Lung Cancer Screening and Nodule Detection

Zheng, Sunyi; van Ooijen, Peter; Oudkerk, Matthijs

Published in:
Artificial Intelligence in Cardiothoracic Imaging

DOI:
10.1007/978-3-030-92087-6_43

Publication date:
2022

Citation for published version (APA):
Zheng, S., van Ooijen, P., & Oudkerk, M. (2022). Lung Cancer Screening and Nodule Detection: The Role of Artificial Intelligence. In C. N. de Cecco, M. van Assen, & T. Leiner (Eds.), *Artificial Intelligence in Cardiothoracic Imaging* (pp. 459-469). (Contemporary Medical Imaging). Humana Cham. https://doi.org/10.1007/978-3-030-92087-6_43

Lung Cancer Screening and Nodule Detection: The Role of Artificial Intelligence

43

Sunyi Zheng, Peter M. A. van Ooijen, and Matthijs Oudkerk

Introduction

Lung cancer accounted for the majority of cancer-related deaths in 2017, more than the combination of the other three leading cancers: colorectal, breast, and prostate cancer [1]. Despite an accelerated decline in the long-term mortality rate, it remains one of the deadliest cancers worldwide. In 2020, it is estimated to cause 135,720 deaths in the United States [1]. The 5-year survival rate for lung cancer at stage IV is only 4% [2]. However, early detection of lung cancer can improve survival rates, as patient treatment plans can be drawn up and implemented at an earlier disease stage [3]. Computed tomography (CT) is an effective method to detect and investigate lung cancer, which generally presents as a pulmonary nodule. CT can also be utilized for the measurement of pulmonary nodule intensity, diameter, or volume, and can provide diagnostic information, based on morphological features, lymph node locations, and distant metastasis, for use in subsequent nodule management. The implementation of lung cancer screening worldwide has led to an ever-increasing workload pressure on radiologists, and the need for accurate nodule detection. To reduce the heavy burden which radiologists face, over the last two decades, artificial intelligence (AI) systems using classic machine learning and advanced deep learning algorithms have been designed [4, 5]. These systems were designed with the aim of achieving a high sensitivity and a low false-positive rate in nodule identification [6]. Systems implemented using classic machine learning techniques more rely on human intervention to define feature extraction algorithms, while systems with advanced deep learning techniques can independently learn the extraction of features. With more data available, the era of deep learning is leading to a renewed interest in leveraging AI systems for accurate and efficient nodule detection, and the integration of them within reporting workstations.

In this chapter, we will focus on the development of AI detection systems for pulmonary nodules, and discuss factors which affect their performance. Moreover, we look at the comparisons in performance between AI and radiologists on pulmonary nodule detection, and explore the role of AI in lung cancer screening programs.

Lung Cancer Screening

With the aim of detecting lung cancer at an early stage, in populations with a high risk of lung cancer, randomized lung cancer screening trials have been established worldwide. In a recently published review regarding lung cancer screening, multiple large-scale screening programs have provided evidence showing lung cancer-related mortality rate is reduced in patients who have undergone CT screening [7]. Variation was however seen in the definitions of clinically relevant nodules (based on size) in the differing CT-based screening trials. One of the major studies included “relevant” nodules based on diameters >4 mm [8], while other studies adopted diameter in addition to volume measurement for nodule managment [9–11]. In this section we briefly review four large-scale lung cancer screening programs. Comparisons of the four lung cancer screening programs including relevant findings are shown in Table 43.1.

S. Zheng, MSc
Department of Radiation Oncology, University Medical Center Groningen/University of Groningen, Groningen, The Netherlands
e-mail: s.zheng@umcg.nl

P. M. A. van Ooijen, MSc, PhD, CPHIMS
Department of Radiation Oncology and Data Science Center in Health, University Medical Center Groningen, Groningen, The Netherlands
e-mail: p.m.a.van.ooijen@umcg.nl

M. Oudkerk, MD, PhD, FNASCI, EBCR (✉)
Faculty of Medical Sciences, University of Groningen, Groningen, The Netherlands

Institute for Diagnostic Accuracy, Groningen, The Netherlands
e-mail: m.oudkerk@rug.nl

C. N. De Cecco et al. (eds.), *Artificial Intelligence in Cardiothoracic Imaging*, Contemporary Medical Imaging,
https://doi.org/10.1007/978-3-030-92087-6_43

Table 43.1 Comparisons of relevant nodules and findings in various lung cancer screening studies

Screening trials	Size of relevant nodules	Relevant findings
NLST	≥4 mm	Diameter-based LDCT reduced lung cancer mortality (hazard ratio [HR] = 0.80, $P < 0.01$)
NELSON	≥4 mm or ≥30 mm^3	Volume-based LDCT reduced lung cancer mortality (HR = 0.76 in men, $P = 0.01$; HR = 0.67 in women)
UKLS	≥3 mm or ≥15 mm^3	LDCT can detect lung cancer (85.7% stage I or II) at an early stage. Mortality results to be published
MILD	≥60 mm^3	Prolonged LDCT screening reduced lung cancer mortality (HR = 0.61, $P < 0.02$)

The National Lung Cancer Screening Trial (NLST) was initiated in 2002, and in total 53,454 participants were enrolled [8]. Half of the participants received LDCT scans and half underwent single-review chest radiography. The study demonstrated that screening using low-dose CT (LDCT) could yield a 20% reduction in lung cancer mortality compared to screening with chest radiography (hazard ratio [HR] = 0.80, P < 0.01). However, more false positives were found due to variation in nodule characterization and reporting manners by radiologists. To solve this issue, Lung Reporting and Data System (Lung-RADS) was released by the American College of Radiology, in order to provide clear definitions of positive findings and effective nodule management for LDCT lung cancer screening. Despite a decrease in sensitivity, a study showed Lung-RADS (2014 version), when considering noncalcified nodules at least 4 mm in any diameter, can result in a lower false-positive rate (with vs. without: 12.8% vs. 26.6%) at baseline when applying it retrospectively on the results of the NLST [12]. In the recent 2019 version of Lung-RADS, taking diameter and volume measurement into consideration, nodules larger than 4 mm or 30 mm^3 (regardless of type) with a risk of malignancy greater than 1% should be reported and recommended for short-term follow-up with additional diagnostic testing or tissue sampling [13].

