Assessing the predictive accuracy of lung cancer, metastases, and benign lesions using an artificial intelligence-driven computer aided diagnosis system

Kunwei Li[1,2#], Kunfeng Liu[1#], Yinghua Zhong[1], Mingzhu Liang[1], Peixin Qin[1], Haijun Li[3,4], Rongguo Zhang[5], Shaolin Li[1], Xueguo Liu[1,2]

[1]Department of Radiology, Fifth Affiliated Hospital of Sun Yat-sen University, Zhuhai, China; [2]Guangdong Provincial Key Laboratory of Biomedical Imaging, Fifth Affiliated Hospital of Sun Yat-sen University, Zhuhai, China; [3]Department of Radiology, The First Affiliated Hospital of Nanchang University, Nanchang, China; [4]Jiangxi Province Medical Imaging Research Institute, Nanchang, China; [5]Beijing Infervision Technology Co. Ltd., Beijing, China

[#]These authors contributed equally to this work.

Correspondence to: Prof. Xueguo Liu, MD, PhD. Department of Radiology, Fifth Affiliated Hospital of Sun Yat-sen University, Zhuhai 519000, China; Guangdong Provincial Key Laboratory of Biomedical Imaging, Fifth Affiliated Hospital of Sun Yat-sen University, Zhuhai 519000, China. Email: liuxueg@mail.sysu.edu.cn.

Background: Artificial intelligence (AI) products have been widely used for the clinical detection of primary lung tumors. However, their performance and accuracy in risk prediction for metastases or benign lesions remain underexplored. This study evaluated the accuracy of an AI-driven commercial computer-aided detection (CAD) product (InferRead CT Lung Research, ICLR) in malignancy risk prediction using a real-world database.

Methods: This retrospective study assessed 486 consecutive resected lung lesions, including 320 adenocarcinomas, 40 other malignancies, 55 metastases, and 71 benign lesions, from September 2015 to November 2018. The malignancy risk probability of each lesion was obtained using the ICLR software based on a 3D convolutional neural network (CNN) with DenseNet architecture as a backbone (without clinical data). Two resident doctors independently graded each lesion using patient clinical history. One doctor (R1) has 3 years of chest radiology experience, and the other doctor (R2) has 3 years of general radiology experience. Cochran's Q test was used to assess the performances of the AI compared to the radiologists.

Results: The accuracy of malignancy-risk prediction using the ICLR for adenocarcinomas, other malignancies, metastases, and benign lesions was 93.4% (299/320), 95.0% (38/40), 50.9% (28/55), and 40.8% (29/71), respectively. The accuracy was significantly higher in adenocarcinomas and other malignancies compared to metastases and benign lesions (all P<0.05). The overall accuracy of risk prediction for R1 was 93.6% (455/486) and 87.4% for R2 (425/486), both of which were higher than the 81.1% accuracy obtained with the ICLR (394/486) (R1 *vs.* ICLR: P<0.001; R2 *vs.* ICLR: P=0.001), especially in assessing the risk of metastases (P<0.05). R1 performed better than R2 at risk prediction (P=0.001).

Conclusions: The accuracy of the ICLR for risk prediction is very high for primary lung cancers but poor for metastases and benign lesions.

Keywords: Lung cancer; artificial intelligence (AI); convolutional neural network (CNN); pulmonary nodule; diagnostic

Introduction

In recent years, there has been a marked increase in the use of low-dose computed tomography (CT) screening for lung cancers. This, together with the subsequent early diagnosis and treatment of patients, has dramatically increased radiologists' workload. According to the white paper on medical imaging using artificial intelligence (AI) in China that was issued in 2019, the number of chest CT examinations in China has been increasing at a rate of 30% every year, whereas the number of radiologists has only been increasing at a rate of 4% (1). Therefore, there is a strong demand for the development of accurate computer-aided detection (CAD) tools in this context. At present, CAD products based on AI can detect pulmonary nodules with low false-positive rates (2,3) and provide quantitative information on nodules, such as size, volume, consistency, location, and probability of malignancy, and even compare images from multiple time points to automatically calculate the volume doubling time (1,4-6). AI products are often used as a first-line or second-line radiography reader (7-9). This greatly improves the efficiency of radiologists and the detection rate of pulmonary nodules and reduces the number of missed nodules and misdiagnoses (3,10-13).

AI-driven computer-aided diagnosis (CADx) is a non-invasive, objective solution for assisting radiologists in diagnosing lung nodules. Existing CADx methods fall into two categories: classification models based on hand-crafted features (14) and deep neural networks with automatic feature extraction (15-17). Approaches in the first category typically measure radiological traits, such as nodule size, location, shape, and texture, and adopt a classifier to determine malignancy status. In the second category, models based on deep neural networks can automatically learn features for diagnosis from lung CT images. Both two-dimensional (2D) convolutional neural networks (CNNs) and three-dimensional (3D) CNNs are commonly used deep learning models in AI-driven CADx systems. While 2D CNNs have lower computational complexity, 3D CNNs can better analyze pulmonary nodules' spatial structure (4). Indeed, CADx systems have shown promising prediction accuracy for lung nodules' risk stratification (4,5).

