IMAGE & SIGNAL PROCESSING

CrossMark

Computer-Aided Detection of Pulmonary Nodules in Computed Tomography Using ClearReadCT

Anne-Kathrin Wagner [1,2] · Arno Hapich [3] · Marios Nikos Psychogios [4] · Ulf Teichgräber [1] · Ansgar Malich [2] ·
Ismini Papageorgiou [1,2]

Abstract

This study evaluates the accuracy of a computer-aided detection (CAD) application for pulmonary nodular lesions (PNL) in computed tomography (CT) scans, the ClearReadCT (Riverain Technologies). The study was retrospective for 106 biopsied PNLs from 100 patients. Seventy-five scans were Contrast-Enhanced (CECT) and 25 received no enhancer (NECT). Axial reconstructions in soft-tissue and lung kernel were applied at three different slice thicknesses, 0.75 mm (CECT/NECT $n = 25/6$), 1.5 mm ($n = 18/9$) and 3.0 mm ($n = 43/18$). We questioned the effect of (1) enhancer, (2) kernel and (3) slice thickness on the CAD performance. Our main findings are: (1) Vessel suppression is effective and specific in both NECT and CECT. (2) Contrast enhancement significantly increased the CAD sensitivity from 60% in NECT to 80% in CECT, $P = 0.025$ Fischer's exact test. (3) The CAD sensitivity was 84% in 3 mm slices compared to 68% in 0.75 mm slices, $P > 0.2$ Fischer's exact test. (4) Small lesions of low attenuation were detected with higher sensitivity. (5) Lung kernel reconstructions increased the false positive rate without affecting the sensitivity ($P > 0.05$ McNemar's test). In conclusion, ClearReadCT showed an optimized sensitivity of 84% and a positive predictive value of 67% in enhanced lung scans with thick, soft kernel reconstructions. NECT, thin slices and lung kernel reconstruction were associated with inferior performance.

Keywords nodule classification · segmentation · vessel suppression · background elimination · lung cancer

Highlights
- ClearReadCT performs vessel subtraction in both contrasted and non-contrasted scans
- The optimized CAD sensitivity was 84% with a PPV of 67% in thick slab, soft kernel, contrast-enhanced images
- ClearReadCT is designed for the detection of small lesions with low attenuation values

This article is part of the Topical Collection on *Image & Signal Processing*

✉ Ismini Papageorgiou
ismini.e.papageorgiou@gmail.com

1 Institute of Diagnostic and Interventional Radiology, University Hospital Jena, Am Klinikum 1, 07747 Jena, Germany

2 Institute of Radiology, Südharz Hospital Nordhausen, Dr.-Robert-Koch street 39, 99734 Nordhausen, Germany

3 Department of Thoracic Surgery, Südharz Hospital Nordhausen, Dr.-Robert-Koch street 39, 99734 Nordhausen, Germany

4 Institute of Diagnostic and Interventional Neuroradiology, University Medicine Göttingen, Robert Koch street 40, 37075 Göttingen, Germany

Springer

Abbreviations

16xDEF	Somatom definition AS 16-row CT scanner
16xEMO	Somatom emotion 16-row CT scanner
64xDEF	Somatom definition AS 64-row CT scanner
CAD	Computer-aided detection
CECT	Contrast-enhanced computed tomography
CT	Computed tomography
FN	False negative
FP	False positive
HU	Hounsfield Units
NECT	Non-enhanced computed tomography
PACS	Picture archiving and communication system
PNL	Pulmonary nodular lesions
PPV	Positive predictive value
TN	True negative
TP	True positive
TPR	True positive rate (sensitivity)
V1	CAD version 1
V2	CAD version 2

Introduction

Pulmonary nodular lesions (PNL) are focal, round or oval areas of increased lung opacity smaller than 30 mm that are not associated with pneumonia, atelectasis or lymphadenopathy [1]. The increasing popularity of the computed tomography (CT) in lung diagnostics augments the amount of PNL detections either as incidental findings or as part of cancer screening programs, which are widely supported by many recent international thoracic community positioning statements [2–4]. Not only the first detection [5] but also the rigorous PNL follow-up process are issues of immense public health interest due to the malignant or premalignant PNL-potential in up to 4% of the cases [2, 6].

Computer-aided Detection (CAD) software applications adapt the technological advances in image analysis to the expanding needs and challenges of the quantitative radiology. Various, custom-made or commercially available CAD systems are implemented in the clinical practice, aiming to reduce the reading time, increase the PNL detection sensitivity and objectify the PNL volumetry [7].

In this study, we evaluate the diagnostic accuracy of ClearReadCT (ClearRead™CT, Riverain™ Technologies, Miamisburg, OH, USA) in a retrospective database with a histological ground truth. ClearReadCT implements a pre-trained machine learning algorithm for the PNL segmentation in lung CT scans and was reported to detect 80% of all lung cancers [8]. The segmentation process is facilitated by a background reduction algorithm that subtracts ("suppresses") the vascular structures in both contrast-enhanced (CECT) and non-enhanced (NECT) CT scans [9]. This study aims to define (1) the CAD diagnostic accuracy and the detection rate of

malignant lesions, (2) the scanning parameters that optimize the CAD-performance and (3) the fidelity of the background reduction method.

Materials and methods

Patient selection and study design

The study was retrospective for the time range between 6.6.2007 and 9.2.2016. 1042 patients received a lung CT with the clinical indication "tumor" and/or "pulmonary nodule". A histological ground truth was the single inclusion criterion to select 100 patients aged 63.50 ± 10.05 years with 106 biopsied nodules. Fifty patients were confirmed as bronchial cancer and 11 patients had a metastatic disease of various origin; the rest 39 patients were diagnosed with a benign lesion. Patients with a second or a third, non-biopsied nodule received follow-up CT-scans. Thirteen such additional nodules without histological identity were included in the study as well. The flow of participants is illustrated in Fig. 1. The single exclusion criterion was the technical incompatibility of some datasets with the CAD-software for unknown reasons (detailed reported in Supplementary Table 1). Two different CAD-versions were tested, V1 and V2. Both tested versions were release candidates and the V2 is currently available as commercial software. The scans that were incompatible with only one of the two CAD-releases have been included for the evaluation of the compatible version but excluded from the paired tests between versions.

