

ORIGINAL ARTICLE

Open Access



Artificial intelligence-based automated matching of pulmonary nodules on follow-up chest CT

Nicola Fink^{1,2*} , Jonathan I. Sperl³, Johannes Rueckel^{1,4}, Theresa Stüber^{1,5}, Sophia S. Goller¹, Jan Rudolph¹, Felix Escher¹, Theresia Aschauer¹, Boj F. Hoppe¹, Jens Ricke¹ and Bastian O. Sabel^{1,2,6}

Abstract

Background The growing demand for follow-up imaging highlights the need for tools supporting the assessment of pulmonary nodules over time. We evaluated the performance of an artificial intelligence (AI)-based system for automated nodule matching.

Methods In this single-center study, patients with nodules and ≤ 2 chest computed tomography (CT) examinations were retrospectively selected. An AI-based algorithm was used for automated nodule detection and matching. The matching rate and the causes for incorrect matching were evaluated for the ten largest lesions (5–30 mm in diameter) registered on baseline CT. The dependence of the matching rate on nodule number and localization was also analyzed.

Results One hundred patients (46 females), with a median age of 62 years (interquartile range 57–69), and 253 CTs were included. Focusing on the ten largest lesions, 1,141 lesions were identified, of which 36 (3.2%) were other structures incorrectly identified as nodules (false-positives). Of the 1,105 identified nodules, 964 (87.2%) were correctly detected and matched. The matching rate for nodules registered in both baseline and follow-up scans was 97.8%. The matching rate per case ranged 80.0–100.0% (median 90.0%). Correct matching rate decreased in follow-up examinations to over 50 nodules ($p = 0.003$), with an overrepresentation of missed matching. Matching rates were higher in parenchymal (91.8%), peripheral (84.4%), and juxtavascular (82.4%) nodules than in juxtaphrenic nodules (71.1%) ($p < 0.001$). Missed matching was overrepresented in juxtavascular, and incorrect assignment in juxtaphrenic nodules.

Conclusion The correct automated-matching rate of metastatic pulmonary nodules in follow-up examinations was high, but it depends on localization and a number of nodules.

Relevance statement The algorithm enables precise follow-up matching of pulmonary nodules, potentially providing a solid basis for standardized and accurate evaluations. Understanding the algorithm's strengths and weaknesses based on nodule localization and number enhances the interpretation of AI-based results.

Key Points

- The AI algorithm achieved a correct nodule matching rate of 87.2% and up to 97.8% when considering nodules detected in both baseline and follow-up scans.
- Matching accuracy depended on nodule number and localization.
- This algorithm has the potential to support response evaluation criteria in solid tumor-based evaluations in clinical practice.

*Correspondence:

Nicola Fink

nicola.fink@med.uni-muenchen.de

Full list of author information is available at the end of the article

Keywords Artificial intelligence, Diagnostic techniques and procedures, Lung neoplasms, Multiple pulmonary nodules, Tomography (x-ray computed)

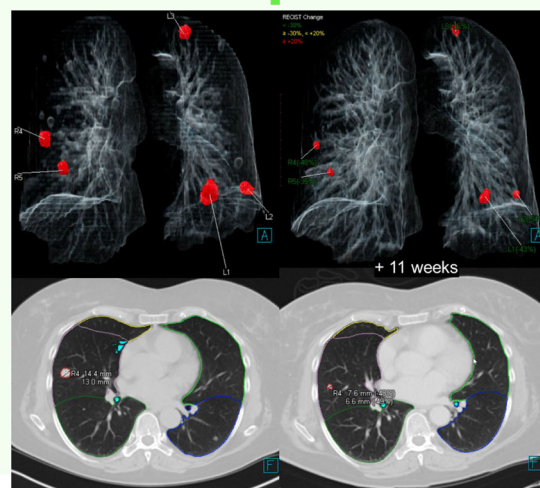
Graphical Abstract

Artificial intelligence-based automated matching of pulmonary nodules on follow-up chest CT

ESR
EUROPEAN SOCIETY
OF RADIOLOGY

- 100 patients and 253 chest CTs were analysed using an artificial intelligence (AI)-based algorithm for automated nodule detection and matching.
- The AI achieved a 97.8% correct nodule matching rate in follow-up scans.
- Matching accuracy depends on nodule number, remaining high with < 50 nodules.
- Nodule localization affected accuracy, with better matching in parenchymal, peripheral, and juxtavascular nodules.

This AI algorithm enabled precise matching of pulmonary nodules in follow-up CT scans



European
Radiology
EXPERIMENTAL

**Eur Radiol Exp (2025) Fink N, Sperl JI, Rueckel J et al;
DOI: 10.1186/s41747-025-00579-w**

Background

Computed tomography (CT) is considered the standard of care for detecting pulmonary nodules [1]. In recent years, CT technology has undergone several improvements, including higher image resolution and low-dose acquisition, resulting in a more sensitive detection of pulmonary nodules and a higher number of scans being performed [2, 3]. Thus, radiologists are increasingly confronted with incidental findings, including pulmonary nodules that may require follow-up to further determine their nature [4–6]. Although approximately 95% of those nodules are benign [5, 7], it is crucial to assess all of them appropriately, if necessary with follow-up CT scans [4], in order to detect the malignant ones at an early stage.

Furthermore, the incidence of lung cancer continues to rise, further driving the need for follow-up CT examinations to monitor pulmonary nodules. This gets even more relevant in nodules detected as part of CT-based lung cancer screening performed in high-risk individuals. Several randomized trials have demonstrated the added value of lung cancer screening using low-dose CT by reducing lung cancer mortality significantly [8–13] and

thus support its implementation for a clearly defined risk population. At the same time, this potentially results in a relatively high number of false-positive results [14] and subsequently an increasing number of follow-up examinations that require standardized assessment. A systematic review of, among others, eight randomized lung cancer trials showed that pulmonary nodules were detected in up to 51% of the included patients [15].