The availability of AI-driven lung nodule detection products has gradually increased and is now widely used in the clinical frontline to reduce radiologists' workload and improve diagnostic efficiency. On July 3rd, 2020, a deep learning-based AI product developed by Beijing Infervision Technology Co., Ltd. (China) became the first

U.S. FDA-approved automated lung nodule detection product. Understanding the advantages and disadvantages of this AI product is crucial to its clinical applicability. The use of AI products in the clinic has primarily focused on automatic pulmonary nodule detection (3,5,11) and classification (2,4,5,10) to identify primary lung tumors (4,5). Furthermore, AI products have mainly been trained using screening cohorts or public databases, such as the Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI) database. However, their performance and accuracy in risk prediction of nodules or mass lesions with different underlying pathology, especially metastases and benign lesions, has not been examined in clinical practice.

This investigation used real-world data to evaluate risk prediction accuracy for a wide range of abnormalities in a clinical cohort using an AI-driven lung nodule algorithm provided by Beijing Infervision Technology Co., Ltd. This AI model was based on a 3D CNN with DenseNet architecture as a backbone. The accuracy of risk prediction by the algorithm was compared to that of two resident doctors in our department. This data will help to improve our understanding of the current AI products available and their use in the clinic.

Methods

Case selection

The ethics committee approved this retrospective study of the Fifth Affiliated Hospital of Sun Yat-sen University. Patient informed consent was waived as the study had minimal risk and would not adversely affect the patient's rights or welfare. Two datasets were compiled using an in-house-developed Radiology Information System/Picture Archiving and Communication System search engine. Patients who underwent lung resection in our institution between September 2015 and November 2018 were identified. The following inclusion criteria were applied: (I) patients with surgically resected and histologically proven lung nodules or masses; and (II) surgery within one month from the last CT scan. The following exclusion criteria were applied: (I) no available presurgical non-contrast thin-section (≤ 2 mm) chest CT scan; (II) diffuse disease; (III) CT images with severe breathing or other motion artifacts; (IV) presurgical chemotherapy or radiotherapy treatment; (V) lesions not detected by the AI algorithm; and (VI) lesions with detection errors when using the AI algorithm
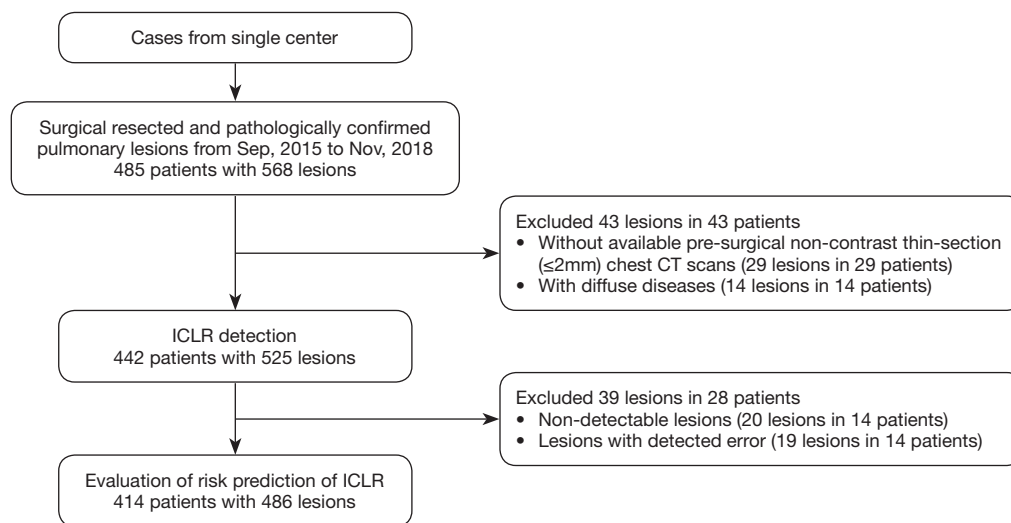
```
┌─────────────────────────────────┐
│      Cases from single center   │
└─────────────────────────────────┘
                │
                ▼
┌─────────────────────────────────┐
│ Surgical resected and pathologically confirmed │
│ pulmonary lesions from Sep, 2015 to Nov, 2018  │
│      485 patients with 568 lesions             │
└─────────────────────────────────┘
                │
                │──────────────────────►  ┌──────────────────────────────────────────┐
                │                          │ Excluded 43 lesions in 43 patients        │
                │                          │ • Without available pre-surgical non-contrast thin-section │
                │                          │   (≤2mm) chest CT scans (29 lesions in 29 patients) │
                │                          │ • With diffuse diseases (14 lesions in 14 patients) │
                │                          └──────────────────────────────────────────┘
                ▼
┌─────────────────────────────────┐
│        ICLR detection           │
│    442 patients with 525 lesions│
└─────────────────────────────────┘
                │
                │──────────────────────►  ┌──────────────────────────────────────────┐
                │                          │ Excluded 39 lesions in 28 patients        │
                │                          │ • Non-detectable lesions (20 lesions in 14 patients) │
                │                          │ • Lesions with detected error (19 lesions in 14 patients) │
                │                          └──────────────────────────────────────────┘
                ▼
┌─────────────────────────────────┐
│ Evaluation of risk prediction of ICLR │
│     414 patients with 486 lesions     │
└─────────────────────────────────┘
```

Figure 1 Study workflow for patient recruitment. CT, computed tomography; ICLR, InferRead CT Lung Research.