Image acquisition

Lung CT-imaging was performed using multi-slice spiral scanners: a Somatom Definition AS plus 64-row CT-scanner (64xDEF, $N = 82$), a Somatom Emotion 16-row CT-scanner (16xEMO, $N = 15$ patients) and a Somatom Definition 16-row CT-scanner (16xDEF, $N = 3$), all purchased from Siemens (Siemens Healthcare GmbH, Erlangen, Germany). The scanning protocols are reported in detail in (Table 1). Seventy-five lung CT scans were Contrast-Enhanced (CECT) with Ultravist 300® (Bayer Vital GmbH, Leverkusen, Germany), 1 ml/kg i.v. at 3 ml/min flow rate followed by 30 ml of saline chaser. The bolus was tracked in the ascending aorta with a threshold of 100 Hounsfield Units (HU). Twenty-five patients received no contrast enhancer (Non-Enhanced CT, NECT).

Image analysis

The observational image analysis was performed by two radiologists: an assistant with two to five years of experience and a consultant with more than 15 years of experience in analyzing lung CT-scans. Due to lack of access to
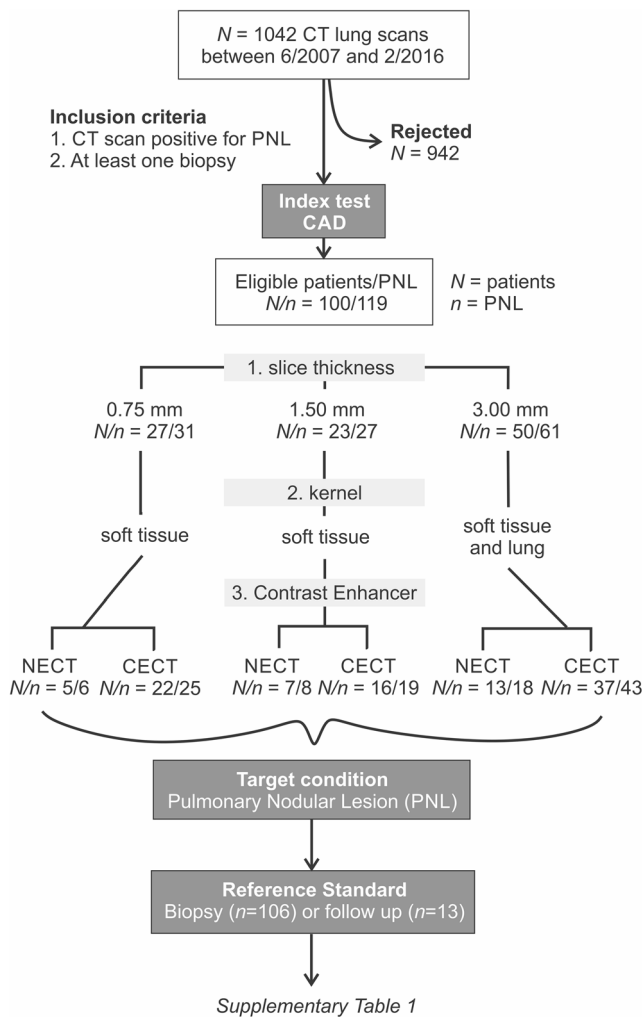
Fig. 1 Standards of Reporting of Diagnostic Accuracy (STARD) flow diagram. Note that the STARD diagram summarizes all available patients/nodules. Deviations between STARD and the "total PNL" number in Supplementary Table 1 reflect the patients/lesions that were not processed for technical reasons ("patient not readable")

the separate observer's results, the evaluation was based on the common report. PNLs were thoroughly described

Table 1 CT Imaging technical characteristics

Scanner model	64xDEF	16xEMO	16xDEF
Tube voltage (kV)	120-130	120-130	120-130
Pitch factor	1.2	1.25	1.25
Collimation	38.4	19.2	19.2
Soft tissue kernel	B31f	B41f/s	B41f/s
Lung kernel	B70f	B70f	B70f
Image size (pixel)	512 × 512	512 × 512	512 × 512

Siemens scanner models: 64-row Somatom Definition AS (64xDEF), 16-row Somatom Emotion (16xEMO) and a 16-row Somatom Definition AS (16xDEF). The preoperative imaging protocol included homogeneous scanning parameters and a standard contrast enhancer type, dose and flow rate wherever applicable

and at least one PNL-per-patient was subjected to biopsy within three months from the diagnosis. ClearReadCT was coupled to the SECTRA Picture Archiving and Communication System (PACS) Version 15.1.22.2 (Sectra AB, Linköping, Sweden) and the dedicated Sectra Radiological Information System (RIS) Version 4.3.1.27 (Sectra AB). The CAD has an integrated background subtraction step that recognizes and "suppresses" the vascular tree in non-enhanced (Fig. 2a-d) and contrast-enhanced scans (Fig. 2e-h). A descriptive report of the PNL diameter, volume, and mean density is integrated into the graphical report (Fig. 2c, d, g, h). The maximum diameter was kept as a PNL-size metric in order to conform to the widely accepted guidelines [10].