Standardized reporting of changes in size, number, and morphologic features plays an important role not only in indeterminate pulmonary nodules [16] but also in assessing treatment response in cancer patients to guide further patient management [17–19]. While computer-aided detection (CAD) systems primarily increase nodule detection sensitivity and already have been shown to reduce the individual reading time [20–22], matching and comparing a pulmonary nodule between two successive examinations is particularly time-consuming and tedious. Ideally, this should be conducted according to standardized reporting systems, such as the lung imaging reporting and data system (Lung-RADS) [23] and the response evaluation criteria in solid tumors (RECIST)

[24]. In this context, it has already been demonstrated that the involvement of a specialized technologist, performing the subsequent measurement during follow-up instead of the radiologist, reduces radiologists' reading time by 87% [25]. Using an artificial intelligence (AI) algorithm instead of an additional human would make this task even more cost-effective, as AI-based automated matching also improves diagnostic efficiency [26].

Hence, the fully automatic identification of the same nodule(s) over multiple time points, hereinafter referred to as nodule matching, has the potential to further improve the follow-up of pulmonary nodules, and thus may increase radiologists' efficiency and accuracy when interpreting serial chest CT scans.

This study aimed to evaluate the performance of an AI-based system performing automated matching of pulmonary nodules in successive CT scans of the chest and to assess its dependence on nodule localization and number.

Methods

The protocol of this retrospective, single-center study was approved by the local institutional Review Board, which waived requirements for informed consent (see "Declarations").

Patient population

Data from patients with an underlying malignant disease, at least one pulmonary nodule, and two or three successive chest CT scans were retrospectively and randomly selected. Patient selection was performed by screening our Picture Archiving and Communication System (PACS) using keywords indicative of pulmonary nodules. The cases were manually reviewed to confirm the presence of pulmonary nodules and to ensure that all inclusion criteria were met. Cases with exclusion criteria were removed from this cohort. Baseline characteristics of the study cohort were derived from medical records, including patient demographics and the underlying malignant disease. The exclusion criteria were as follows: (i) age < 18 years and (ii) thoracic surgery in between the included CT scans. Due to the inclusion criteria, included nodules were primarily metastatic.

Image acquisition and reconstruction

CT scans were performed on Somatom Force, Somatom Definition AS+, Somatom Definition Flash, Somatom Drive (all Siemens Healthineers, Forchheim, Germany), or Optima 660 (GE Healthcare, Chicago, IL, USA) scanners. Patients underwent either an unenhanced high-resolution or a contrast-enhanced CT scan using a standardized institutional protocol. Key parameters included a tube voltage of 100–120 kVp, tube current modulation

(100–300 mAs), rotation time of 0.5–0.6 s, and pitch of 0.9–1.2. Reconstructed in-plane resolution for axial images ranged from 0.5×0.5 mm to 0.7×0.7 mm, depending on the scanner and patient size. Images were reconstructed using iterative reconstruction techniques, including advanced modeled iterative reconstruction—ADMIRE on Siemens Healthineers scanners and Adaptive Statistical Iterative Reconstruction—ASiR on the GE Optima 660 scanner. Standard soft tissue kernels were used for general assessment, while high-spatial-resolution lung kernels were applied for detailed nodule evaluation. Image analysis was conducted on a RadiForce RX370 monitor (EIZO, Hakusan, Ishikawa, Japan) (3 megapixels, Digital Imaging and Communications in Medicine (DICOM)-calibrated, maximum brightness 1,000 cd/m², contrast ratio 1,000:1) under electronically controlled optimal lighting conditions (ambient light < 25 lux).

Automatic nodule matching

All data sets were analyzed using the cloud-based prototype "AI-Rad Companion Research Chest CT Explore" (Siemens Healthineers), involving four subsequent steps:

1. AI-based detection of lung nodules (LungCAD, VD20);
2. AI-based extraction of anatomical landmarks, including a mesh of the lung lobes [27–29];
3. affine co-registration of the landmarks over multiple time points using an iterative closest point algorithm [30]; and
4. identification of nodule pairs in the coregistered space.

The third step is a data-minimalistic approach suitable for cloud-based processing where the full three-dimensional data set of the prior study is no longer available in the cloud at the time point of the follow-up exam. In the fourth step, the Euclidean distance is calculated for the coregistered nodules, and a pair is considered a match if the distance is below 15 mm. This distance threshold was established by analyzing the matching performance on an independent cohort. Figure 1 illustrates a sample case for the visualization of the AI-based results. This algorithm was a prototype at the time of initial evaluation but is now commercially available as part of the AI-Rad Companion Chest CT product (version VA20, Siemens Healthineers). The nodule detection component has been shown to have high sensitivity and specificity in prior studies [31–33].

Overall, the underlying AI is the anatomical landmark detection, as well as lung lobe segmentation. Anatomical landmark detection has been trained on over 15,000 CT data sets [27]. The lung lobe segmentation U-Net has been trained on over 5,000 CT data sets [31].