(for example, one large lesion was mis-detected as multiple lesions). Moreover, this study utilized the Faster RCNN model to detect lesions. The study workflow is displayed in *Figure 1*.

The management protocols were as follows: (I) the I-ELCAP protocol was used for nodule screening and detection, and the Fleischner society and the National Comprehensive Cancer Network (NCCN) guidelines were applied for incidental nodules; (II) a multidisciplinary team (MDT) discussion on growing subsolid nodules was conducted during follow-up; (III) the patient's wishes were accommodated where possible, especially those who are nervous and anxious. Subsolid nodules <10 mm were resected for the following reasons: (I) the smaller accompanying lesions were removed together with, the larger lesions, and this accounted for most cases; and (II) the patients strongly requested surgery.

Some of the cases included in this current investigation were also in our previous studies, which focused on the quantitative radiomic model for predicting solid nodules' (SNs) malignancy (18-20).

CT image acquisition

Non-contrast CT scans were acquired in a single-breath hold during full inspiration using three CT scanners, Somatom Sensation 16 (S16), Definition Flash (DF) (Siemens Medical Solutions, Forchheim, Germany), and Uct760 (United Imaging; Shanghai, China). The scanning range was from the lung apex to the base. All images

were obtained with a standard dose scanning protocol and reconstructed at 1.0-mm or 2.0-mm slice-thicknesses with 0.7-mm to 1-mm increments, 512×512 matrix, and a moderate or high reconstruction kernel (b50f, b60f, b70f for S16 and DF; sharp for Uct760). The lung window setting used a window level of -600 Hounsfield units (HU) and a window width of 1,500 HU. The mediastinal window setting used a window level of 40 HU and a window width of 400 Hu.

Evaluation by a computer-aided diagnostic system

The AI-driven commercial CAD product (InferRead CT Lung Research, ICLR), which utilizes a detection and risk prediction model, was provided by Beijing Infervision Technology Co., Ltd. (Beijing, China).

To train the detection model, a total of 11,205 CT scans with 3,527,048 image slices were collected from multiple hospitals in China. The scanning devices included GE MEDICAL SYSTEMS, Siemens, Philips, and TOSHIBA. Two radiologists (radiologists A and B) with approximately 10 years' experience independently reviewed the CT scans and the corresponding radiology reports to label the locations and specific attributes of the pulmonary nodules or mass lesions in the image slices. In the case of disagreements, the annotation was checked by a third radiologist (radiologist C) with 15 years' experience, and a consensus was reached by discussion. This study utilized the Faster RCNN model (21) to detect the nodules. The AI model utilized in the current study was a Region-based

CNN for object detection, which consisted of two modules. The first was the regional proposal network (RPN), a fully convolutional network for generating object proposals. The second was the Fast R-CNN detector which aimed to refine the proposals generated by the first module. Compared to the previous two versions (i.e., R-CNN and Fast R-CNN), the Faster R-CNN had better performance and faster processing speed, with the less computational burden (21,22).

For training of the risk prediction model, training data including 5,000 benign and 3,604 malignant nodules or mass lesions were prepared. Among the malignant training samples, adenocarcinomas accounted for 93.8% of all lesions. Squamous cell carcinomas, other primary malignant tumors, and metastases accounted for 4.2%, 1.0%, and 1.0% of all lesions, respectively. Biopsy or surgery was performed in all malignant cases to determine the characteristics of the lesion. Among benign samples, approximately 21% were highly suspected of being malignant and were biopsy- or surgery-proven. The remaining 79% were confirmed through long-term follow-up and judged as benign by experienced radiologists. These lesions had obvious benign features, and most were small nodules less than 20 mm. This study utilized a CNN based on the ResNet-34 framework as the risk prediction model. The fully connected layer's output was fed into the sigmoid layer to acquire the risk probability of lung nodule malignancy. A detailed description relating to the risk prediction model can be found in our previous study (23).

When radiologists used the CAD product, the imaging data were transmitted to an ICLR workstation post-anonymization. After processing by the workstation, the probability of malignancy was recorded. The probability was divided into three levels, namely, low-risk (<50%), moderate-risk (50–70%), and high-risk (>70%). The division criteria were determined based on the training model results derived from the larger data set and internal verification.

Image interpretation by radiologists

Two thoracic radiologists, each with 12 and 17 years of experience, were blinded to the ICLR system's results. Each independently interpreted the CT images using an institutional digital database system (PACS, V5.5.4.50720, Neusoft, Shenyang, China). Any disagreements were resolved by a third radiologist with 30 years of experience. Consistency, size, and distribution were recorded for each lesion. If the lung parenchyma within the entire nodule was obscured, it was classified as a SN, even if there was external or internal cystic airspace or internal cavitation. If the underlying parenchyma was visible except for branching blood vessels within the nodule, it was classified as a nonsolid nodule (NS). If the nodule had nonsolid components and solid components, it was classified as a part-solid nodule (PS) (24). The overall size was determined based on the maximum diameter on 3D images. Lesions were divided into four subgroups ($size_{10mm}$, $size_{20mm}$, $size_{30mm}$, and $size_{>30mm}$) based on size, representing lesions of ≤10, 10–20, 20–30, and >30 mm, respectively. Lesions in the lower lobes were regarded as lower lobe lesions, and lesions in both the upper and right middle lobes were classified as upper lobe lesions.