The Dutch–Belgian randomized controlled lung-cancer screening trial (NELSON) was started in 2003 and recruited 15,789 individuals aged between 50 and 74. Half of the participants underwent volume-based LDCT screening, and the other participants received no screening (control group). Recently published results showed that among men, volume-based screening can substantially reduce the mortality rate of lung cancer, when compared to participants who did not undergo screening (cumulative rate ratio for lung cancer-specific mortality = 0.76, 95% CI: 0.61–0.94, $P < 0.05$ at 10 years) [9]. To support the successful implementation of the NELSON study and other lung cancer screening programs in Europe, the European position statement on lung cancer screening was released [14]. The statement indicated that solid non-calcified nodules larger than or equal to 4 mm or 30 mm^3 and sub-solid nodules greater than 5 mm are clinically relevant in screening programs.

The UK Lung Cancer Screening (UKLS) pilot trial randomized 4055 patients, aged 50–75, into CT screening and no screening groups. The UKLS trial showed that more than 80% of early stage lung cancers could be detected and delivered potentially curative treatment [10]. This study focused on nodules which were larger than 3 mm or 15 mm^3.

The multicenter Italian Lung Detection (MILD) trial recruited 2376 individuals who underwent annual or biennial LDCT. Ten-year results of the MILD trial provided evidence that prolonged LDCT screening led to a 39% mortality reduction in lung cancer (HR = 0.61, 95% CI: 0.39–0.95, $P < 0.02$) [15]. Besides, the lung cancer-specific 10-year mortality was similar in both biennial and annual screening groups. However, LDCT screening occurring every 2 years instead of 1 year could save 44% of follow-up scans for individuals who had a negative baseline examination [11]. A negative case in the MILD trial was classified as a nodule with the volume less than 60 mm^3.

Despite the increased opportunity for early detection of lung cancer through the implementation of lung cancer screening programs, lung cancer may still be missed or detected at a late stage. This is due to the complexity of nodule characterization and their anatomical structures, an increased number of CT slices to be viewed, inconsistent levels of observer reading experience, fatigue, etc. A potential solution to this problem is the use of a well-preforming AI-based system to assist radiologists for nodule detection.

AI-Based Pulmonary Nodule Detection

Machine intelligence, better known as artificial intelligence (AI), is used to describe machines which are designed to mimic and help humans to achieve certain goals. AI has been attempting to support radiologists since the end of the twentieth century. The first AI system for lung nodule detection on chest radiographs was published in 1988. Due to insufficient computational power, poor image quality, and immature detection algorithms, the system yielded a high false-positive rate and its generalizability was unknown [16]. With the developments in computational resources and digital imaging, computers are able to process more difficult tasks, and the introduction of CT imaging brings a clearer view of the three-dimensional bodily structures. Consequently, AI has achieved great success in nodule detection in recent years, especially during the era of deep learning.

Aims of AI

The reasons for implementing an AI nodule detection system for radiologists are as follows:

1. To avoid missing clinically relevant nodules. Nodules have large variations in size, characteristics, and anatomical location, which can easily lead to false-negative outcomes.
2. To reduce the number of false positives that might result in unnecessary follow-up, incorrect clinical staging and treatment, etc.
3. To save the time of radiologists concentrating on diagnosis and subsequent management of nodule characteristics, if the AI system is efficient and can accurately process scans.

Stages of Nodule Detection Systems

An AI nodule detection system can consist of five stages. Firstly, images should be retrieved from the picture archiving and communication system. Then, the image will be preprocessed in appropriate window settings. The next (optional) step is to segment lung parenchyma, so that the system can focus solely on the region of interest and not be disrupted by other organs, clothes, or tables. The most essential two stages come after the segmentation of lung parenchyma, as shown in Fig. 43.1. At the nodule candidate detection stage, AI will try to find as many nodule candidates as possible based on features defined by algorithms. Corresponding nodule coordinates and size can be further calculated to extract features of nodule candidates. In the end, by analyzing extracted features, false-positive candidates will be removed from the final output at the false-positive reduction stage.

Detection Systems Using Classic Machine Learning Algorithms

Machine learning (ML), as a subset of AI, can achieve a good performance on small datasets, but reply on the experience of humans. This might limit the generalizability of the machine learning-based system if it is validated on an external set, or even an internal set. Moreover, due to inadequate computational power, most systems have been established based on small private databases, and their performance varies. In a literature survey regarding the progress of machine learning-based systems on nodule detection, the sensitivity of the systems varied from 70.0% to 90.0%, whereas the false-positive rate ranged from 0.5 to 15 per scan [17].

In one of the fully automated detection systems, multiple thresholds were applied to segment the lung region, and initial nodule candidates were selected using an 18-point algorithm [18]. For nodule candidate characterization, a feature vector was computed based on the mean HU value, standard deviation, volume, radius, sphericity, eccentricity, and circularity. In 43 evaluated CT scans with 171 nodules, the AI system had a sensitivity of 70.0% with 1.5 false positives detected in every scan. A different study dedicated to detecting isolated, juxtapleural, and juxtavascular nodule types using 2D and 3D operations accordingly [19] used 164 nodules detected in 20 patients by 2 radiologists as a reference standard. A sensitivity of 97.4% was achieved for isolated nodules, whereas a sensitivity of 93.0% was achieved for juxtapleural and juxtavascular nodules. Using a novel quantized convergence index filter to enhance round lesions, an automatic system was able to achieve a high sensitivity of 90% on five clinical datasets [20]. Eight rule-based features, such as the output level of quantized convergence index filter, gray-level gradient, and lengths, were applied for false-positive reduction by linear discriminant analysis. Another