For estimating the diagnostic accuracy, CAD segmentations were classified as follows:
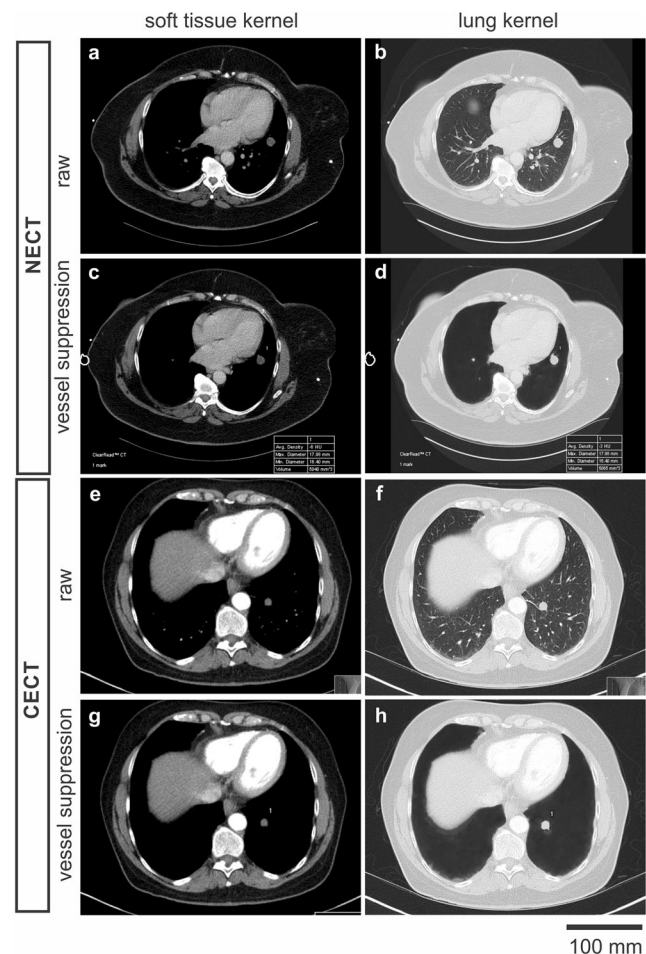

Fig. 2 Vessel suppression with and without contrast enhancer. Sample images of non-enhanced conmputed tomography (NECT) lung scan raw data reconstructed in soft-tissue (a) and lung (b) kernel. In (c) and (d) sample images after vessel suppression. For all segmented areas ClearReadCT provides a volumetric and densitometric report (lower right corner in c and d). Sample images of contrast-enhanced computed tomography (CECT) lung scan raw data reconstructed in soft-tissue (e) and lung (f) kernel. In (g) and (h) sample images after vessel suppression

1. True Positives (TP) were the segmentations with a malignant or benign ground truth validated by histology or by consecutive scans
2. False Positives (FP) were the unspecific segmentations that did not correspond to PNLs. FP received an observational post-hoc identity by two radiologists with intermediate and large experience in interpreting lung scans, both blinded to the initial report.
3. False Negatives (FN) were the non-segmented PNLs with a malignant or benign ground truth.
4. Additional PNLs (oversights) were segmented by the CAD system. These were retrospectively confirmed by two radiologists with intermediate and large experience in interpreting lung scans, both blinded to the initial report.

Unremarkable scans were not included in the current study.

Statistical analysis and graphics

For the statistical analysis, we used the Statistical Package for Social Sciences (SPSS, IBM Corporation, NYC, USA). The chi-squared test and Fischer's exact test were applied for non-parametric data. ClearReadCT V1 and V2 and different kernels were processed pairwise with the McNemar's test. The linear regression model, Pearson's and Spearman's algorithm were applied for correlation analysis of parametric data. Parametric data differences were tested with one-way ANOVA or with a Mann-Whitney rank-sum test. Receiver Operative Curve (ROC) analysis was used to assess the CAD sensitivity in relation to the PNL volume and opacity. Statistics are powered to 80% for $\alpha = 0.05$. Normality was proofed using the Shapiro-Wilk method. The unweighted Cohen's kappa coefficient was used to evaluate the inter-rater agreement between CAD versions. Box plots illustrate the median and inner quartiles, error bars include the 5th-95th percentile and dots correspond to outliers. Graphical work and halftones were processed with the freeware platform Inkscape 0.92.

Results

ClearReadCT is a commercially available CAD system with a pre-trained, machine learning classifier for PNL detection in CT scans with or without contrast enhancement. In this retrospective diagnostic accuracy study, we aimed to (1) proof the fidelity of two CAD releases (V1 and V2) and (2) define the scanning and image reconstruction parameters for optimal CAD-performance, such as the slice thickness, kernel and contrast-enhancement.

Vessel suppression is efficient in contrast-enhanced and non-enhanced images

ClearReadCT implements a background subtraction algorithm that detects and subtracts the vascular tree from CECT and NECT, soft and lung kernel reconstructions, referred to as "vessel suppression" (Fig. 2). We asked whether vessel suppression could "suppress" pulmonary nodules as well, hence jeopardizing the detection of potential cancers. From a total of 53 non-segmented (FN) nodules (V1, V2 and not readable scans), only one was eliminated during the vessel suppression step – the rest was detectable in the background-suppressed images (Fig. 2c, d, g, h) and the segmentation failed for other reasons. Interestingly, most FNs were localized juxtapleural (29/53, 54.72%), 11 FN were juxtahilar (20.76%), 10 FN juxtavascular (18.87%) and only in a small FN proportion was adjacent to scar tissue (3/53, 5.67%).

ClearReadCT versions reveal a low inter-rater agreement

The retrospective dataset of this study derives from three CT-scanners (Siemens) with different technical features: a 64-row model (64xDEF) and two different 16-row models (16xEMO and 16xDEF). Patients were distributed as 82/15/3 amongst 64xDEF/16xEMO/16xDEF. In order to merge data from different scanners we compared the PNLs for attenuation and volume, finding no apparent differences between different scanner ($P > 0.05$ Kruskal-Wallis ANOVA on ranks).

V1 and/or V2 segmented together 387 PNLs (Supplementary Table 2), albeit with a poor overlap. Cohen's statistic reveals no significant agreement between ClearReadCT V1 and V2 (55.97%) with an unweighted kappa coefficient of −0.02. The character of the non-overlapping segmentations between versions was independent of the scanner type, contrast enhancement, volume or attenuation.