Observer study

To compare the deep learning system with human performance, two radiology resident doctors [one with 3 years of chest radiology training (R1) and one with 3 years of general radiology training (R2)], were blinded to the results of the ICLR system and pathology. Both independently reviewed and graded each lesion. Lesions were divided into three categories, namely, malignant (high-risk, >70%), suspicious for malignancy (moderate-risk, 50–70%), and benign (low-risk, <50%). The doctors were given access to associated patient demographics, clinical history, and prior CT images.

Histological classification

Four histological types, including adenocarcinoma, other malignancy, metastatic, and benign, were evaluated in our study. The histopathologic classification of adenocarcinomas was based on the IASLC/American Thoracic Society/ERS classification (25) and included preinvasive lesions [atypical adenomatous hyperplasia (AAH) and adenocarcinoma in situ (AIS)], minimally invasive adenocarcinomas (MIA), and invasive adenocarcinomas (IA). Another malignancy refers to the 2015 World Health Organization classification of malignant lung tumors, except for adenocarcinomas and metastatic tumors (26).

Statistical analysis

Categorical variables were summarized as percentages. Continuous variables were summarized as means ± standard deviations. Differences were evaluated with Chi-square

tests for categorical variables. Differences between groups were tested with Bonferroni correction. To calculate malignancy risk prediction accuracy, a true result was defined as a benign lesion that was predicted as low-risk or a malignant lesion that was predicted as moderate- or high-risk. Otherwise, it was defined as a false result. Cochran's Q test was used to compare the malignancy risk prediction performance between ICLR, R1, and R2. A P value <0.05 was considered statistically significant. Statistical analysis was performed using IBM SPSS Statistics for Windows, Version 25.0 (IBM Corp., Armonk, NY, USA).

Results

The detection rate of nodules

A total of 485 patients with 568 lesions met the inclusion criteria. Among them, 29 lesions in 29 patients without available presurgical non-contrast thin-section chest CT, 14 lesions in 14 patients with diffuse disease, 20 non-detectable lesions in 14 patients, and 19 lesions in 14 patients with detection errors in the AI algorithm were excluded. The non-detectable lesions included 6 endobronchial lesions, 3 perihilar lesions, 7 adhesive lesions (lesions were attached to the mediastinum in 3 cases, attached to the pleural effusion in 2 cases, attached to the inflammatory exudate in 2 cases), a large mass attached to the costal pleura in 1 case, 2 cases with patchy appearance, and 1 case with a 3-mm NS nodule. Detailed information regarding the 20 non-detectable lesions and 19 lesions with detection errors are shown in Tables S1 and S2, respectively. Examples are shown in Figures S1 and S2.

The remaining 486 lesions in 414 patients (including 218 males and 196 females) were used for malignancy risk evaluation using the ICLR. The detection rate was 92.6% (486/525), suggesting that the Faster RCNN utilized in this study had good lung nodule detection performance. There were 342, 52, and 10 patients with 1, 2, and 3 resected lesions, respectively. The average age was 58.4±11.0 years (range, 28 to 81 years). The average nodule size was 20.7±13.7 mm (range, 3.0–78.7 mm), and there were 90 NS, 75 PS, and 321 SN lesions. A total of 320 adenocarcinomas, 40 other malignancies, 55 metastases, and 71 benign lesions were detected, with mean sizes of 20.1±12.1 mm (3.0–78.7 mm), 37.7±16.6 mm (11.0–77.6 mm), 15.3±10.5 mm (3.0–69.0 mm), and 17.3±13.7 mm (3.0–73.0 mm), respectively. Characteristics of the patient cohort are shown in Table 1.

Accuracy of malignancy-risk prediction by the ICLR

Out of the 486 lesions, 79 were classified by the ICLR as low-risk, 40 were moderate-risk, and 367 were high-risk. The malignancy risk prediction categories obtained from the ICLR and the corresponding pathological classification (benign and malignant) for each category are shown in Table 2. The overall accuracy of risk prediction was 81.1% (394/486). The accuracy of risk prediction was significantly different based on size, consistency, and pathology, but no significant difference was observed among different scanners, slice thicknesses, and lobe distributions. Factors affecting the accuracy of malignancy-risk prediction by the ICLR are summarized in Table 3.

The accuracy of malignancy-risk prediction in the $size_{10mm}$, $size_{20mm}$, $size_{30mm}$, and $size_{>30mm}$ subgroups was 74.5% (82/110), 77.5% (145/187), 87.8% (86/98), and 89.0% (81/91), respectively. The risk prediction accuracy for lesions larger than 20 mm was slightly higher than that for lesions smaller than 20 mm. However, there was no significant difference in the accuracy of prediction in any of the size subgroups.

The accuracy of malignancy-risk prediction for NS, PS, and SN lesions was 94.4% (85/90), 94.7% (71/75), and 74.1% (238/321), respectively. The risk prediction accuracy for NS and PS lesions was significantly higher than that for SN lesions (NS vs. SN, P<0.05; PS vs. SN, P<0.05).