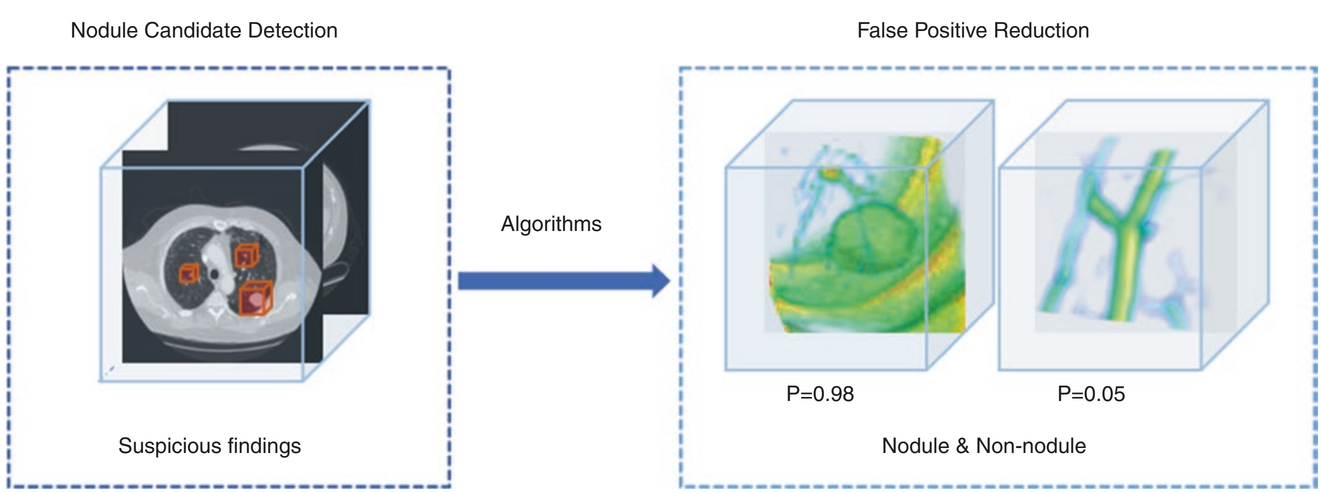

Fig. 43.1 Two main stages in a nodule detection system

interesting study designed an algorithm that included a multi-scale Laplacian of Gaussian filter, prior shape information, and circularity to detect nodule candidates, before applying deep learning techniques for false-positive reduction [21]. One hundred percent of nodules on the 888 CT scans were found, although an average false-positive rate of 50.3 appeared in each scan.

Since many AI systems used different data for development, it is difficult to make a fair comparison. Hence, the public dataset ANODE09 was released, to benchmark the performance of different systems [22]. This dataset included 55 scans collected in the NELSON study. The organizer compared the performance of six AI systems and found the combination of the results can boost the final sensitivity from 20.8% to 72.5% with one false positive per scan [23]. More specially, with the same false-positive rate, the sensitivities of identifying small, large, isolated, vascular, pleural, and peri-fissural nodules were 76.1%, 67.8%, 73.8%, 72.1%, 62.7%, and 82.9%, respectively. More nodules could have been detected if a higher false-positive rate was tolerated. They also studied false-positive findings and found the main causes to be vessel branchings, small lesions such as scar tissue, and apparent protrusions mimicking nodules in areas where high-density body structures are close or in contact with the pleural surface of the lung. As false positives are often small, an extra algorithm might be helpful which discards findings less than 4 mm, in order to further reduce the false-positive rate of the system.

Detection Systems Applying Advanced Deep Learning Algorithms

Deep learning (DL) is a subcategory of machine learning utilizing artificial neural networks to learn presentative features of targets. Compared to ML algorithms, DL is more advanced with automated feature extraction and possibly unlimited accuracy, although a larger training set and more powerful computers are required. A recent systematic review paper showed DL systems can achieve high accuracy from 82.2% to 97.6% [4].

For the purpose of facilitating the development of computer-aided detection systems using DL techniques, large datasets are available, such as the Lung Image Database Consortium and Image Database Resource Initiative (LIDC/IDRI) with 1018 annotated thoracic CT scans collected from 7 participated academic centers, [24] and the National Lung Cancer Screening Trial (NLST) dataset with LDCT scans from 26,254 patients [25]. Furthermore, many competitions have been held for the implementation of DL-based nodule detection systems. For example, the LUng Nodule Analysis 2016 (LUNA16) competition built a benchmark dataset with 888 CT scans which were acquired in thin slices and selected from the LIDC-IDRI database for evaluation [26]. In 2017, a Kaggle's Data Science Bowl (DSB) challenge was organized for lung cancer diagnosis with a subset of the NLST set [27]. Two years later, the Lung Nodule Database (LNDb) competition released 294 annotated CT scans retrospectively collected in Porto for automated nodule management based on the 2017 Fleischner society pulmonary nodule guidelines [28, 29].

In recent years, a large number of researchers have utilized the LUNA16 dataset to build up AI detection systems. In one study, results of five complete AI detection systems and three systems developed only for removal of false positives were combined. The nodule detection rate at the candidate detection stage was increased up to 98.3%, which was higher than that of any single system. At the false-positive reduction stage, probabilities of nodule candidates were simply averaged. As a result, the merged system achieved a sensitivity of 96.9% with one false positive per scan [30]. In another study, a deep learning approach was implemented inspired by the maximum intensity projection (MIP) technique which is one of the useful routine clinical procedures [31]. Their results showed that convolutional neural networks were capable of detecting more small nodules (85.5–88.5%) between 3 and 10 mm on three different MIP slices (5, 10, and 15 mm) than on 1 mm axial slices (79.4%). Examples of various MIP images are shown in Fig. 43.2. After results on various slices were merged, 95.4% of nod-

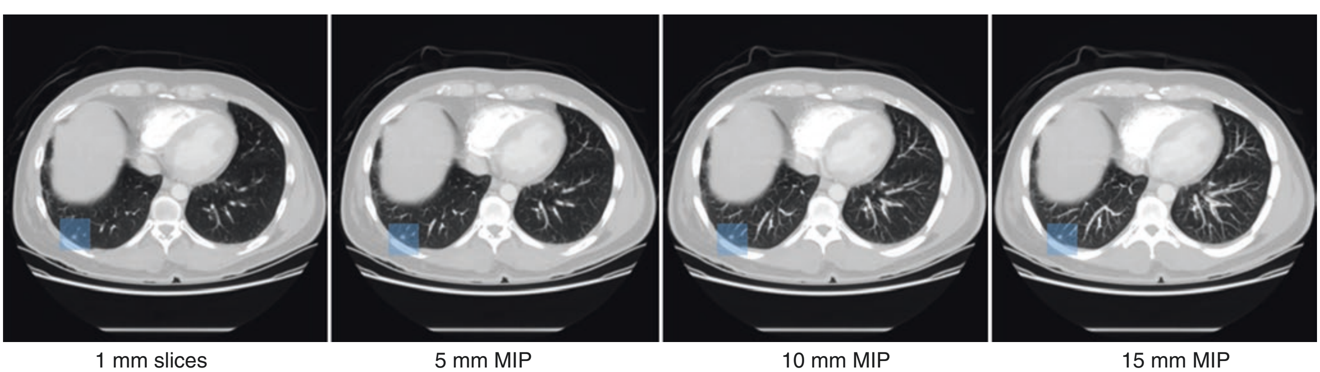

Fig. 43.2 Comparisons between different images generated by maximum intensity projection (MIP) and 1 mm axial slices. Nodules are enhanced and vessels are more continuous in the MIP images

ules (regardless of size) were detected. Moreover, using two kinds of cubic patches for the elimination of false positives resulted in a final sensitivity of 92.7% with one false positive per scan. When the system took 16 MIP slices into consideration, the performance was further improved to 94.4% at a false-positive rate of one [32]. Furthermore, a study proposed a 3D Feature Pyramid Network with 96 × 96 × 96 volumetric patches and an HS2 network to reduce false positives by tracking changes of consecutive slices where the nodule candidate is [33]. Aided by applying spatial features, the system can reach a sensitivity of 90.4% with a very low false-positive rate of one eighth. Additionally, this study implemented a faster region-based convolutional neural network with two regions [34]. To integrate spatial information with the aim of refining results, three continuous slices containing nodule candidates were fed into three classification models, and the final probability was averaged. If the models misclassified any nodules, this data was used again for retraining. The designed system finally found 85.2% of nodules with also one false positive per scan.