CAD-assisted volumetry

The maximum diameter (Dmax in mm) was favored as a PNL-size metric in order to conform to the clinical guidelines for risk stratification and PNL follow-up [10]. However, CAD-implementation allows for a volumetric analysis that might be beneficial towards the classical metrics. A volume-maximum diameter regression study (Supplementary Fig. 1 a) reveals a highly significant volume-diameter correlation for both software versions, which supports that the Dmax-based statistics correlate well to the volume-based statistics in this database, $P << 0.001$, ANOVA. Nevertheless, the low goodness of fitting (Supplementary Fig. 1 a) for both V1 and V2 (R^2 0.525 and 0.607, respectively) points towards PNL shape irregularities which might convey relevant information. A shape analysis revealed that the distribution of the CAD-

detected PNLs did not differ between the tested Versions (Supplementary Fig. 1 b and 1c), $P > 0.05$, Mann-Whitney rank-sum test between V1 and V2.

CAD sensitivity is optimal in thick-slab, soft-tissue kernel reconstructions of contrast-enhanced CT scans

We opted to analyze how the slice thickness, contrast enhancement and reconstruction kernel could influence the diagnostic accuracy of ClearReadCT. For this we compared volume data reconstructed with soft or lung kernel at 0.75 mm, 1.5 mm and 3.0 mm slabs in CECT and NECT (Supplementary Table 1). Technically incompatible scans are referred to as "patient not readable".

The sensitivity (true positive rate, TPR) of V1 in soft-tissue CECT was 68/74/84% for a slice thickness of 0.75/1.5/3.0 mm respectively. The performance of V2 in CECT was comparable to V1, with a TPR of 86/68/83% for the corresponding slice thickness. The TPR was significantly reduced in NECT, 33/75/63% for V1 and 67/38/63% for V2, $P = 0.025$, Fischer's exact test with Yates correction (Supplementary Table 1 and Fig. 3a). Despite the positive trend, the slice thickness did not significantly influence the TPR, $P > 0.05$ Fischer's exact test (Supplementary Table 1 and Fig. 3 a). Overall there was no statistically significant benefit of the optimized V2 towards V1, $P > 0.05$ McNemar's test for CECT, NECT, and all slice thicknesses (Supplementary Table 1).

Thick slabs (3.0 mm) were reconstructed in both soft-tissue and lung kernel. The TPR in CECT scans was inferior in the lung- (maximum 63%) compared to the soft-tissue kernel (maximum 84%). However, the lung kernel allowed for a better PNL segmentation in NECT scans (TPR approximately 83% in lung versus 63% in soft-tissue NECT). The increased TPR in NECT lung kernel data was achieved, however, at the cost of an increased false positive rate (Supplementary Table 1, Fig. 3b).
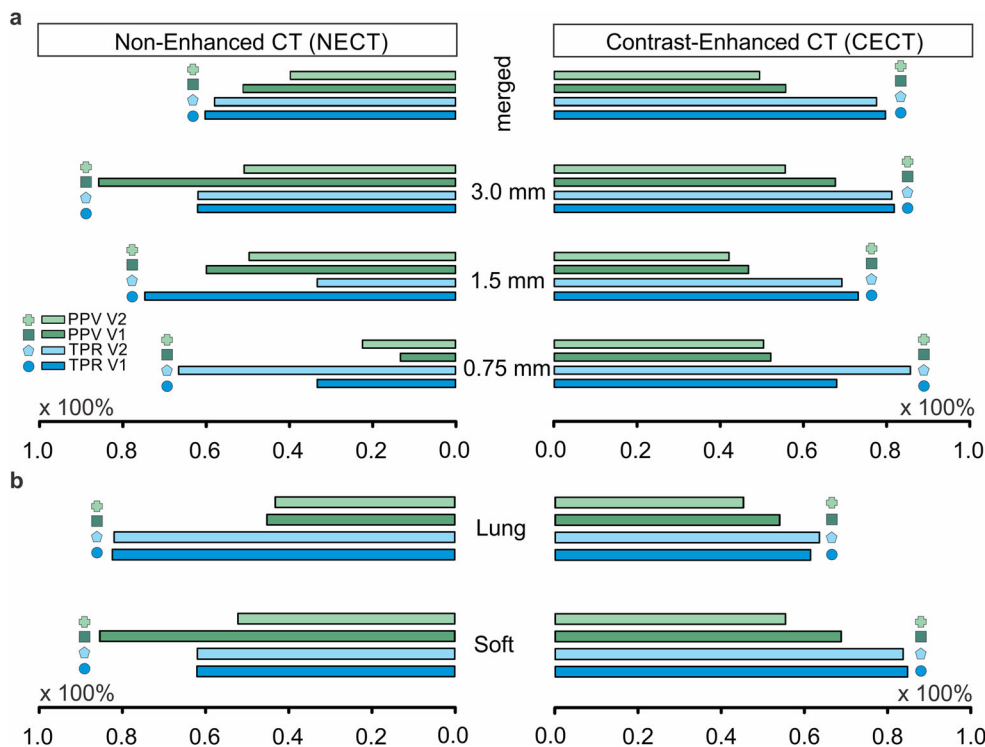

Fig. 3 **a** ClearReadCT sensitivity and positive predictive value. Tornado plot for the true positive rate (TPR, blue) and positive predictive value (PPV, green) of ClearReadCT for PNL segmentation in soft-tissue kernel reconstructed CT scans. We compared different slice thickness (0.75, 1.5 and 3.0 mm) reconstructions with (CECT, right-handed bar plots) and without (NECT, left-handed bar plot) contrast-enhancement. Across different slice thickness, the difference was not statistically significant in CECT and NECT, $P > 0.05$, Fischer's exact test. Versions V1 and V2 performed equally when tested in similar conditions (kernel, slice thickness), $P > 0.05$, McNemar's test. Comparison between CECT and NECT (all slice thickness merged) revealed a significant sensitivity advantage of V2 upon contrast application, $P = 0.025$, Fischer's exact test. V1 showed only a mild trend with no statistical significance for better sensitivity in CECT scans ($P = 0.137$, Fischer's exact test). **b** Lung kernel increases the segmentation rate of false positives without influencing the sensitivity. Tornado plot for the TPR (blue) and PPV (green) of ClearReadCT for PNL segmentation in soft-tissue and lung kernel reconstructed CT scans of 3.0 mm slice thickness. We compared CECT (right-handed bar plots) and NECT (left-handed bar plot). Versions V1 and V2 performed equally when tested in similar conditions (image contrast, kernel), $P > 0.05$, McNemar's test. The kernel comparison revealed no statistically significant advantage of the soft-tissue over the lung kernel, $P > 0.05$, McNemar's test. Reconstruction in lung kernel increased the false positive rate thus negatively influencing the PPV especially in NECT (Supplementary Table 1). V1, CAD version 1; V2, CAD version 2

Summarizing the above, PNL detection can be optimized to a TRP of 84% and a PPV of 67% by using ClearReadCT in thick, soft-tissue reconstructed CECT scans. While the TPR was positively influenced by the contrast-enhancement, the slice thickness was no significant modulator. Lung kernel was not beneficial in CECT but increased the CAD sensitivity in NECT at the cost of an increased false positive rate. Both CAD versions were implemented with equal results.