The accuracy of malignancy-risk prediction for adenocarcinomas, other malignancies, metastases, and benign lesions was 93.4% (299/320), 95.0% (38/40), 50.9% (28/55), and 40.8% (29/71), respectively. The risk prediction accuracy for primary lung cancer was significantly higher than that for metastases and benign lesions (all P<0.05).

Accuracy of risk prediction of the ICLR in lesions with different histological types and sizes

The accuracy of risk prediction for adenocarcinomas in the $size_{10mm}$, $size_{20mm}$, $size_{30mm}$, and $size_{>30mm}$ subgroups was 87% (59/68), 91% (116/127), 99% (72/73), and 100% (52/52), respectively. For other malignancies, the accuracy of risk prediction for the different size subgroups was 0% (0/0), 88% (7/8), 100% (6/6), and 96% (25/26), respectively. For metastases, the accuracy of risk prediction for the different size subgroups was 20% (4/17), 60% (15/27), 70% (5/7), and 100% (4/4), respectively, and for benign lesions, it was 76% (19/25), 28% (7/25), 25% (3/12), and 0% (0/9), respectively.

Table 1 Patient characteristics

Characteristics	Validation sets
Patient demographics	
No. of patients	414
Age, mean ± SD, years	58.4±11.0
Female	196 (44.3)
Smoking history	
Non-smoker	278 (67.2)
Current smoker	82 (19.8)
Former smoker	54 (13.0)
Tumor history	96 (23.2)
Image quality distribution	
CT scanner	
Siemens DF	382 (78.6)
Siemens S16	76 (15.6)
uCT 760	28 (5.8)
Slice thickness	
1 mm	394 (81.1)
2 mm	92 (18.9)
Dose	
Standard	476 (97.9)
Low	10 (2.1)
Pulmonary lesion information	
No. of lesions	486
Location	
Right upper lobe	144 (29.6)
Right middle lobe	46 (9.5)
Right lower lobe	98 (20.2)
Left upper lobe	115 (23.7)
Left lower lobe	83 (17.1)
Size, mean ± SD, mm	20.7±13.7
Size subgroup	
≤10 mm	110 (22.6)
10–20 mm	187 (38.5)
20–30 mm	98 (20.2)
>30 mm	91 (18.7)

Table 1 (*continued*)

Table 1 (*continued*)

Characteristics	Validation sets
Consistency	
Nonsolid	90 (18.5)
Part-solid	75 (15.4)
Solid	321 (66.1)
Histological type	
Adenocarcinoma	320 (65.8)
Atypical adenomatous hyperplasia	7 (1.4)
Adenocarcinoma in situ	1 (0.2)
Minimally invasive adenocarcinoma	15 (3.1)
Invasive adenocarcinoma	297 (61.1)
Other malignancy	40 (8.2)
Squamous	29 (6.0)
Adenosquamous	2 (0.4)
Sarcomatoid carcinoma	3 (0.6)
Lymphoepithelioma-like carcinomas	3 (0.6)
Neuroendocrine carcinoma	1 (0.2)
Epithelioid hemangioendothelioma	1 (0.2)
Follicular dendritic cell tumor	1 (0.2)
Metastasis	55 (11.3)
Benign	71 (14.6)
Congenital cystic adenomatoid malformation	2 (0.2)
Tuberculosis	12 (2.5)
Chronic granuloma	15 (3.1)
Cryptococcus	7 (1.4)
Fungus	2 (0.2)
Hamartoma	7 (1.4)
Sclerosing pneumocytoma	5 (1.0)
Inflammatory pseudotumor	2 (0.4)
Intrapulmonary lymph node	5 (1.0)
Bronchocele with inflammation	1 (0.2)
Other non-specific	13 (2.7)

Unless otherwise stated, each cell indicates the number (percentage). SD, standard deviation; CT, computed tomography.

Table 2 The frequency of malignancy risk prediction categories obtained from the ICLR and the corresponding pathological classification under different conditions

Variable	Low (n=79)	B/M	Moderate (n=40)	B/M	High (n=367)	B/M
Scanner						
Siemens DF	59 (15.4)	22/37	31 (8.1)	4/27	292 (76.4)	56/326
Siemens S16	8 (10.5)	3/5	7 (9.2)	0/7	61 (80.3)	7/54
Uct760	12 (42.9)	4/8	2 (7.1)	1/1	14 (50.0)	0/14
Slice thickness						
1 mm	69 (17.5)	25/44	31 (7.9)	5/26	294 (74.6)	29/265
2 mm	10 (10.9)	4/6	9 (9.8)	0/9	73 (79.3)	8/65
Lobe distribution						
Upper	44 (14.5)	14/30	26 (8.6)	3/23	234 (77.0)	19/215
Lower	35 (19.2)	15/20	14 (7.7)	2/12	133 (73.1)	18/115
Size						
≤10 mm	41 (37.3)	19/22	19 (17.3)	1/18	50 (45.5)	5/45
10–20 mm	31 (16.6)	7/24	17 (9.1)	3/14	139 (74.3)	15/124
20–30 mm	6 (6.1)	3/3	1 (1.0)	0/1	91 (92.9)	9/82
>30 mm	1 (1.1)	0/1	3 (3.3)	1/2	87 (95.6)	8/79
Consistency						
Nonsolid	4 (4.4)	1/3	14 (15.6)	0/14	72 (80.0)	2/72
Part-solid	0 (0)	0/0	2 (2.7)	0/2	73 (97.3)	4/69
Solid	75 (23.4)	28/47	24 (7.5)	5/19	222 (69.2)	31/191
Histological types						
Adenocarcinoma	21 (6.6)	0/21	26 (8.1)	0/26	273 (85.3)	0/273
Other malignancy	2 (5.0)	0/2	2 (5.0)	0/2	36 (90.0)	0/36
Metastasis	27 (49.1)	0/27	7 (12.7)	0/7	21 (38.2)	0/21
Benign	29 (40.8)	29/0	5 (7.0)	5/0	37 (52.1)	37/0