In addition, nodules have large variations in morphology, which challenges the ability of the system to reduce the number of irrelevant findings. One method took CT patches in multiple scales for training, and combined the strategy that gradually extracted features in a zoom-in and zoom-out manner [35]. As a consequence, the proposed method yielded a sensitivity of 94.7%. Moreover, based on the ResNet-101 which had high accuracy on natural images, a developed nodule detector found 71.9% of nodules on the subset of the DSB challenge [36].

Some studies have used other databases for system implementation and yielded good performance in nodule identification. For example, a retrospective study collected 9225 chest radiographs with nodules and 34,067 negatives cases, between 2010 and 2015 [37]. Thirteen experienced radiologists labeled these images. The proposed DL system was internally and externally assessed on datasets from three hospitals in South Korea and one in the United States. The area under the receiver of AI (0.92–0.99) was superior to 17 out of 18 physicians who were involved in the evaluation. AI was able to detect 96% of nodules which were >30 mm, 74% of nodules 20–30 mm, 68% of nodules 15–20 mm, 55% of nodules 10–15 mm, and 11% of nodules <10 mm. A different study included 2000 CT scans with 2608 nodules from NLST, LIDC-IDRI, and a private dataset [38]. Challenging negative mining and volumetric patches training were applied to facilitate the performance. When the system was evaluated on 354 validation CT scans, a sensitivity of 93% was achieved for nodules larger than 5 mm. A third study used 12,754 CT scans, of which 11,625 were obtained from 3 hospitals for AI development and 1129 were collected from more than 10 centers for validation [39]. The AI system showed a sensitivity of 74.0% when the false-positive rate is one in terms of different nodules. More specifically, AI had a sensitivity of 71.9%, 88.6%, 61.3%, 85.2%, 86.4%, and 75.3% for solid nodules ≤6 mm, solid nodules >6 mm, sub-solid nodules ≤5 mm, sub-solid nodules >5 mm, calcified nodules, and pleural nodules, respectively. In 2019, the Google AI team proposed an end-to-end system that took the volume of interest for cancer detection [40]. The system was first trained on the LIDC-IDRI dataset and subsequently fine-tuned on 70% of 29,541 NLST cases. Tested on 6729 cases, the cancer detector reached a sensitivity of 97.5% for baseline scans and 94.6% for prior scans when the top 2 detections were hit.

Factors Impacting the Performance of AI

The performance of existing AI systems varies from system to system. Many factors such as CT parameters (slice thickness, slab thickness, scan dose), nodule characteristics (nodule location, nodule size, nodule density), and pulmonary vessel can influence the performance of AI.

Slice Thickness

Due to the difference between vendors and reconstruction algorithms, CT scans have thin and thick slices. Thick-section CT scans have the issue of partial volume effect, which causes the information loss of small objects. Evidence demonstrates sensitivity of the AI detection system is decreased when the same AI was assessed in CT scans with thick slices compared to thin slices. To explore the effect of slice thickness on the performance of a computer-aided detection (CAD) system, Kim et al. evaluated an AI system on multi-detector CT scans in 1 mm and 5 mm groups in 10 patients [41]. In the thin-section group, 126 nodules were detected by 2 radiologists, whereas 114 nodules were found in the thick-section group. The AI system detected 120 out of 126 (95.2%) and 101 out of 114 (88.6%) nodules in two groups separately. The evidence was also supported by Narayanan et al. who assessed a CAD system in 192 CT scans with a slice thickness of 1.25 mm from the LUNA16 dataset [42]. At a down-sampling ratio of 2, 4, and 8, images with a slice thickness of 2.5, 5, and 10 mm were stimulated. The study exemplified the system tested on 1.25 or 2.5 mm slices outperformed the one on 5 or 10 mm slices. The performance of using different slice thickness settings is shown in Fig. 43.3.

Slab Thickness in Maximum Intensity Projections

Maximum intensity projections (MIPs) are one of the clinical procedures used by radiologists to efficiently detect

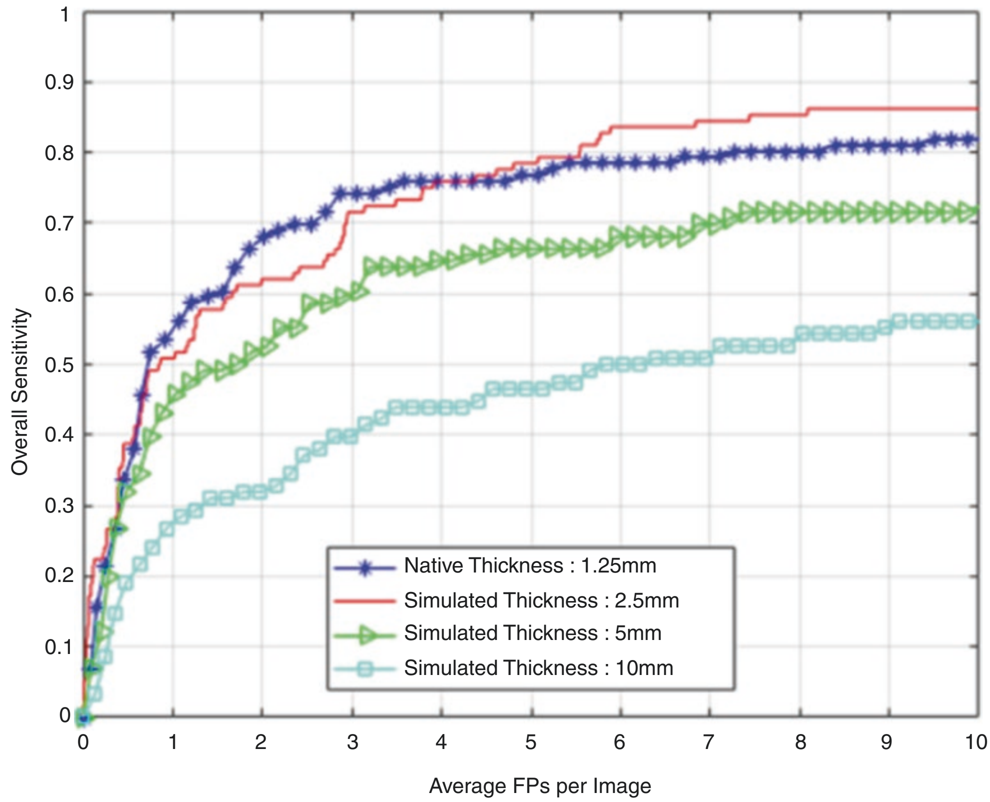

Fig. 43.3 Performance of CAD in different slice thickness settings for nodule detection. (Reprinted with permission of Barath Narayanan. From: Narayananet al. [42])