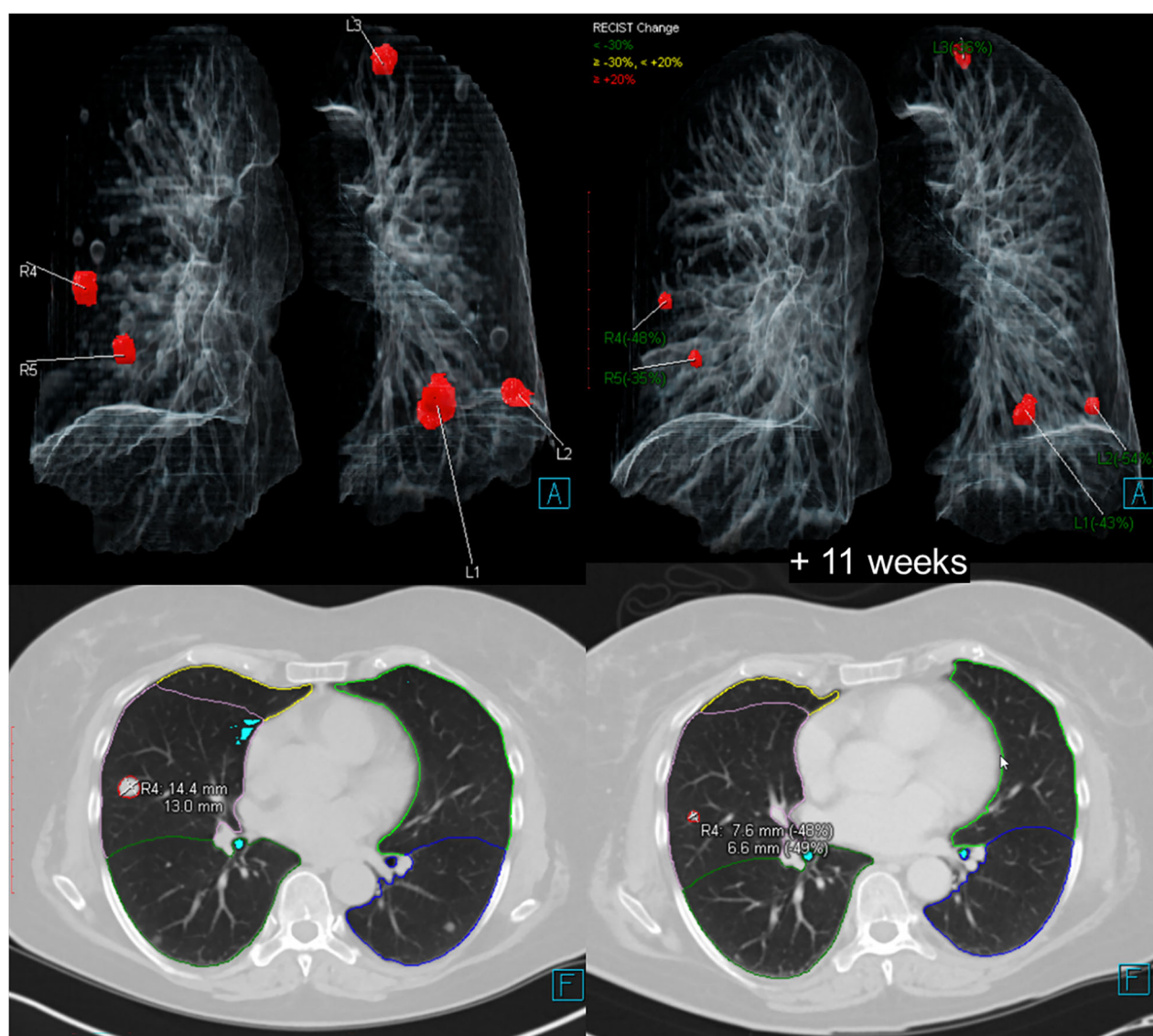


Fig. 1 Sample case of the visualization of the AI-based results for automated matching

Radiological reading and algorithm performance

Regardless of whether the included patients received two or three CT scans, each follow-up was evaluated separately, resulting in 153 follow-up examinations. The entire process of nodule detection and segmentation was fully automated. The validation of the AI algorithm's results was performed by a radiologist with an experience of five years in thoracic imaging (N.F.), whose role was to identify false positives, to verify whether the nodules matched by the AI corresponded to the same nodules on follow-up scans and to identify the root causes of incorrect matching. Nodules missed during detection were not included in the analysis, as the primary aim of this study was to assess the algorithm's

matching performance. The radiologist received the anonymized CT data as axial reconstructions, as well as AI results (Fig. 1), and was allowed to perform various analyses, utilizing standard tools available in a conventional DICOM viewer, such as size measurement, region-of-interest placement, and density analysis. The analysis was focused on a maximum of ten pulmonary nodules (size between 5 mm and 30 mm in diameter) found by the AI algorithm in the baseline CT scan. Nodules that had completely disappeared on follow-up CT and subsolid nodules, including mixed (solid and subsolid) nodules, were excluded. For each lesion assigned as "incorrectly detected/matched" the root cause was documented as follows:

Table 1 Type of CT examinations and radiation doses

	Total scans (n = 253)
Use of intravenous contrast agent	
Unenhanced	51 (20.2)
Contrast-enhanced	202 (79.8)
Initial examination range	
Neck and thorax	3 (1.2)
Neck, thorax, and abdomen	36 (14.2)
Thorax and abdomen	129 (51.0)
Thorax	85 (33.6)
Radiation dose	
CTDI (mGy)	6.7 (5.1–8.5)
DLP	387.0 (267.9–527.6)

Values are median with interquartile range or n (%)
CTDI/ Computed tomography dose index, DLP Dose length product

1. false-positive detection (other, non-nodular structures, such as atelectasis, vertebral bodies, or ribs, incorrectly identified as pulmonary nodules by the algorithm)
2. missed matching (nodule registered at baseline but not in follow-up CT despite its presence); and
3. incorrect assignment (nodule present in both baseline and follow-up scans but incorrectly linked to a different nodule instead of its true counterpart).

In addition, the total number of pulmonary nodules was documented, and the selected pulmonary nodules were classified according to their localization as parenchymal, peripheral, juxtavascular, or juxtaphrenic. Peripheral nodules were defined as being located within 10 mm of the pleura, juxtavascular and juxtaphrenic nodules as directly adjacent to a vessel and the diaphragm, respectively.

Statistical analysis

The software GraphPad Prism (version 8.4.2, GraphPad, San Diego, CA, USA) and R-Studio (version 1.4.1717, RStudio Inc., Boston, MA, USA) were used for statistical analyses and graphical illustration. Continuous variables are reported as median with interquartile range, categorical variables as absolute frequencies and relative frequencies.

Based on a binary outcome (correct or incorrect), we evaluated the matching rate depending on nodule number and localization. The association between nodule number and matching rate, as well as between nodule localization and matching rate, each including root causes for incorrect matching, has been screened for independence by χ^2 analysis based on contingency tables including Pearson residuals. Mann–Whitney *U*-test was used to evaluate differences in matching rates between follow-ups with different numbers of pulmonary nodules.

Any test result with a *p*-value smaller than 0.05 was considered statistically significant.

Results

Patient population

The study population included 100 patients (46 females) with a median age of 62 years (interquartile range 57–69), with an underlying malignant disease and pulmonary nodules. The most frequent origin of the underlying malignant disease was colorectal (*n* = 24), followed by thyroid carcinoma (*n* = 17), sarcoma (*n* = 12%), and malignant melanoma (*n* = 11). Six patients had signs of lung emphysema, but none of the included patients had lung fibrosis. A total of 253 thoracic CT scans were included (Table 1). With some patients having two and some three CT scans, this resulted in a total of 153 follow-up examinations. The median time between scans per follow-up was 13.7 (range 11.1–18.5) weeks. Detailed baseline characteristics of all included patients are shown in Table 2.