Each cell indicates the number (percentage) of patients in the corresponding malignancy risk prediction categories (low, moderate, high). B/M, benign/malignant according to final pathology results; ICLR, InferRead CT Lung Research.

For adenocarcinomas and metastases, the accuracy of risk prediction increased significantly with an increase in lesion size (P=0.005 and P=0.016, respectively). For adenocarcinomas, the accuracy of risk prediction was significantly higher in the size$_{30mm}$ and size$_{>30mm}$ subgroups than the size$_{10mm}$ subgroup (both P<0.05). For benign lesions, the accuracy of risk prediction decreased significantly with an increase in lesion size (P<0.001). For other malignant lesions, there was no significant difference among the size subgroups (P=0.513) (*Figure 2*).

Accuracy of risk prediction of the ICLR in lesions with different histological types and consistencies

The respective accuracy of risk prediction in NS, PS, and SN lesions was 97% (84/87), 100% (71/71), and 89% (144/162) for adenocarcinomas; and 33.3% (1/3), 0% (0/4), and 43.8% (28/64) for benign lesions. For adenocarcinoma, the accuracy of risk prediction was slightly lower in SN lesions than in PS and NS lesions. However, the difference was only significant between SN and PS lesions (P<0.05). No significant difference based on consistency was observed

Table 3 Factors affecting the accuracy of malignancy risk prediction by the ICLR

Variable	False (n=92)	True (n=394)	P value
Scanner			0.157
Siemens DF	71 (18.6)	311 (81.4)	
Siemens S16	12 (15.8)	64 (84.2)	
Uct760	9 (32.1)	19 (67.9)	
Slice thickness			0.313
1 mm	78 (20.0)	316 (80.0)	
2 mm	14 (15.2)	78 (84.8)	
Lobe distribution			0.184
Upper	52 (17.1)	252 (82.9)	
Lower	40 (22.0)	142 (78.0)	
Size			0.011
[a]≤10 mm	28 (25.5)	82 (74.5)	
[a]10–20 mm	42 (22.5)	145 (77.5)	
[a]20–30 mm	12 (12.2)	86 (87.8)	
[a]>30 mm	10 (11.0)	81 (89.0)	
Consistency			<0.001
[a]Nonsolid	5 (5.6)	85 (94.4)	
[a]Part-solid	4 (5.3)	71 (94.7)	
[b]Solid	83 (25.9)	238 (74.1)	
Histological types			<0.001
[a]Adenocarcinoma	21 (6.6)	299 (93.4)	
[a]Other malignancy	2 (5.0)	38 (95.0)	
[b]Metastasis	27 (49.1)	28 (50.9)	
[b]Benign	42 (59.2)	29 (40.8)	

Each cell indicates the number (percentage) of patients in the corresponding performance of risk prediction (false, true). [a,b]For groups with the same superscript letters, the difference between the two subgroups was not statistically significant. For groups with different superscript letters, the difference between the two subgroups was statistically significant. ICLR, InferRead CT Lung Research.

in benign lesions (P=0.277). Only SN lesions were observed among other malignancies and metastases.

A comparison of the accuracy of risk prediction between the ICLR and the radiologists

The overall accuracy of risk prediction for R1 and R2 was 93.6% (455/486) and 87.4% (425/486), respectively, which were significantly higher than that for the ICLR (P<0.001 and P=0.001, respectively). R1 performed better than R2 (P=0.001). Representative examples of the performance of the ICLR and the radiologists are shown in *Figure 3*. A comparison of the performance between the ICLR and the two radiologists based on size, consistency, and pathology is shown in *Table 4*.

There were significant differences in the accuracy of malignancy risk prediction between the ICLR and the radiologists in the $size_{10mm}$, $size_{20mm}$, and $size_{30mm}$ subgroups (P=0.005, P<0.001, and P=0.017, respectively; *Figure 4A*). R1 significantly outperformed the ICLR in all three size subgroups (P=0.003, P<0.001, and P=0.014, respectively). R2 only outperformed ICLR in the $size_{20mm}$ subgroup P=0.015. There was no significant difference in accuracy between the radiologists and the ICLR in the $size_{>30mm}$ subgroup (P=0.062).

In terms of lesion consistency, only SN lesions showed a significant difference in the accuracy of risk prediction between the ICLR and the radiologists (*Figure 4B*). The performance of the ICLR was again poorer than that of the two radiologists (both P<0.001). There was no significant difference in accuracy among lesions of other consistencies (NS, P=0.459; PS, P=0.247).