nodules in vessel-enhanced slices. Human observation using MIPs with a slab thickness of 10 mm showed a higher nodule detection rate than using 5, 15, and 20 mm MIPs [43]. However, the principle of detecting nodules between radiologists and AI is not completely the same. Ten-mm might not be the optimal slab thickness for AI to detect lung lesions. To determine the optimal slab thickness for AI, one study investigated the effect of different slab thicknesses at the nodule candidate detection stage using the large LUNA16 dataset [32]. To avoid partial volume effect, scans with a slice thickness >2.5 mm were excluded. All scans were interpolated to 1 mm slices, and MIPs with a slab thickness from 3 to 50 mm at an interval of 5 mm were generated by the 1 mm slices. The system was developed and validated with one type of MIP each time, and results were acquired by tenfold cross-validation. With increasing slab thickness, the detection rate of the AI system initially gradually increased up to 90% with a slab thickness of 10 mm, and thereafter decreased slowly. Regarding false positives, the number decreased with slab thickness from 1 mm to 25 mm, and remained stable thereafter. When determining the appropriate method for detecting as many nodules as possible at the nodule candidate detection stage, 10 mm was the optimal parameter of MIP for AI.

Radiation Dose

Radiation dose is related to the tube current and voltage, and by lowering one of these values, radiation dose can be reduced. However, in doing so the image quality can be also decreased which may cause more false positives. To attempt to reduce the dose which patients receive from screening, while keeping a high detection rate of lung cancer, studies have been conducted to evaluate the effect of using low-dose CT (LDCT) or ultra-low-dose CT (ULDCT) on the performance of AI detection. It is evident that in spite of an increasing number of false positives, AI can detect a similar number of nodules on CT scans performed at a lower dose. For example, Young et al. collected 481 LDCT scans from the NLST screening and simulated corresponding scans with 50% and 25% of the original dose [44]. On a patient level, the AI system detected 38%, 37%, and 38% of patients who had at least one nodule at the original, 50%, and 25% dose, respectively. Although the sensitivity was comparable, the AI system found 13.5 more false positives per scan with 25% dose, compared to 5 false positives per scan with 50% dose and 3 false positives per scan with routine dose. Therefore, the authors suggested that it might be possible to reduce radiation dose of LDCT by a half in the current NSTL protocol. A different study included scans of 25 patients, and image acquisitions were performed at a dose of 2.2, 1.2, 0.9, and 0.6 mSv by reducing the tube current [45]. When CT scans were reconstructed using the filtered back-projection technique, a similar number of true positives were detected by the AI system at different dose levels, while ten more false positives appeared on the scans at a lower dose compared to routine scans. Furthermore, in a large evaluation study, an AI system was tested on 187 LDCT scans and 942 regular CT scans [39]. There was no significant difference between dose levels for AI when detecting true positives.

Nodule Location

According to anatomical location, nodules can be categorized into pleural, isolated, and juxtavascular types. Some studies have shown that nodule location can affect the performance of AI. In the research of Bae et al., authors found that isolated nodules were easier to detect using the system than juxtavascular and pleural nodules [19]. This was due to isolated nodules having a clear contrast between nodule boundary and lung parenchyma. In addition, these nodules were differentiable on a single slice without the need for comparison between contiguous slices. By contrast, the other two types of nodules, having more complex morphological features, were connected to disturbing vessels or tissues. This trend was also proved by the study of van Ginneken et al. who evaluated an AI system on the ANODE09 dataset [23]. The sensitivities of detection of isolated, juxtavascular, and pleural nodules were 73.8%, 72.1%, and 62.7%, respectively. Another reason for the good accuracy demonstrated when detecting isolated nodules is the high prevalence of these nodules in the training set. This meant that the system had learned more features of isolated nodules, and hence tended to detect them.

Nodule Size

Some AI systems are prone to missing small nodules, since they only show up in a few slices and noise can highly impact the definition of the nodule boundary. In the work of Zheng et al., nodules were stratified according to the Lung-RADS guideline [46]. The majority of missed nodules (15/23) had a diameter between 3 and 6 mm at the nodule candidate detection stage. A similar result was found in the work of Xie et al. [34]. The detection rate was only 44.4% for nodules smaller than 10 mm, whereas it was 86.9% for larger nodules (10–30 mm). Contrarily, van Ginneken et al. found that five out of six of the evaluated AI systems presented lower sensitivities in identifying large nodules at different false-positive rates [23]. A possible explanation from the authors was that small nodules appeared more frequently in the training dataset, which results in a conflict expectation.

Nodule Density

In general, solid nodules are more differentiable than part-solid and non-solid nodules in the lung window settings. A study undertaken by Setio et al. analyzed the density of detected nodules for seven AI nodule detection systems which participated in the LUNA16 challenge [30]. All of the systems showed a similar detection trend; the fewer the solid components in nodules, the greater the sensitivity of the system. In other words, among the three types of nodules, non-solid nodules had the lowest sensitivity. They normally represent areas of ground-glass opacity, which increases the difficulty for AI to differentiate them from almost identical lung parenchyma or focal pneumonia. Another study also investigated the performance of AI on solid and sub-solid nodules [39]. The detection rate of solid nodules was 72.9% (556/763) in LDCT scans, whereas that of sub-solid nodules was only 65.0% (256/394).