Automatic nodule matching

Focused on the ten largest lesions registered in baseline CT scan per follow-up examination by the algorithm, a total of 1,141 lesions were included for further analysis. Among these, 36 lesions (3.2%) were false positives, meaning they were other, non-nodular structures (*e.g.*, atelectasis, vertebral bodies, ribs) incorrectly identified as pulmonary nodules in the baseline scan. Among the remaining 1,105 pulmonary nodules, 141 (12.8%) were not correctly matched at follow-up, with the following root causes: 119 (10.8%) were missed at follow-up (detected at the baseline but not in the follow-up scan despite being present; “missed matching”) and 22 (2.0%) were incorrectly assigned (two different nodules were erroneously matched; “incorrect assignment”). This resulted in 964 nodules that were both correctly identified at baseline and correctly matched at follow-up, leading to a correct nodule matching rate of 87.2% (964/1105). The median correct matching rate per follow-up was 90.0% (80.0–100.0%).

Additionally, considering only nodules that were detected as pulmonary nodules in both CT scans (baseline and follow-up), a total of 986 nodules were included. Among these, 964 were correctly matched, resulting in a correct nodule matching rate of 97.8% (964/986).

Figure 2 shows a flowchart that illustrates nodule categorization. The median size of included pulmonary nodules was 9.0 [IQR 7.1–11.7] mm.

Performance depending on the number and localization of pulmonary nodules

Among the 153 included follow-up examinations, 93 (60.8%) had less than 20 and 31 (20.3%) had 20 to 50, and

Table 2 Patient and pulmonary lesion characteristics

Clinical characteristics		Total (n = 100)
Age in years, median (IQR)		62.0 (57.0–69.0)
Sex, n (%)	Female	46 (46.0)
	Male	54 (54.0)
Underlying malignant disease, n (%)	Colorectal cancer	24 (24.0)
	Thyroid cancer	17 (17.0)
	Sarcoma	12 (12.0)
	Malignant melanoma	11 (11.0)
	Hepatocellular/cholangiocarcinoma	6 (6.0)
	Pancreatic cancer	6 (6.0)
	Urologic cancer	6 (6.0)
	Breast cancer	5 (5.0)
	Pharyngeal cancer	3 (3.0)
	Hemangioendothelioma	3 (3.0)
	Gynecological cancer	2 (2.0)
	Upper gastrointestinal cancer	2 (2.0)
	Cancer of unknown primary	1 (1.0)
Pulmonary lesions	Parotid cancer	1 (1.0)
	Prostate cancer	1 (1.0)
	Total	(n = 1,141)
	False-positives	36 (3.2)
	Pulmonary nodules	1,105 (96.8)
	Parenchymal	549 (49.7)
	Peripheral	346 (31.3)
	Juxtavascular	165 (14.9)
	Juxtaphrenic	45 (4.1)

IQR Interquartile range

29 (19.0%) had more than 50 pulmonary nodules. The correct matching rate differed depending on the number of pulmonary nodules ($p = 0.003$). The matching rate was similar in follow-ups with less than 20 nodules (median 100.0%, range 80.0–100.0%) and 20–50 nodules (90.0%, 80.0–100.0%) ($p = 0.372$). It was significantly lower in examinations with more than 50 pulmonary nodules (median 80.0%, 65.0–90.0%) than in those with < 20 nodules ($p < 0.001$) or those with 20–50 nodules ($p = 0.025$). Figure 3 illustrates independence testing of nodule number per follow-up and matching rate, as well as root causes for incorrect matching. Missed matching was significantly overrepresented in follow-ups with more than 50 nodules and significantly underrepresented in follow-ups with less than 20 nodules. At the same time, false-positive detection of non-nodule structures was overrepresented in follow-ups with less than 20 and underrepresented in those with more than 50 nodules.

Among the pulmonary nodules correctly identified as such on baseline CT ($n = 1,105$), 549 (49.7%) were parenchymal, 346 (31.3%) peripheral, 165 (14.9%) juxtavascular, and 45 (4.1%) juxtaphrenic nodules. The correct matching rate of pulmonary nodules differed significantly depending on the localization ($p < 0.001$). While parenchymal nodules were correctly matched in 91.8% (504/549), the matching rate of peripheral and juxtavascular nodules was 84.4% (292/346) and 82.4% (136/165), respectively. Juxtaphrenic nodules showed the lowest matching rate (71.1%, 32/45). Figure 4 shows independency testing of the nodule localization and the matching rate, as well as the root causes for incorrect matching. While mis-matching was significantly underrepresented in parenchymal nodules, it was overrepresented in juxtavascular nodules. Incorrect assignment of two non-identical nodules was significantly overrepresented in juxtaphrenic nodules.

Discussion

This study evaluated the performance of the cloud-based prototype AI-Rad Companion Research Chest CT Explore (Siemens Healthineers) for automated matching of pulmonary nodules in serial CT scans. The main findings are as follows: first, this algorithm showed a high correct matching rate; second, the correct matching rate was significantly lower in examinations with more than 50 pulmonary nodules; third, the correct matching rate was significantly higher in parenchymal, peripheral, and juxtavascular than in juxtaphrenic nodules with an overrepresentation of missed matching in juxtavascular and incorrect assignment in juxtaphrenic nodules.

Previous studies evaluated such automated nodule matching in a limited number of patients (from 11 to 40) [34–36], whereas this study investigated matching rates, as well as detailed root causes of decreasing accuracy in a larger cohort of 100 patients. With a median per case matching rate of 90.0%, the algorithm evaluated in the present study showed high performance regarding automated nodule matching in a heterogeneous cohort in terms of localization and the number of pulmonary nodules. Considering only lesions in which the algorithm registered nodules on both CT scans, the nodule matching rate was 97.8%.

This high matching accuracy has important clinical implications. Reliable tracking of nodule changes over time supports treatment decisions by ensuring consistency in monitoring growth or stability and facilitates adherence to standardized protocols such as RECIST and Lung-RADS. Additionally, automated matching reduces the risk of false positives and false negatives, helping to minimize unnecessary interventions and improve diagnostic confidence.