There were significant differences in the radiologists' accuracy and the ICLR in predicting the risks of adenocarcinomas, metastases, and benign lesions (P<0.001, P<0.001, and P=0.012, respectively; *Figure 4C*). R1 significantly outperformed the ICLR in all three histological types (P=0.001, P=0.001, and P=0.010, respectively). R2 only outperformed the ICLR in metastases (P<0.001). There was no significant difference in the accuracy of other malignancies (P=0.247). For adenocarcinomas, a significant difference was observed between the ICLR and the radiologists in assessing solid type lesions (P<0.001) but not in subsolid type lesions (P=0.078).

R1 outperformed R2 in three subgroups ($size_{20mm}$, P=0.027; SN lesions, P=0.008; and adenocarcinoma, P=0.004).

Discussion

The AI-driven commercial CADx product used in this study applied a 3D CNN with DenseNet architecture as a backbone to determine the malignant probability of pulmonary lesions. The experimental results showed that the AI model had promising applications in clinical practice.
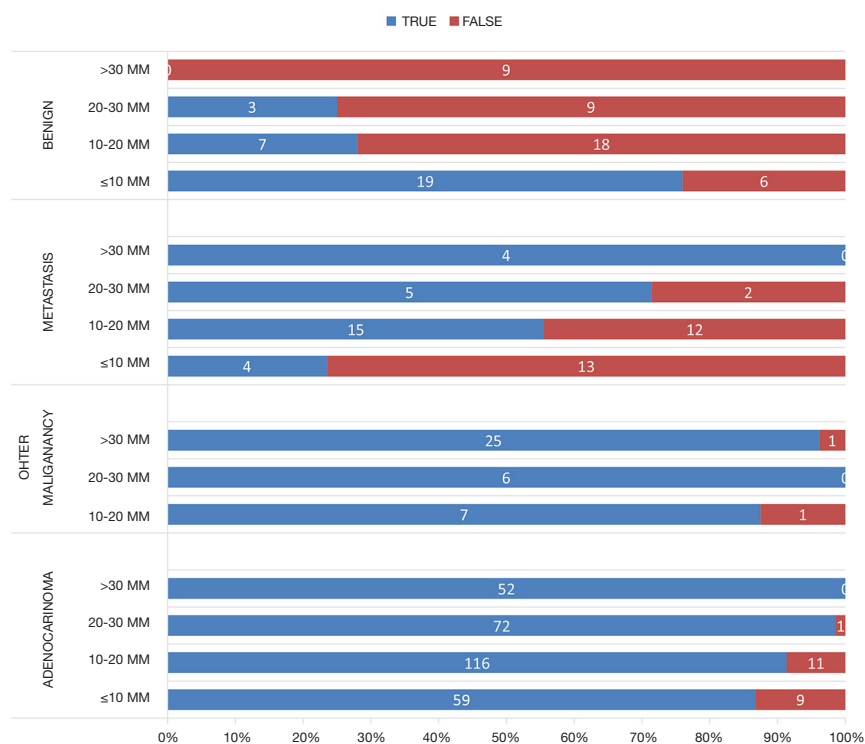
Figure 2 Accuracy of the ICLR for risk prediction of lesions with different pathological types and sizes. For adenocarcinomas and metastases, the accuracy of risk prediction significantly increased with an increase in size (P=0.005 and P=0.016, respectively). The accuracy of risk prediction was significantly higher in the $size_{30mm}$ and $size_{>30mm}$ subgroups than in the $size_{10mm}$ subgroup (both P<0.05). For benign lesions, the accuracy of risk prediction significantly decreased with an increase in size (P<0.001). For other malignant lesions, there was no significant difference in accuracy based on size (P=0.513). ICLR, InferRead CT Lung Research.

The model achieved high performance for detection (92.6%) and risk probability prediction (81.1% with metastasis, 84.9% without metastasis) of pulmonary nodules and masses. This AI product was accurate in the risk prediction of primary lung cancers (93.6%). The accuracy was especially high in adenocarcinomas manifesting as NS or PS nodules (98.1%, 155/158), whereas it was less accurate in solid adenocarcinomas (89%, 139/157). This study is an important supplement to prior investigations using the same software (23). In a previous report, researchers used the National Lung Screening Trial (NLST) database as the training set to build the model and the LIDC-IDRI and Infervision Multi-Center (IMC) databases for verification. They showed excellent performances for the classification of malignancy of lung nodules with the area under the curve (AUC) values of 0.91, 0.86, and 0.95 for receiver operating characteristic curves on the NLST dataset, LIDC-IDRI dataset, and IMC dataset, respectively. The accuracy for classification of benign and malignant lesions using deep learning-based CADx systems reported in the LIDC-IDRI was between 86.84% and 92.3%, with an average AUC of 0.956 (4).