Pulmonary Vessels

Pulmonary vessels are one of the main false-positive findings in the results of AI systems. When nodules are close or attached to vessels, it becomes challenging for AI to select these nodules. If vessels are removed from the lung while preserving nodules completely, nodule detection could be much easier for both human readers and AI, with less false-positive results. Inspired by this idea, Gu et al. developed a deep learning-based detection system using the pulmonary vessel suppression technique, and explored the effect of removal of vessels on the performance of AI [47]. The ablation experiment demonstrated that sensitivity improved from 96.9% to 98.6%, with a false-positive rate that dropped from 7.65 to 0.92 per scan after ruling out pulmonary vessels for the same candidate detection and false-positive reduction algorithms. Another recent study showed that the elimination of pulmonary vessels can reduce suspicious regions [48]. Although most nodules (94.0%) can be detected by AI, the false-positive rate, 15.1 per scan, was still high. The reason for this may be that the authors only used 2D slices instead of 3D cubic patches for false-positive elimination; therefore, nodular findings could not be effectively removed.

Role of AI in Nodule Detection

Although AI can achieve promising performance in the detection of different sizes and types of pulmonary nodules, the effect of its performance in the reading of CT scans when compared to or used alongside radiologists should be further studied before applying it in real clinical practice. The performance of AI, such as the percentage of truly detected nodules, the number of false positives, and the order of using aided detection, highly impacts the role that AI plays in clinical settings. For the purpose of exploring the feasibility of applying AI systems in clinical practice, a large number of studies have been conducted to compare their performance with that of radiologists. Despite a few researchers finding that AI failed in improving diagnostic accuracy [49], most results have shown that AI can be a good second or concurrent reader, when results of AI are shown to radiologists subsequently or concurrently. The reading principles of using AI or computer-aided detection (CAD) systems are presented in Fig. 43.4.

In the subsequent model, AI is generally considered as a second reader, and radiologists are only allowed to see the results of computerized detection after finishing their first read. After the initial independent radiologist reads, the computerized findings for the same scans are displayed to radiologists for the purpose of checking whether nodules reported are indeed true nodules. Thus, the consensus panel, including nodules detected by both radiologists and AI, will be included in the reference standard. In the first phase of a comparison study, the authors analyzed the effect of utilizing AI in the subsequent model for nodule detection on thin-section CT scans [50]. When AI was used as a second reader, the final sensitivities were 82%, 97%, and 82% for solid, part-solid, and GGO nodules, respectively. However, the sensitivities were only 57%, 81%, and 69% for corresponding nodules while reading without AI. To assess the effectiveness of AI as a second reader, Awai et al. collected 50 chest CT scans with nodules ranging from 3 mm to 29 mm [51]. Five experienced radiologists and five radiology residents were asked to rate nodule presence with and without using a computer-aided detection system. The observation results showed that there was a significant detection difference between reading alone and using the system afterward, for both experienced and less experienced readers. The study concluded that the use of AI can enhance nodule detection for both reader groups.

In the concurrent model, radiologists can take image interpretation of AI into account while reading the CT scans. Studies have shown that when AI is used as a concurrent reader, radiologists are capable of detecting more nodules than reading alone. The study of Matsumoto et al. compared the performance of nodule detection on 50 CT scans with and without AI. With the help of the second AI reader, 66.5% of nodules were found by radiologists [52]. In contrast, only 56.5% of nodules were detected in unaided reading. Importantly, radiologists took only approximately 10 seconds longer reading each scan, and the sensitivity of nodule detection was improved.

The concurrent model obviously saves more time than the subsequent model since radiologists do not have to go through slices again without AI-aided detection. This assumption was proved by Matsumoto et al. who compared the performance of AI on 50 CT scans in both models [52]. The results showed that the mean reading time in the concurrent model (132 ± 70 s) was significantly shorter than that of the subsequent model (210 ± 103 s). A similar trend was reported in the study of Beyer et al. where the reading time was 274 seconds vs. 337 seconds on average for the concurrent model vs. the subsequent model [53]. Matsumoto et al. also found no significant difference in the performance of radiologists in these two models. Conversely, Beyer et al., who also reviewed 50 scans, reported the sensitivity when using AI as a second reader was significantly higher than the sensitivity in the case of applying it concurrently [53]. A possible explanation for the varying results is that the AI system of Matsumoto et al. achieved a higher stand-alone sensitivity of 63% with 3.5 false positives per scan, whereas Beyer et al. had a stand-alone sensitivity of 43% with 1.3 false positives per scan. The results of Matsumoto et al. were more conducive to the effectiveness of using AI.

AI detection systems can also detect nodules previously missed in the consensus of radiologists, and could be utilized as an assisting reader. Findings could be referenced by radiologists in lung cancer screening programs, if a higher sensitivity and a lower false-positive rate can be achieved. In the study of Zhao et al., an AI system was evaluated on 400 LDCT scans selected from the NELSON screening study [54]. Five out of one hundred and fifty-one nodules were only detected by double reading, whereas

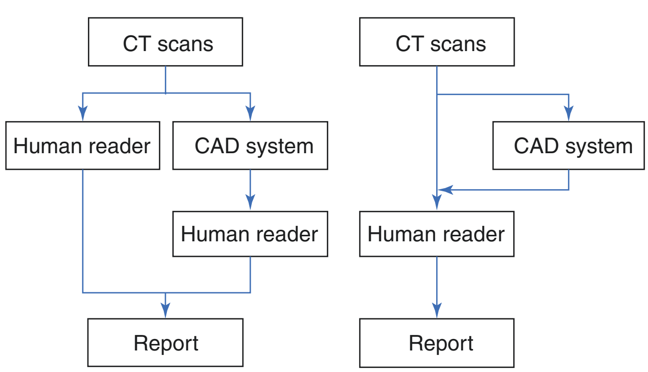

Fig. 43.4 Reading principles with the CAD integrated. The CAD is a second reader in the left workflow, while the CAD is a concurrent reader at the right side

thirty-three true nodules were only identified by AI. Notably, one nodule overlooked by radiologists but found by AI was lung cancer. A second study compared the detection performance of AI and double reading on 346 CT scans retrospectively collected in a lung cancer screening program between March and November 2017 [55]. The ratio of men to women was 221:125, and the average age was 51 years old. The AI system had a higher sensitivity than radiologists, not only for nodules larger than 5 mm (96.5% vs. 88.0%) but also for small nodules (84.3% vs. 77.5%), although the system yielded more false positives for radiologists to check. Seventeen nodules were confirmed as missed by the readers but detected by AI after a third reading by an experienced radiologist. A third comparison study evaluated three AI systems on the large LIDC-IDRI dataset, with 777 nodules accepted by 4 radiologists [56]. The best system reached a sensitivity of 82% with 1108 false-positive findings. After the exclusion of the obvious non-nodules by a researcher, 269 findings were assessed by 4 radiologists. There were 45 marks accepted by 4 radiologists as nodules, of which 70% were not obvious pulmonary nodules. Thirty-seven nodules were larger than 4 mm and required follow-up. These findings indicate AI has the potential to detect subtle nodules previously missed by radiologists. Furthermore, an AI-based system was evaluated on the NLST screening chest radiographs from 5485 participants [57]. The sensitivity and specificity when using AI were 100.0% and 90.9% for malignant nodule detection in digital radiographs. Compared to results of AI, radiologists had a lower sensitivity of 94.1% but a slightly higher specificity of 91.0%. AI also found seven nodules or masses missed by radiologists. The study concluded that the AI system outperformed NLST radiologists in detecting malignant nodules, and might help radiologists to detect lung cancer if used as a second reader.