Oversight detection

ClearReadCT segmented many overlooked PNLs. Oversights were retrospectively verified by two radiologists blinded to the original report. The dignity of oversights is unknown; therefore they are dealt with as "additional PNLs" (Supplementary Table 1). In the thin slab (0.75 and 1.5 mm) soft kernel reconstructions, ClearReadCT segmented approximately 40-50% and 20% oversights in CECT/NECT. Increasing the slice thickness to 3.0 mm reduced the additional PNLs to 6-14%, which might partly reflect the increased PNL conspicuity and the consecutive improved observer's sensitivity in thicker reconstructions [11]. The lung kernel increased the proportion of oversights in thick slabs to approximately 20 – 30% (compared to 6-14% in the soft-tissue kernel of the same patients) in both CECT and NECT (Supplementary Table 1). Interestingly, the oversights (abbreviated as ADD in Fig. 4a and b) were significantly smaller in diameter and less opaque compared to the PNLs detected by both the observer and CAD (TP). Moreover, a shape analysis reveals that oversights were rather irregular PNLs (Supplementary Fig. 1b and c, Table 2).

Consecutively, CAD is advantageous as a second-read method for the detection of small, irregular, subsolid and ground-glass oversights, especially in thin slice and lung kernel reconstructions. The distinction between oversights and FP segmentations remains however a critical process that should be carefully reviewed by a qualified radiologist.

ClearReadCT selectively detects small nodules with attenuation values lower than 0 HU

CAD-classifiers might show different sensitivities for particular lesion sizes and opacities [12]. In order to reveal a possible volume-class predilection of the ClearReadCT classifier, we analyzed the CAD sensitivity and specificity for different nodule volumes. As a convention and for compatibility reasons with the existing literature, we used the maximal diameter as an indirect volume indicator. Regardless of the CAD version applied, the median FN-diameter was above 30 mm (Fig. 4a), the median FN-opacity around 0 HU (Fig. 4b) and both were significantly higher compared to the corresponding TP-segmentations, $P < 0.0001$ Mann-Whitney rank-sum test. Accordingly, ClearReadCT is sensitive for the detection of small nodules with attenuation values lower than 0 HU (Table 2).

Fig. 4 ClearReadCT segments small nodules with low attenuation values. Box plots for the maximal diameter (a) and opacity (b) of the segmented (true positive, TP), non-segmented (false negative, FN) and additionally detected (ADD) nodules by ClearReadCT V1 and V2. In panel (b), a custom lung window (width 1500 HU, level − 600 HU) and soft tissue window (width 350 HU, level 50 HU) are illustrated as dark and light grey zones next to the y-axis, respectively. The asterisk (*) denotes statistical significance, $P < 0.001$ one-way ANOVA and $P < 0.05$, Dunn's pairwise post-hoc test. Receiver operating curves for opacity and maximal diameter for V1 (c) and V2 (d) reveal that both features are negative modulators of the CAD diagnostic accuracy. A, Area under the curve; HU, Hounsfield Units; V1, CAD version 1; V2, CAD version 2
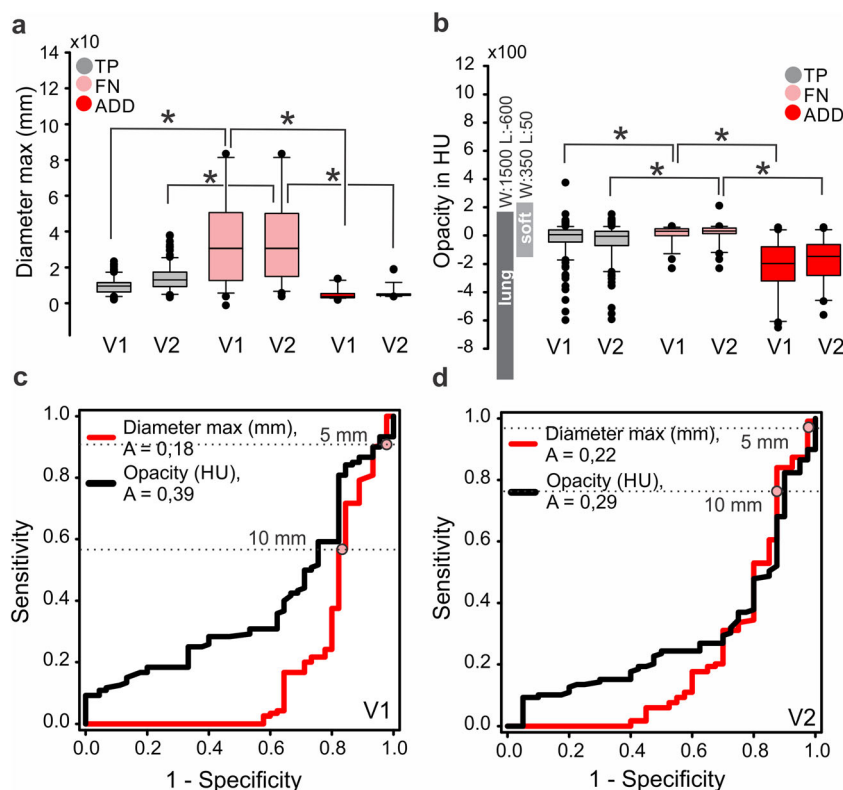
Table 2 Characteristics of detected nodules, CAD and observer's oversights