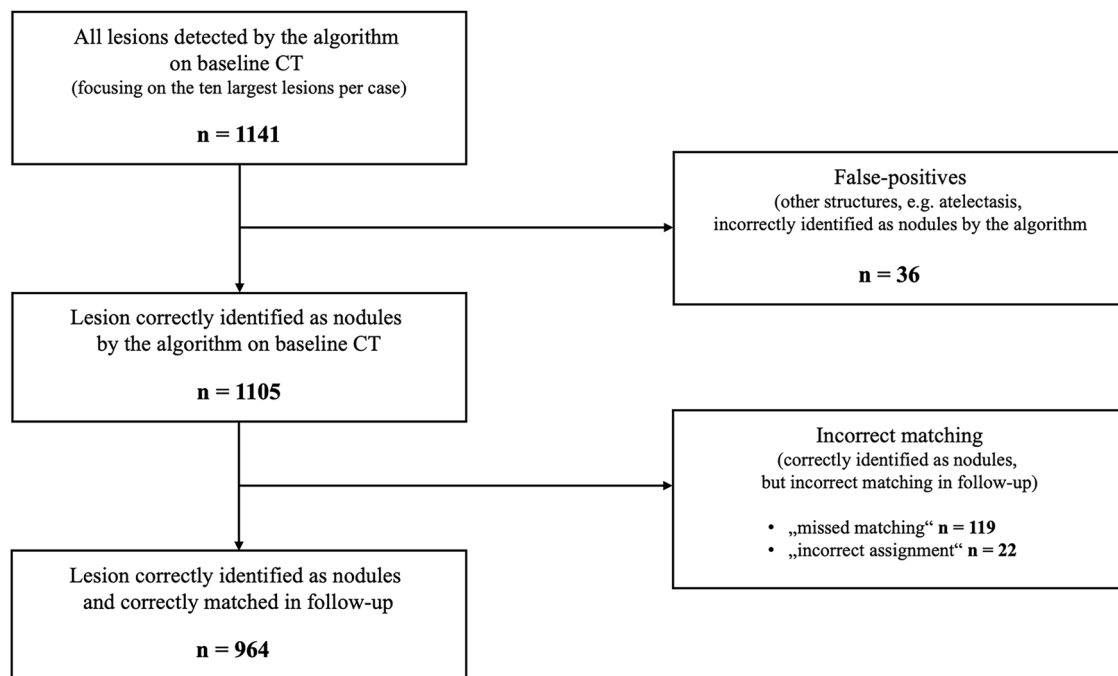


Fig. 2 Flowchart illustrating nodule categorization



Fig. 3 Matching depending on the number of nodules: **a** distribution of the matching (correct versus incorrect) per number of pulmonary nodules illustrated as a balloon plot. **b** Cohen-friendly association plot for the relation between nodule number per follow-up and matching rate (1 = correct; 2 = incorrect), as well as root causes for incorrect matching (2a = missed matching; 2b = incorrect assignment; 2c = false-positive detection) illustrating the deviation from the statistically expected contingency table (over-/underrepresentation in red/blue; independency testing using χ^2 analysis)

Those matching rates are comparable to or even higher than those reported in previous studies, which were 92.7% [36], 86.3% [34], and 66.7% [35]. Tao et al [36] already attributed their relatively high matching rate compared with other studies to differences in the study population: While their population was based on a lung cancer screening cohort, the other two studies included cancer patients in whom the lung parenchyma may be more damaged due to pre-existing conditions and previous

treatments and thus more difficult to assess than in screening patients, most of whom are likely to be relatively healthy. However, in the present study, the correct matching rate did not significantly differ between patients with or without lung parenchymal changes, such as scars and atelectasis. Only a small subset of our study cohort had such conditions, and in cases where they were present, the extent was relatively low. For example, lung fibrosis was not detected in any of the patients. Compared

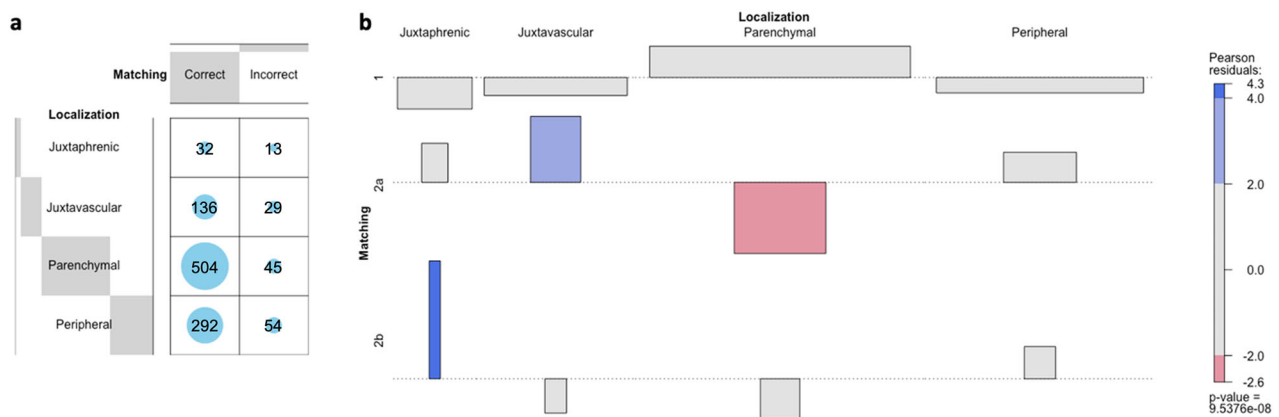


Fig. 4 Matching depending on the localization of nodules: **a** distribution of the matching (correct *versus* incorrect) per localization of the pulmonary nodules illustrated as a balloon plot. **b** Cohen-friendly association plot for the relation between nodule localization and matching rate (1 = correct; 2 = incorrect), as well as root causes for incorrect matching (2a = missed matching; 2b = incorrect assignment) illustrating the deviation from the statistically expected contingency table (over-/underrepresentation in red/blue; independency testing using χ^2 analysis)

to the results presented in the present study, Lee et al [35] obtained a lower matching rate of 66.7% and only succeeded in raising this to a comparable value of 82.4% when considering only patients with unchanged findings.