Despite studies showing that AI can boost radiologist's performance, the feasibility of applying AI systems in daily clinical practice and lung cancer screening has not been sufficiently proven. It is necessary to validate AI systems on screening data from multiple institutes and provide more evidence on the detection of those clinically relevant nodules. In addition, not only technicians but also radiologists should work cooperatively to build up a larger and more widely distributed nodule dataset, which is shared with the community in order to further stimulate the development of nodule detection systems.

Conclusions

Lung cancer accounts for the majority of cancer-related deaths, and the 5-year survival rate is only 4% when detected at stage IV. For the purpose of detecting lung cancer at an early stage, randomized lung cancer screening trials have been established all over the world. The results from some large-scale screening studies, such as NELSON and NLST, show that the mortality rate of patients who undergo CT screening was reduced. These developments give lung cancer patients a greater chance of survival, but as a consequence the reading workload of radiologists is increasing. Artificial intelligence detection systems have been designed with the aim of reducing the pressure on radiologists, so that they can focus more on diagnoses. Despite many factors, such as CT parameters, nodule characteristics, and pulmonary vessels, affecting the sensitivity and false-positive rates of AI, AI can still provide substantial assistance for radiologists in the clinic, both as a second reader and concurrent reader. A few studies have been conducted evaluating AI on lung cancer screening data, and show AI can achieve good performance on nodule detection. In the future, the effect of using AI in lung cancer screening programs should be further studied. AI could be utilized as an assisting reader whose findings can be referenced by radiologists in screening settings, so long as the sensitivity can be continuously improved and the false-positive rate can be minimized.

References

1. Siegel RL, Miller KD, Jemal AJCacjfc. Cancer statistics. 2020. 2020;70(1):7–30.
2. Miller KD, Nogueira L, Mariotto AB, et al. Cancer treatment and survivorship statistics. 2019. 2019;69(5):363–85.
3. Oncology NLSTRTJJoT. Lung cancer incidence and mortality with extended follow-up in the national lung screening trial. 2019;14(10):1732–42.
4. Pehrson LM, Nielsen MB, Ammitzbøl Lauridsen CJD. Automatic pulmonary nodule detection applying deep learning or machine learning algorithms to the LIDC-IDRI database: a systematic review. 2019;9(1):29.
5. Li D, Mikela Vilmun B, Frederik Carlsen J, et al. The Performance of Deep Learning Algorithms on Automatic Pulmonary Nodule Detection and Classification Tested on Different Datasets That Are Not Derived from LIDC-IDRI: A Systematic Review. 2019;9(4):207.
6. Setio AAA, Ciompi F, Litjens G, et al. Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks. 2016;35(5):1160–1169.
7. Oudkerk M, Liu S, Heuvelmans MA, Walter JE, Field JKJNRCO. Lung cancer LDCT screening and mortality reduction—evidence, pitfalls and future perspectives. 2020:1–17.
8. Medicine NLSTRTJNEJo. Reduced lung-cancer mortality with low-dose computed tomographic screening. 2011;365(5):395–409.
9. de Koning HJ, van der Aalst CM, de Jong PA, et al. Reduced lung-cancer mortality with volume CT screening in a randomized trial. 2020;382(6):503–13.
10. Field JK, Duffy S, Baldwin DR, 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. 2016;71(2):161–70.
11. Pastorino U, Sverzellati N, Sestini S, et al. Ten-year results of the Multicentric Italian Lung Detection trial demonstrate the safety and efficacy of biennial lung cancer screening. 2019;118:142–48.