	True positive nodules (TP)		False negative nodules (FN) = CAD oversights		Additional nodules (ADD) detected by CAD = observer's oversights	
	V1 ($n = 120$)	V2 ($n = 119$)	V1 ($n = 45$)	V2 ($n = 40$)	V1 ($n = 30$)	V2 ($n = 28$)
Size, Dmax (mm)	3.17/24.7, 11.14 ± 0.44	4.45/39.23, 15.68 ± 0.72	0.01/122, 36.27 ± 4.07	5.00/122.0, 39.35 ± 4.68	3.03/15.0, 6.07 ± 0.58	4.93/20.37, 7.46 ± 0.75
Opacity (HU)	−604/813, −23.68 ± 13.62	−598/148, −47.33 ± 12.07	−236/62, −0.56 ± 11.34	−236/1013, 40.45 ± 27.43	−657/−54, −222.7 ± 36.57	−567/−55, −178.07 ± 31.85
Shape index (mm^{-2})	0.6/145.16, 14.71 ± 2.00	1.5/1531.8**, 32.96 ± 12.89	n.a.	n.a.	1.65/136.67, 47.73 ± 5.56	7.3/192.5, 68.08 ± 8.25

Summary table of all nodules detected in soft tissue and lung kernel, with and without enhancer. Nodule features: size as maximal diameter (Dmax) in mm, average opacity in HU and shape index (Dmax/volume in mm^{-2}). For all metrics: $P < 0.001$, one-way ANOVA and $P < 0.05$, Dunn's post-hoc test between TP and TN, TN and ADD, TP and ADD. Between V1 and V2, $P >> 0.05$ for all metrics, Mann-Whitney rank sum test. CAD, Computer-aided Detection; HU, Hounsfield Units; V1, CAD version 1; V2, CAD version 2

* All metrics provided in (min/max, average ± SEM)

** Outlier

ROC illustrate that large PNL diameter and high attenuation values are negative modifiers of the CAD sensitivity with Areas Under Curve (AUC) significantly below 0.5 (Fig. 4c and d). For PNLs with a maximum diameter around 5 mm the sensitivity/ specificity was 90/7% for V1 and 99/3% for V2. For 10 mm lesions, the sensitivity of V1/V2 was 57/78% and the specificity in both cases lower than 15%, reflecting the high FP-rate. The high FP-rate is a major drawback that should be addressed in future versions.

The character of false positives

An important withdrawal of ClearReadCT is the high FP-rate especially in thin, lung kernel slabs (Fig. 3a, b, Supplementary Table 1). FP segmentations (red crosses and pink dots in Fig. 5a) are significantly less opaque compared to the TP ones regardless of their size ($P < 0.05$, 2-tailed t-test or Mann-Whitney rank-sum test) for almost all protocols and CAD-versions (Fig. 5b). The FP segmentations were classified into three groups: (1) infiltration/ scar/atelectasis, (2) vessel or (3) no correlate (Fig. 5c). All categories showed similar opacity values and shared a similar distribution between lung (Fig. 5c, lower panel) and soft-tissue kernel (Fig. 5c, upper panel), $P = 0.52$ One Way ANOVA. Note that the detection of vessel fragments reflects a leaking activity of the vessel suppression algorithm (Fig. 5 c, middle column in both panels). Consequently, FP segmentations occur preferentially at lower attenuation values and their character, predominantly scars, dystelectasis and infiltrates, is not significantly affected by the slice thickness, contrast enhancement or kernel selection.

Discussion

In this retrospective study, we evaluate the diagnostic accuracy of ClearReadCT (Riverain). Our main finding is that the CAD sensitivity can be optimized to 84% with a PPV of 67% by using contrast-enhanced scans with thick slabs and soft-tissue kernel.

ClearReadCT performance is comparable to similar, commercially available CAD applications. In a previous study of diagnostic accuracy in low-dose CT scans [5], Lung VCAR (General Electric Deutschland Holding GmbH, Frankfurt am Main, Germany), ImageChecker® CT lung System (R2 Technology Inc., Sunnyvale, CA, USA), Syngovia Via Va20 (Siemens) and Cornell Via (Cornell University, Ithaca, NY, USA) showed sensitivities between 56 and 70% for 5 mm PNLs and 74-82% for PNLs with a maximal diameter of approximately 11 mm. The optimized sensitivity of ClearReadCT for 5 mm and 10 mm lesions was 98/78%, however at the cost of a very low specificity (5% and 15%). Prakashini et al. [13] evaluated the CT Lung Nodule Assessment (LNA) and Comparison Option (Philips GmbH, Hamburg, Germany) in 1.4 mm slabs from a 64-row scanner and found a TPR/PPV of approximately 91/67%. It is worth noticing that the cumulative TPR/PPV of ClearReadCT for an equal slab thickness was 75/45%.