The matching rate in the present study was dependent on the number of pulmonary nodules, with a significantly lower matching rate in follow-up examinations with more than 50 nodules. Lee et al [35] already demonstrated decreasing accuracy in automated nodule matching with an increasing number of nodules, mainly caused by unmatched (“missed matching”) rather than incorrectly assigned nodules most likely due to the occupancy of lung parenchyma by a large number of nodules that complicate nodule registration by covering anatomic landmarks. We support this finding with a significant overrepresentation of missed matching in follow-ups with more than 50 nodules, due to the increased complexity associated with a larger number of nodules, possibly resulting in a more challenging anatomical registration during image processing. At the same time, follow-up examinations with less than 20 nodules showed an overrepresentation of false-positives, primarily based on the inclusion of the ten largest lesions at baseline: examinations with a low number of nodules are more likely to contain false-positives among these ten lesions.

Our matching rate was highly associated with the nodule’s localization, with the highest rate of 91.8% in parenchymal nodules, mainly surrounded by low-attenuating lung parenchyma, which provides a strong contrast for both initial detection and subsequent matching. Conversely, juxtavascular nodules exhibited a lower matching rate (due to their proximity to blood vessels), as previously reported in CAD system performance studies [37, 38]. The difficulty in detection may

compromise the subsequent registration process. Juxtaphrenic nodules presented the lowest matching rate, primarily due to the effects of diaphragm motion during breathing, explaining the overrepresentation of incorrect assignment. Image-based registration might show superior results in this case, but is not suitable for the cloud-based data-minimalistic approach of the AI-Rad Companion, as the full three-dimensional data set of the baseline exam is not available in the cloud at the time of analysis of the follow-up exam. In contrast to the present study, Beyer et al [34] did not observe differences in matching rates depending on nodule localization. However, they only investigated nodule localization in terms of lung side (right or left lung) and area (upper, middle, or lower third; central or peripheral), not their relation to adjacent structures.

Beyond the high matching accuracy demonstrated in this study, the AI system holds potential for substantial time savings in clinical workflows. Prior research demonstrated that delegating nodule comparison tasks to technologists reduced radiologists’ time by up to 87% [39]. Fully automated AI solutions could achieve similar or greater efficiencies, particularly in high-nodule-count scenarios [25]. Future research should aim to directly quantify these time savings and assess their impact on clinical workflow.

Overall, this automated matching algorithm has the potential to support clinical assessment of incidental and metastatic pulmonary nodules on follow-up examinations. Despite a slight decrease in the correct matching rate in cases with high nodule count or juxtaphrenic nodules, this algorithm works reliably and accurately in most cases. In addition, we included data from multiple CT scanners and imaging protocols, reflecting the

variability encountered in routine clinical practice that potentially influences nodule detection and characterization [40, 41]. Despite these challenges, the algorithm demonstrated robust matching performance, which highlights its applicability in heterogeneous clinical environments.

In a real-world scenario, using this algorithm may not only reduce reporting time and improve reporting quality but also provide a basis for standardized evaluations such as Lung-RADS and RECIST-based reporting, in particular in patients undergoing serial imaging for metastatic disease. This approach would reduce the time radiologists spend visually comparing scans and ensure more standardized and reproducible assessments, and minimize variability in follow-up recommendations, enhancing decision-making and facilitating multidisciplinary communication.

The current AI system presented here is designed as a fully automated solution, but could also be adapted for semi-automated workflows that integrate manual input, such as RECIST-based identification of target lesions. By providing a standardized and reliable matching process, the algorithm may help track changes in nodule size more consistently, supporting more accurate assessments of disease progression or treatment response. Future studies should explore how such AI systems can further streamline follow-up evaluations, enhance decision-making accuracy, and ultimately contribute to improved patient outcomes.

This study has several limitations. First, only the ten largest solid nodules per follow-up with a size between 5 mm and 30 mm were included to enable a systematic and reproducible evaluation of the algorithm's performance. However, since some of the included patients have more than several hundred pulmonary nodules, this may not fully capture the algorithm's ability to match all detected nodules. Automated matching of very small, very large, or ground-glass nodules may be more complicated and should be assessed in future studies. Second, as this study analyzed the correct matching of pulmonary nodules between two successive CT scans, nodules that disappeared in the follow-up scan were excluded from further analysis. Additionally, while longitudinal tracking across multiple follow-ups is clinically important, only a small subset of our cohort allowed for such analysis, which limited the statistical power for meaningful conclusions. Longitudinal evaluations also involve additional complexities, such as managing new or disappearing nodules and accounting for imaging variability, requiring dedicated analytical frameworks. Third, we did not evaluate the impact of inspiration level or nodule size and growth on the matching rate. The analyzed patient cohort did not include cases with fibrosis, and such parenchymal

changes may impact the algorithm's performance, which should be investigated in future studies. In addition, the distribution of malignancies in this study cohort does not specifically reflect cancer prevalences in the general population, which may affect algorithm performance. Fourth, the validation of the algorithm's results was conducted by a single experienced radiologist, who focused on cross-checking the AI results rather than independent nodule identification and analysis. Fifth, analyses were only retrospectively performed and based on single-center data. Sixth, the impact of automated nodule matching on clinical workflow and individual patient management was not investigated and should be part of further analyses. Seventh, this study is based on different CT scanners from different CT vendors, as variability between scans often reflects real-world scenarios due to factors such as scanner availability and location. However, it should be emphasized that this study attempts to address the generalizability of our algorithm performance considering scanner variability. Eighth, some patients had multiple follow-up scans; for example, 53 patients had three consecutive CT scans in our study, which may introduce variance and bias in contrast to patients scanned only once or twice, thus potentially affecting the overall algorithm performance.

In conclusion, this study demonstrated that the accuracy of this algorithm (AI-Rad Companion Research Chest CT Explore [Siemens Healthineers]) in automated matching of pulmonary metastatic nodules is high, when focusing on the ten largest nodules per case. However, the algorithm's performance depends on the localization and number of nodules. Especially the decreasing matching rate in CT scans with high nodule count has to be taken into account when interpreting AI-based results. Overall, the analyzed AI-based solution for automated matching of pulmonary nodules on follow-up chest CT can provide valuable support for standardized reporting in clinical practice.