12. Pinsky PF, Gierada DS, Black W, et al. Performance of Lung-RADS in the National Lung Screening Trial: a retrospective assessment. 2015;162(7):485–91.
13. Lung ct screening reporting & data system (lung-rads). https://www.acr.org/Clinical-Resources/Reporting-and-Data-Systems/Lung-Rads. Accessed 10 Oct 2020.
14. Oudkerk M, Devaraj A, Vliegenthart R, et al. European position statement on lung cancer screening. 2017;18(12):e754-66.
15. Pastorino U, Silva M, Sestini S, et al. Prolonged lung cancer screening reduced 10-year mortality in the MILD trial: new confirmation of lung cancer screening efficacy. 2019;30(7):1162–9.
16. Giger ML, Doi K, MacMahon HJMP. Image feature analysis and computer-aided diagnosis in digital radiography. 3. Automated detection of nodules in peripheral lung fields. 1988;15(2):158–66.
17. Li QJCMI, Graphics. Recent progress in computer-aided diagnosis of lung nodules on thin-section CT. 2007;31(4–5):248–57.
18. Armato III SG, Giger ML, MacMahon HJMp. Automated detection of lung nodules in CT scans: preliminary results. 2001;28(8):1552–61.
19. Bae KT, Kim J-S, Na Y-H, Kim KG, Kim J-HJR. Pulmonary nodules: automated detection on CT images with morphologic matching algorithm—preliminary results. 2005;236(1):286–93.
20. Matsumoto S, Kundel HL, Gee JC, Gefter WB, Hatabu HJMIA. Pulmonary nodule detection in CT images with quantized convergence index filter. 2006;10(3):343–352.
21. Zhang J, Xia Y, Zeng H, Zhang YJN. NODULe: Combining constrained multi-scale LoG filters with densely dilated 3D deep convolutional neural network for pulmonary nodule detection. 2018;317:159–67.
22. The ANODE09 dataset. https://anode09.grand-challenge.org/Home/. Accessed 12 Oct, 2020.
23. Van Ginneken B, Armato III SG, de Hoop B, et al. Comparing and combining algorithms for computer-aided detection of pulmonary nodules in computed tomography scans: the ANODE09 study. 2010;14(6):707–22.
24. Armato III SG, McLennan G, Bidaut L, et al. The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans. 2011;38(2):915–31.
25. The National Lung Screening Trial dataset. https://wiki.cancerimagingarchive.net/display/NLST/National+Lung+Screening+Trial. Accessed 11 Oct 2020.
26. LUNA16 grand challenge. https://luna16.grand-challenge.org/. Accessed 11 Oct 2020.
27. Data Science Bowl 2017. https://www.kaggle.com/c/data-science-bowl-2017. Accessed 11 Oct 2020.
28. Pedrosa J, Aresta G, Ferreira C, et al. Lndb: A lung nodule database on computed tomography. 2019.
29. Fleischner 2017 guideline. https://radiologyassistant.nl/chest/plumonary-nodules/fleischner-2017-guideline. Accessed 11 Oct 2020.
30. Setio AAA, Traverso A, De Bel T, et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge. 2017;42:1–13.
31. Zheng S, Guo J, Cui X, Veldhuis RN, Oudkerk M, Van Ooijen PM. Automatic pulmonary nodule detection in CT scans using convolutional neural networks based on maximum intensity projection. IEEE Trans Med Imaging. 2019;39(3):797–805.
32. Zheng S, Cui X, Vonder M, et al. Deep learning-based pulmonary nodule detection: effect of slab thickness in maximum intensity projections at the nodule candidate detection stage. Computer Methods Programs in Biomedicine. 2020;196
33. Liu J, Cao L, Akin O, Tian Y. 3DFPN-HS2: 3D feature pyramid network based high sensitivity and specificity pulmonary nodule detection. Paper presented at: International Conference on Medical Image Computing and Computer-Assisted Intervention. 2019.
34. Xie H, Yang D, Sun N, Chen Z, Zhang Y. Automated pulmonary nodule detection in CT images using deep convolutional neural networks [published online ahead of print January 2019]. Pattern Recogn. 2019;85:109–19.
35. Kim B-C, Yoon JS, Choi J-S, Suk H-IJNN. Multi-scale gradual integration CNN for false positive reduction in pulmonary nodule detection. 2019;115:1–10.
36. Kuan K, Ravaut M, Manek G, et al. Deep learning for lung cancer detection: tackling the kaggle data science bowl 2017 challenge. 2017.
37. Nam JG, Park S, Hwang EJ, et al. Development and validation of deep learning–based automatic detection algorithm for malignant pulmonary nodules on chest radiographs. 2019;290(1):218–28.
38. Kaluva KC, Vaidhya K, Chunduru A, Tarai S, Nadimpalli SPP, Vaidya S. An automated workflow for lung nodule follow-up recommendation using deep learning. Paper presented at: International Conference on Image Analysis and Recognition. 2020.
39. Liu K, Li Q, Ma J, et al. Evaluating a fully automated pulmonary nodule detection approach and its impact on radiologist performance. 2019;1(3):e180084.
40. Ardila D, Kiraly AP, Bharadwaj S, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. 2019;25(6):954–61.
41. Kim J-S, Kim J-H, Cho G, Bae KTJR. Automated detection of pulmonary nodules on CT images: effect of section thickness and reconstruction interval—initial results. 2005;236(1):295–9.
42. Narayanan BN, Hardie RC, Kebede TMJJMI. Performance analysis of a computer-aided detection system for lung nodules in CT at different slice thicknesses. 2018;5(1):014504.
43. Li W, Chu Z, Zhang Y, Li Q, Zheng Y, Lv F. Effect of slab thickness on the detection of pulmonary nodules by use of CT maximum and minimum intensity projection. Am J Roentgenol. 2019;213(3):562–7.
44. Young S, Lo P, Kim G, et al. The effect of radiation dose reduction on computer-aided detection (CAD) performance in a low-dose lung cancer screening population. 2017;44(4):1337–46.
45. Den Harder AM, Willemink MJ, Van Hamersvelt RW, et al. Effect of radiation dose reduction and iterative reconstruction on computer-aided detection of pulmonary nodules: Intra-individual comparison. 2016;85(2):346–51.
46. Zheng S, Cornelissen LJ, Cui X, et al. Deep convolutional neural networks for multi-planar lung nodule detection: improvement in small nodule identification. Med Phys. 2020.
47. Gu X, Xie W, Fang Q, Zhao J, Li QJMp. The effect of pulmonary vessel suppression on computerized detection of nodules in chest CT scans. Med Phys. 2020;47(10):4917–27.
48. Jiang H, Ma H, Qian W, Gao M, Li YJIjob, informatics h. An automatic detection system of lung nodule based on multigroup patch-based deep learning network. 2017;22(4):1227–37.
49. Nietert PJ, Ravenel JG, Taylor KK, Silvestri GAJJoti. Influence of nodule detection software on radiologists' confidence in identifying pulmonary nodules with computed tomography. 2011;26(1):48.
50. Godoy MC, Kim TJ, White CS, et al. Benefit of computer-aided detection analysis for the detection of subsolid and solid lung nodules on thin-and thick-section CT. 2013;200(1):74–83.
51. Awai K, Murao K, Ozawa A, et al. Pulmonary nodules at chest CT: effect of computer-aided diagnosis on radiologists' detection performance. 2004;230(2):347–52.
52. Matsumoto S, Ohno Y, Aoki T, et al. Computer-aided detection of lung nodules on multidetector CT in concurrent-reader and second-reader modes: a comparative study. 2013;82(8):1332–7.
53. Beyer F, Zierott L, Fallenberg E, et al. Comparison of sensitivity and reading time for the use of computer-aided detection (CAD)

of pulmonary nodules at MDCT as concurrent or second reader. 2007;17(11):2941–7.

54. Zhao Y, de Bock GH, Vliegenthart R, et al. Performance of computer-aided detection of pulmonary nodules in low-dose CT: comparison with double reading by nodule volume. 2012;22(10):2076–84.
55. Li L, Liu Z, Huang H, Lin M, Luo DJTc. Evaluating the performance of a deep learning-based computer-aided diagnosis (DL-CAD) system for detecting and characterizing lung nodules: Comparison with the performance of double reading by radiologists. 2019;10(2):183–92.
56. Jacobs C, van Rikxoort EM, Murphy K, Prokop M, Schaefer-Prokop CM, van Ginneken BJEr. Computer-aided detection of pulmonary nodules: a comparative study using the public LIDC/IDRI database. 2016;26(7):2139–47.
57. Yoo H, Kim KH, Singh R, Digumarthy SR, Kalra MKJJNO. Validation of a Deep Learning Algorithm for the Detection of Malignant Pulmonary Nodules in Chest Radiographs. 2020;3(9):e2017135.