The high FP-rate and the low PPV in the current and previously published CAD metrics bring into attention the persisting and challenging problem of the FP detections by CAD systems [7]. The tested CAD software (ClearReadCT, Riverain) implements a pre-processing step for vascular tree subtraction to increase the signal-to-noise ratio and thus facilitate the effectiveness of the classifier. The background subtraction algorithm showed precise results in both contrast-enhanced and non-contrasted scans without suppressing the juxtavascular PNLs. Although some "vessel remnants" were detected as a low "leaking" activity of the background suppression algorithm, ClearReadCT FPs were in the vast majority unspecific, small size atelectasis or scars with low attenuation values. Many research groups suggest innovating algorithms for increasing the CAD specificity while preserving the
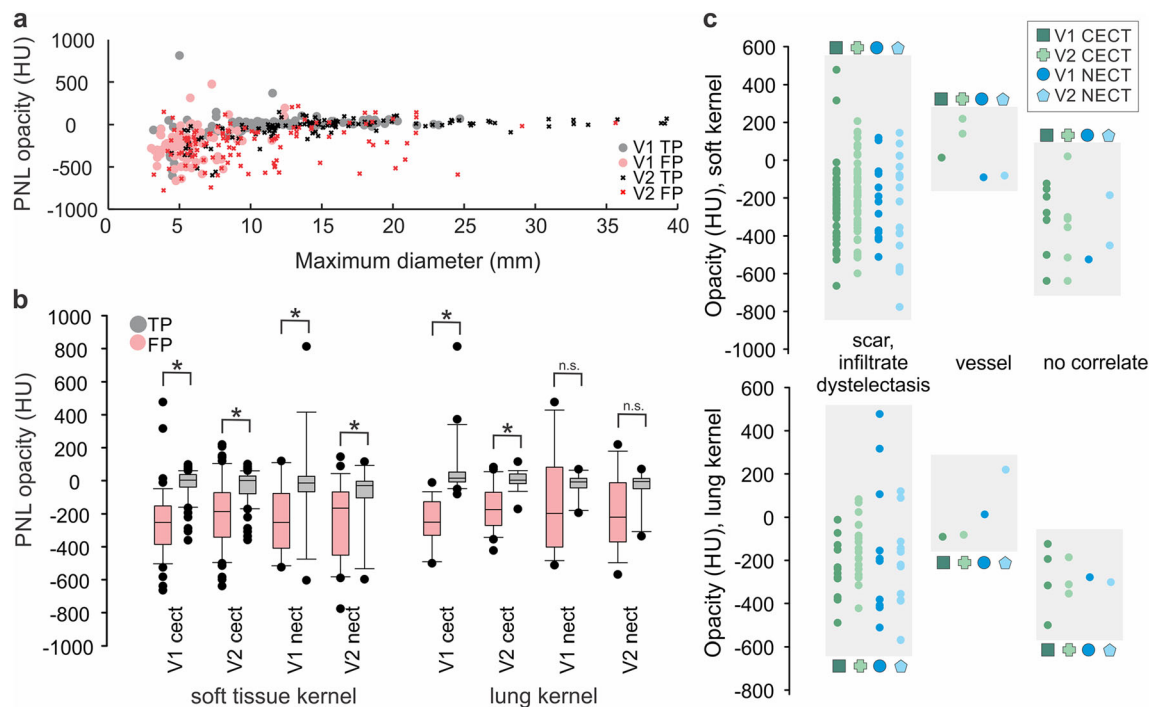
Fig. 5 Character and features of false positives **a** Scatter plot of the maximum diameter (mm) and opacity (HU) for True Positive (TP) and False Positive (FP) PNLs of mixed dignity as segmented by ClearReadCT V1 and V2. FP lesions show a lower density compared to the TP ones and are scattered across lower attenuation values regardless of their size, $P < 0.001$, Mann-Whitney rank-sum test. **b** Box plot for the opacity of FP (*pink*) and TP (*grey*) lesions in the soft-tissue kernel (*left-handed panel*) and lung kernel (*right-handed panel*). FP segmentations reveal significantly lower opacity compared to TP ones in the soft-tissue kernel and CECT lung kernel reconstructions, *$P < 0.05$ Mann-Whitney

rank-sum test. In NECT, no significant opacity difference was detected between TP and FP, $P > 0.05$ Mann-Whitney rank-sum test; n.s., non-significant. **c** The character of FP lesions was determined by observation. Vertical scatter plot for FPs in soft-tissue (*upper panel*) and lung kernel (*lower panel*) for CECT (*green*) and NECT (*blue*) scans. $P = 0.52$ One Way ANOVA, power < 0.80. Note that some vessels "escape" the vessel suppression algorithm and are segmented as FP lesions in both soft-tissue and lung kernel. CECT, Contrast-Enhanced Computed Tomography; NECT, Non-Enhanced Computed Tomography; HU, Hounsfield Units; V1, CAD version 1; V2, CAD version 2

high sensitivity [14–16]. Deep learning solutions, 2D and 3D convolutional neuronal networks [17–19] and radiomic approaches [20] are independently suggested to achieve sensitivities above the 95th percentile for pulmonary lesion segmentation in reference, big-data databases. The plethora of suggested solutions and the exponentially rising number of publications signify the refractoriness of the problem and the abundance of open challenges in the CAD field. Some authors recognize individualities in different nodule size-groups and suggest combinations of classifiers, each one dedicated to a particular size-group [12]. An interesting approach for follow-up scans was recently published by the group of Terasawa et al. implementing temporal subtraction of consecutive scans as a solution for background subtraction and automated calculation of the VDT [21]. This method was effectively implemented for the detection of vertebral metastasis [22], the resilience of the lung parenchyma and its high shape-dependence on the respiratory movements might, however, perplex the applicability of temporal subtraction in PNL detection [21].

The current approach reveals some important features of ClearReadCT that might improve its clinical

implementation. By using biopsied PNLs we eliminated the ground truth bias. The acquired level of evidence is nevertheless hampered by the retrospective character of the study and relatively low number of patients in some subgroups (NECT, thin slab reconstructions, Supplementary Table 1). The main results of this study are powered to 80% and support a clinically meaningful result. Within the disadvantages of the current study one should mention the lack of iterative image reconstruction which could improve the CAD metrics [23, 24]. Moreover, this study does not include low-dose lung scans. Since the low-dose lung CT defines a new trajectory in the field of lung cancer screening [25–28] the diagnostic accuracy of ClearReadCT should be tested in a corresponding database in a future study. Finally, a withdrawal of the current study is the unbalanced distribution of the PNL size with the opacity, dignity and neighboring structures. The team of Kobayashi et al. approaches this problematic by suggesting a uniform randomized nodular lesion injection model into unremarkable scans [29]. This standardization allows for a more precise and unbiased quantification of different CAD classifiers.

Conclusions

ClearReadCT showed a sensitivity of 84% with a positive predictive value of 67% when applied to contrast-enhanced CT scans with thick slab, soft kernel reconstructions. The implemented classifier is sensitive for small nodules with attenuation values lower than 0 HU and was efficient in oversight detection especially in thin slab reconstructions. The major withdrawal is the universal CAD-problem of a high FP rate [7]. The need for a post-hoc observer evaluation restricts the CAD applicability to a second read mode. Similar clinical studies shall provide the developers with an unbiased feedback towards CAD-optimization and promote the smooth integration of machine learning applications into the clinical practice.