Abbreviations

AI	Artificial intelligence
CAD	Computer-aided detection
CT	Computed tomography
DICOM	Digital imaging and communications in medicine
Lung-RADS	Lung imaging reporting and data system
RECIST	Response evaluation criteria in solid tumors

Acknowledgements

Large Language Models were not used to write this paper.

Author contributions

NF designed the study, performed data analysis and interpretation, and drafted the manuscript. BOS designed the study, supervised the study's conception, performed data interpretation, and substantially revised the manuscript. JIS assisted with software requirements and revised the manuscript. JRue

supported the study's methodology, contributed to data interpretation, and revised the manuscript. TS conducted statistical analyses and revised the manuscript. SSG and FE substantially revised the manuscript. JRud, TA, BFH, and JRI supported data analysis and revised the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported within a research cooperation with Siemens Healthineers. Open Access funding enabled and organized by Projekt DEAL.

Data availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethical approval and consent to participate

Institutional Review Board (University Hospital LMU Munich, Germany; number 18-399, 07/02/2018) approval was obtained. Written informed consent was waived by the Institutional Review Board.

Consent for publication

Not applicable.

Competing interests

JIS is affiliated with Siemens Healthineers. The remaining authors declare that they have no competing interests.

Author details

¹Department of Radiology, University Hospital, LMU Munich, Munich, Germany. ²Comprehensive Pneumology Center (CPC-M), Member of the German Center for Lung Research (DZL), Munich, Germany. ³Siemens Healthineers, Erlangen, Germany. ⁴Institute of Neuroradiology, University Hospital, LMU Munich, Munich, Germany. ⁵Department of Statistics, Statistical Learning & Data Science, LMU Munich, Munich, Germany. ⁶Department of Radiology, Asklepios Lung Clinic Munich-Gauting, Gauting, Germany.

Received: 10 October 2024 Accepted: 18 March 2025

Published online: 02 May 2025

References

- Bueno J, Landeras L, Chung JH (2018) Updated Fleischner Society guidelines for managing incidental pulmonary nodules: common questions and challenging scenarios. *Radiographics* 38:1337–1350. <https://doi.org/10.1148/rg.2018180017>
- Smith-Bindman R, Miglioretti DL, Johnson E et al (2012) Use of diagnostic imaging studies and associated radiation exposure for patients enrolled in large integrated health care systems, 1996–2010. *JAMA* 307:2400–2409. <https://doi.org/10.1001/jama.2012.5960>
- Smith-Bindman R, Kwan ML, Marlow EC et al (2019) Trends in use of medical imaging in US health care systems and in Ontario, Canada, 2000–2016. *JAMA* 322:843–856. <https://doi.org/10.1001/jama.2019.11456>
- MacMahon H, Naidich DP, Goo JM et al (2017) Guidelines for management of incidental pulmonary nodules detected on CT images: from the Fleischner Society 2017. *Radiology* 284:228–243. <https://doi.org/10.1148/radiol.2017161659>
- Anderson IJ, Davis AM (2018) Incidental pulmonary nodules detected on CT images. *JAMA* 320:2260–2261. <https://doi.org/10.1001/jama.2018.16336>
- Gould MK, Tang T, Liu ILA et al (2015) Recent trends in the identification of incidental pulmonary nodules. *Am J Respir Crit Care Med* 192:1208–1214. <https://doi.org/10.1164/rccm.201505-0990OC>
- McWilliams A, Tammemagi MC, Mayo JR et al (2013) Probability of cancer in pulmonary nodules detected on first screening CT. *N Engl J Med* 369:910–919. <https://doi.org/10.1056/NEJMoa1214726>
- The National Lung Screening Trial Research Team (2011) Reduced lung-cancer mortality with Low-dose computed tomographic screening. *N Engl J Med* 365:395–409. <https://doi.org/10.1056/NEJMoa1102873>
- de Koning HJ, van der Aalst CM, de Jong PA et al (2020) Reduced lung-cancer mortality with volume CT screening in a randomized trial. *N Engl J Med* 382:503–513. <https://doi.org/10.1056/NEJMoa1911793>
- Paci E, Puliti D, Lopes Pegna A et al (2017) Mortality, survival and incidence rates in the ITALUNG randomised lung cancer screening trial. *Thorax* 72:825–831. <https://doi.org/10.1136/thoraxjnl-2016-209825>
- Oudkerk M, Liu S, Heuvelmans MA, Walter JE, Field JK (2021) Lung cancer LDCT screening and mortality reduction—evidence, pitfalls and future perspectives. *Nat Rev Clin Oncol* 18:135–151. <https://doi.org/10.1038/s41571-020-00432-6>
- Pastorino U, Silva M, Sestini S et al (2019) Prolonged lung cancer screening reduced 10-year mortality in the MILD trial: new confirmation of lung cancer screening efficacy. *Ann Oncol* 30:1162–1169. <https://doi.org/10.1093/annonc/mdz117>
- Becker N, Motsch E, Trotter A et al (2020) Lung cancer mortality reduction by LDCT screening—results from the randomized German LUSI trial. *Int J Cancer* 146:1503–1513. <https://doi.org/10.1002/ijc.32486>
- Wender R, Fontham ETH, Barrera E Jr et al (2013) American cancer society lung cancer screening guidelines. *CA Cancer J Clin* 63:106–117. <https://doi.org/10.3322/caac.21172>
- Bach PB, Mirkin JN, Oliver TK et al (2012) Benefits and harms of CT screening for lung cancer: a systematic review. *JAMA* 307:2418–2429. <https://doi.org/10.1001/jama.2012.5521>
- Kim TJ, Kim CH, Lee HY et al (2020) Management of incidental pulmonary nodules: current strategies and future perspectives. *Expert Rev Respir Med* 14:173–194. <https://doi.org/10.1080/17476348.2020.1697853>
- Gohagan JK, Marcus PM, Fagerstrom RM et al (2005) Final results of the lung screening study, a randomized feasibility study of spiral CT versus chest x-ray screening for lung cancer. *Lung Cancer* 47:9–15. <https://doi.org/10.1016/j.lungcan.2004.06.007>
- Swensen S, Jett J, Sloan J et al (2002) Screening for lung cancer with low-dose spiral computed tomography. *Am J Respir Crit Care Med* 165:508–513. <https://doi.org/10.1164/ajrccm.165.4.2107006>
- Rajasooriyar C, Sritharan T, Chenthruran S et al (2020) The role of staging computed tomography on detection of occult metastasis in asymptomatic breast cancer patients. *Cancer Rep* 3:e1247. <https://doi.org/10.1002/cnr.2.1247>
- Yacoub B, Varga-Szemes A, Schoepf UJ et al (2022) Impact of artificial intelligence assistance on chest CT interpretation times: a prospective randomized study. *AJR Am J Roentgenol* 219:743–751. <https://doi.org/10.2214/AJR.22.27598>
- Setio AAA, Traverso A, de Bel T et al (2017) Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge. *Med Image Anal* 42:1–13. <https://doi.org/10.1016/j.media.2017.06.015>
- Murchison JT, Ritchie G, Senyszak D 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:e0266799. <https://doi.org/10.1371/journal.pone.0266799>
- Chelala L, Hossain R, Kazerooni EA et al (2021) Lung-RADS version 1.1: challenges and a look ahead, from the AJR special series on radiology reporting and data systems. *AJR Am J Roentgenol* 216:1411–1422. <https://doi.org/10.2214/AJR.20.24807>
- Eisenhauer EA, Therasse P, Bogaerts J et al (2009) New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* 45:228–247. <https://doi.org/10.1016/j.ejca.2008.10.026>
- Beaumont H, Bertrand AS, Klifa C et al (2019) Radiology workflow for RECIST assessment in clinical trials: Can we reconcile time-efficiency and quality? *Eur J Radiol* 118:257–263. <https://doi.org/10.1016/j.ejrad.2019.07.030>
- Koo CW, Anand V, Girvin F et al (2012) Improved efficiency of CT interpretation using an automated lung nodule matching program. *AJR Am J Roentgenol* 199:91–95. <https://doi.org/10.2214/AJR.11.7522>
- Ghesu FC, Georgescu B, Zheng Y et al (2019) Multi-scale deep reinforcement learning for real-time 3D-landmark detection in CT scans. *IEEE Trans Pattern Anal Mach Intell* 41:176–189. <https://doi.org/10.1109/TPAMI.2017.2782687>

