginneken B, Nair A, Baldwin D. Use of volumetry for Lung nodule management: Theory and Practice 1 STATE OF THE ART: Volumetry for Lung Nodule Management Devaraj et al. *radiology.rsna.org n Radiology*. 2017; 284(3).
15. Heuvelmans MA, Walter JE, Vliegenthart R, van Ooijen PMA, de Bock GH, de Koning HJ, et al. Disagreement of diameter and volume measurements for pulmonary nodule size estimation in CT lung cancer screening. *Thorax*. 2018; 73(8). <https://doi.org/10.1136/thoraxjnl-2017-210770> PMID: 29056601
16. Naidich DP, Bankier AA, MacMahon H, Schaefer-Prokop CM, Pistolesi M, Goo JM, et al. Recommendations for the management of subsolid pulmonary nodules detected at CT: A statement from the Fleischner Society. Vol. 266, *Radiology*. 2013. <https://doi.org/10.1148/radiol.12120628> PMID: 23070270
17. Armato SG, McLennan G, Mcnitt-Gray MF, Meyer CR, Yankelevitz D, Aberle DR, et al. Lung image database consortium: Developing a resource for the medical imaging research community. *Radiology*. 2004; 232(3). <https://doi.org/10.1148/radiol.2323032035> PMID: 15333795
18. Lo SCB, Freedman MT, Gillis LB, White CS, Mun SK. Computer-aided detection of lung nodules on CT with a computerized pulmonary vessel suppressed function. *American Journal of Roentgenology*. 2018; 210(3).
19. Roos JE, Paik D, Olsen D, Liu EG, Chow LC, Leung AN, et al. Computer-aided detection (CAD) of lung nodules in CT scans: Radiologist performance and reading time with incremental CAD assistance. *European Radiology*. 2010; 20(3). <https://doi.org/10.1007/s00330-009-1596-y> PMID: 19760237

20. Brown MS, Lo P, Goldin JG, Barnoy E, Kim GHJ, McNitt-Gray MF, et al. Toward clinically usable CAD for lung cancer screening with computed tomography. *European Radiology*. 2014; 24(11). <https://doi.org/10.1007/s00330-014-3329-0> PMID: 25052078
21. Wagner AK, Hapich A, Psychogios MN, Teichgräber U, Malich A, Papageorgiou I. Computer-Aided Detection of Pulmonary Nodules in Computed Tomography Using ClearReadCT. *Journal of Medical Systems*. 2019; 43(3). <https://doi.org/10.1007/s10916-019-1180-1> PMID: 30706143
22. Benzakoun J, Bommart S, Coste J, Chassagnon G, Lederlin M, Boussouar S, et al. Computer-aided diagnosis (CAD) of subsolid nodules: Evaluation of a commercial CAD system. *European Journal of Radiology*. 2016; 85(10). <https://doi.org/10.1016/j.ejrad.2016.07.011> PMID: 27666609
23. Liu K, Li Q, Ma J, Zhou Z, Sun M, Deng Y, et al. Evaluating a Fully Automated Pulmonary Nodule Detection Approach and Its Impact on Radiologist Performance. *Radiology: Artificial Intelligence*. 2019; 1(3). <https://doi.org/10.1148/ryai.2019180084> PMID: 33937792
24. Zhao YR, van Ooijen PMA, Dorrius MD, Heuvelmans M, de Bock GH, Vliegenthart R, et al. Comparison of three software systems for semi-automatic volumetry of pulmonary nodules on baseline and follow-up CT examinations. *Acta Radiologica*. 2014; 55(6). <https://doi.org/10.1177/0284185113508177> PMID: 24132766
25. Jeon KN, Goo JM, Lee CH, Lee Y, Choo JY, Lee NK, et al. Computer-aided nodule detection and volumetry to reduce variability between radiologists in the interpretation of lung nodules at low-dose screening computed tomography. *Investigative Radiology*. 2012; 47(8). <https://doi.org/10.1097/RLI.0b013e318250a5aa> PMID: 22717879
26. Kim H, Park CM, Gwak J, Hwang EJ, Lee SY, Jung J, et al. Effect of CT reconstruction algorithm on the diagnostic performance of radiomics models: A task-based approach for pulmonary subsolid nodules. *American Journal of Roentgenology*. 2019; 212(3). <https://doi.org/10.2214/AJR.18.20018> PMID: 30476456
27. 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. *European Radiology*. 2016; 26(7). <https://doi.org/10.1007/s00330-015-4030-7> PMID: 26443601
28. Setio AAA, Traverso A, de Bel T, Berens MSN, Bogaard C van den, Cerello P, et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge. *Medical Image Analysis*. 2017;42.
29. Zou KH, Warfield SK, Bharatha A, Tempany CMC, Kaus MR, Haker SJ, et al. Statistical Validation of Image Segmentation Quality Based on a Spatial Overlap Index. *Academic Radiology*. 2004; 11(2). [https://doi.org/10.1016/s1076-6332\(03\)00671-8](https://doi.org/10.1016/s1076-6332(03)00671-8) PMID: 14974593
30. Martins Jarnalo CO, Linsen PVM, Blazís SP, van der Valk PHM, Dickerscheid DBM. Clinical evaluation of a deep-learning-based computer-aided detection system for the detection of pulmonary nodules in a large teaching hospital. *Clinical Radiology*. 2021.