Compliance with the ethical standards

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Jeong, Y. J., Yi, C. A., and Lee, K. S., Solitary pulmonary nodules: detection, characterization, and guidance for further diagnostic workup and treatment. AJR Am. J. Roentgenol. 188(1):57–68, 2007.
2. Ruparel, M. et al., Pulmonary nodules and CT screening: the past, present and future. Thorax 71(4):367–375, 2016.
3. Baldwin, D. R., Ten Haaf, K., Rawlinson, J., and Callister, M. E. J., Low dose CT screening for lung cancer. BMJ 359:j5742, 2017.
4. Oudkerk, M. et al., European position statement on lung cancer screening. Lancet Oncol. 18(12):e754–e766, 2017.
5. Liang, M. et al., Low-Dose CT Screening for Lung Cancer: Computer-aided Detection of Missed Lung Cancers. Radiology 281(1):279–288, 2016.
6. Horeweg, N. et al., Lung cancer probability in patients with CT-detected pulmonary nodules: a prespecified analysis of data from the NELSON trial of low-dose CT screening. Lancet Oncol. 15(12):1332–1341, 2014.
7. Girvin, F., and Ko, J. P., Pulmonary Nodules: Detection, Assessment, and CAD. Am. J. Roentgenol. 191(4):1057–1069, 2008.
8. Lo, S. B., Freedman, M. T., Gillis, L. B., White, C. S., and Mun, S. K., JOURNAL CLUB: Computer-Aided Detection of Lung Nodules on CT With a Computerized Pulmonary Vessel Suppressed Function. Am. J. Roentgenol. 210(3):480–488, 2018.
9. Milanese, G., Eberhard, M., Martini, K., Martini, I. V. D., and Frauenfelder, T., Vessel suppressed chest Computed Tomography for semi-automated volumetric measurements of solid pulmonary nodules. Eur. J. Radiol. 101:97–102, 2018.
10. Bankier, A. A. et al., Recommendations for Measuring Pulmonary Nodules at CT: A Statement from the Fleischner Society. Radiology 285(2):584–600, 2017.
11. Rubin, G. D., Lung Nodule and Cancer Detection in CT Screening. J. Thorac. Imaging 30(2):130–138, 2015.
12. Gupta, A., Saar, T., Martens, O., and Moullec, Y. L., Automatic detection of multisize pulmonary nodules in CT images: Large-scale validation of the false-positive reduction step. Med. Phys. 45(3):1135–1149, 2018.
13. Prakashini, K., Babu, S., Rajgopal, K. V., and Kokila, K. R., Role of Computer Aided Diagnosis (CAD) in the detection of pulmonary nodules on 64 row multi detector computed tomography. Lung India Off Organ Indian Chest Soc 33(4):391–397, 2016.
14. Wang, Z. et al., Improved lung nodule diagnosis accuracy using lung CT images with uncertain class. Comput. Methods Prog. Biomed. 162:197–209, 2018.
15. Ali, I. et al., Lung Nodule Detection via Deep Reinforcement Learning. Front. Oncol. 8:108, 2018.
16. Nibali, A., He, Z., and Wollersheim, D., Pulmonary nodule classification with deep residual networks. Int. J. Comput. Assist. Radiol. Surg. 12(10):1799–1808, 2017.
17. Jin, H., Li, Z., Tong, R., and Lin, L., A deep 3D residual CNN for false-positive reduction in pulmonary nodule detection. Med. Phys. 45(5):2097–2107, 2018.
18. da Silva, G. L. F., Valente, T. L. A., Silva, A. C., de Paiva, A. C., and Gattass, M., Convolutional neural network-based PSO for lung nodule false positive reduction on CT images. Comput. Methods Prog. Biomed. 162:109–118, 2018.
19. Gierada, D. S. et al., Quantitative CT Classification of Lung Nodules: Initial Comparison of 2D and 3D Analysis. J. Comput. Assist. Tomogr. 40(4):589–595, 2016.
20. Ma, J. et al., Computerized detection of lung nodules through radiomics. Med. Phys. 44(8):4148–4158, 2017.
21. Terasawa, T. et al., Detection of lung carcinoma with predominant ground-glass opacity on CT using temporal subtraction method. Eur. Radiol. 28(4):1594–1599, 2018.
22. Iwano, S. et al., Thoracic Temporal Subtraction Three Dimensional Computed Tomography (3D-CT): Screening for Vertebral Metastases of Primary Lung Cancers. PLoS One 12(1):e0170309, 2017.
23. Jin, S. et al., Lung nodules assessment in ultra-low-dose CT with iterative reconstruction compared to conventional dose CT. Quant Imaging Med Surg 8(5):480–490, 2018.
24. Nomura, Y. et al., Effects of Iterative Reconstruction Algorithms on Computer-assisted Detection (CAD) Software for Lung Nodules in Ultra-low-dose CT for Lung Cancer Screening. Acad. Radiol. 24(2):124–130, 2017.
25. National Lung Screening Trial Research Team et al., Reduced lung-cancer mortality with low-dose computed tomographic screening. N. Engl. J. Med. 365(5):395–409, 2011.
26. National Lung Screening Trial Research Team et al., Results of initial low-dose computed tomographic screening for lung cancer. N. Engl. J. Med. 368(21):1980–1991, 2013.
27. Yousaf-Khan, U. et al., Final screening round of the NELSON lung cancer screening trial: the effect of a 2.5-year screening interval. Thorax 72(1):48–56, 2017.
28. Field, J. K. et al., UK Lung Cancer RCT Pilot Screening Trial: baseline findings from the screening arm provide evidence for the potential implementation of lung cancer screening. Thorax 71(2):161–170, 2016.
29. Kobayashi, H. et al., A method for evaluating the performance of computer-aided detection of pulmonary nodules in lung cancer CT screening: detection limit for nodule size and density. Br. J. Radiol. 90(1070):20160313, 2017.