28. Ronneberger O, Fischer P, Brox T (2015) U-Net: convolutional networks for biomedical image segmentation. MICCAI 9351:234–241. https://doi.org/10.1007/978-3-319-24574-4_28
29. Fischer AM, Varga-Szemes A, van Assen M et al (2020) Comparison of artificial intelligence-based fully automatic chest CT emphysema quantification to pulmonary function testing. *AJR Am J Roentgenol* 214:1065–1071. <https://doi.org/10.2214/AJR.19.21572>
30. Rusinkiewicz S, Levoy M (2001) Efficient variants of the ICP algorithm. In: Proceedings of the third international conference on 3-D digital imaging and modeling. Princeton University, pp 145–152. <https://doi.org/10.1109/IM.2001.924423>
31. Chamberlin J, Kocher MR, Waltz J et al (2021) Automated detection of lung nodules and coronary artery calcium using artificial intelligence on low-dose CT scans for lung cancer screening: accuracy and prognostic value. *BMC Med* 19:55. <https://doi.org/10.1186/s12916-021-01928-3>
32. Abadia AF, Yacoub B, Stringer N et al (2022) Diagnostic accuracy and performance of artificial intelligence in detecting lung nodules in patients with complex lung disease: a noninferiority study. *J Thorac Imaging* 37:154–161. <https://doi.org/10.1097/rti.0000000000000613>
33. Kocher MR, Chamberlin J, Waltz J et al (2022) Tumor burden of lung metastases at initial staging in breast cancer patients detected by artificial intelligence as a prognostic tool for precision medicine. *Heliyon* 8:e08961. <https://doi.org/10.1016/j.heliyon.2022.e08962>
34. Beyer F, Wormanns D, Novak C et al (2004) Clinical evaluation of a software for automated localization of lung nodules at follow-up examinations. *Rofo* 176:829–836. <https://doi.org/10.1055/s-2004-813144>
35. Lee KW, Kim M, Gierada DS et al (2007) Performance of a computer-aided program for automated matching of metastatic pulmonary nodules detected on follow-up chest CT. *AJR Am J Roentgenol* 189:1077–1081. <https://doi.org/10.2214/ajr.07.2057>
36. Tao C, Gierada DS, Zhu F et al (2009) Automated matching of pulmonary nodules: evaluation in serial screening chest CT. *AJR Am J Roentgenol* 192:624–628. <https://doi.org/10.2214/AJR.08.1307>
37. Bae KT, Kim JS, Na YH et al (2005) Pulmonary nodules: automated detection on CT images with morphologic matching algorithm—preliminary results. *Radiology* 236:286–293. <https://doi.org/10.1148/radiol.2361041286>
38. Brown MS, Goldin JG, Suh RD et al (2003) Lung micronodules: automated method for detection at thin-section CT—initial experience. *Radiology* 226:256–262. <https://doi.org/10.1148/radiol.2261011708>
39. Hsu HH, Ko KH, Chou YC et al (2021) Performance and reading time of lung nodule identification on multidetector CT with or without an artificial intelligence-powered computer-aided detection system. *Clin Radiol* 76:626.e23–626.e32. <https://doi.org/10.1016/j.crad.2021.04.006>
40. Erdal BS, Demirer M, Little KJ et al (2020) Are quantitative features of lung nodules reproducible at different CT acquisition and reconstruction parameters? *PLoS One* 15:e0240184. <https://doi.org/10.1371/journal.pone.0240184>
41. Yao Y, Su X, Deng L et al (2024) Effects of tube voltage, radiation dose and adaptive statistical iterative reconstruction strength level on the detection and characterization of pulmonary nodules in ultra-low-dose chest CT. *Cancer Imaging* 24:123. <https://doi.org/10.1186/s40644-024-00770-z>

Publisher's Note

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