However, this AI product's performance for risk prediction of metastases and benign lesions was poor, with accuracy rates of 50.9% and 40.8%, respectively. The major reason for this poor performance is likely due to bias in the training dataset. With the increase in the availability of CT equipment and lung cancer screening in recent years, the detection of asymptomatic lung cancer has increased. Adenocarcinoma accounts for the vast majority of surgically resected cases, and thus, adenocarcinomas are the main component of the training set. As computer output always comes from input, CADx performance heavily depends on the difficulty and diversity of the training and testing datasets (27). The low accuracy may also be due to the lack of malignant features of metastases, especially when they are small, round and regular, less lobular, and less spiculated, which does not match the AI's malignant features.
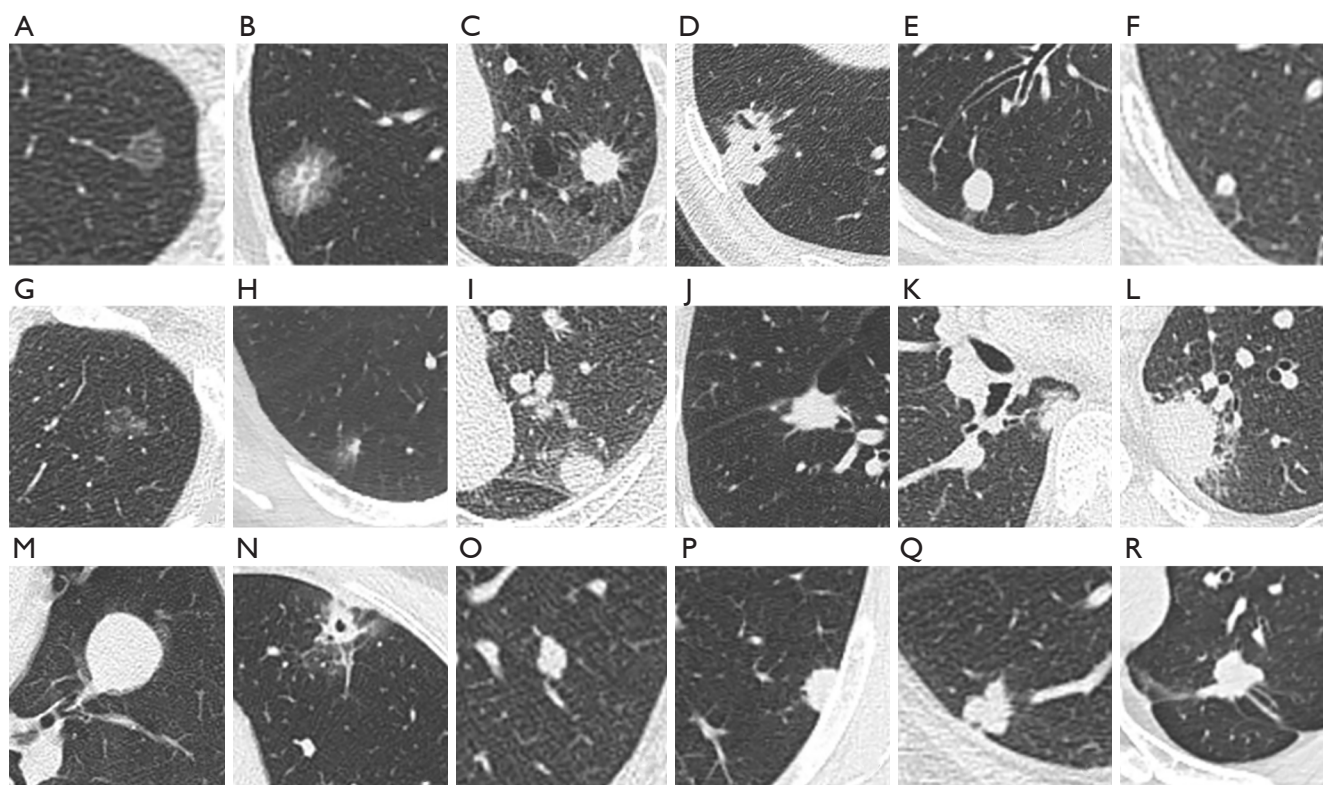
Figure 3 A comparison of the accuracy between the ICLR and the radiologists. The ICLR and both radiologists (R1 and R2) provided accurate risk assessments for the following lesions: (A) 10-mm nonsolid adenocarcinoma; (B) 20-mm part-solid adenocarcinoma; (C) 14-mm solid adenocarcinoma; (D) 38-mm solid other malignancy (adenosquamous); (E) 14-mm solid benign lesion (tuberculosis); (F) 5-mm solid benign lesion (lymph node). The ICLR and both radiologists provided inaccurate risk assessments for the following lesions: (G) 14-mm nonsolid benign lesion (other non-specific); (H) 10-mm part-solid benign lesion (other non-specific); (I) 13-mm solid benign lesion (cryptococcus); (J) 21-mm solid benign lesion (inflammation). R1 provided accurate risk assessments for the following lesions, whereas both ICLR and R2 provided inaccurate assessments: (K) 13-mm solid benign lesion (inflammatory pseudotumor); (L) 66-mm solid benign lesion (inflammation). Both radiologists provided accurate assessments for the following lesions, whereas the ICLR provided inaccurate assessments: (M) 28-mm solid benign lesion (sclerosing pneumocytoma); (N) 32-mm part-solid benign lesion (inflammation); (O) 11-mm solid metastasis; (P) 14-mm solid adenocarcinoma; (Q) 15-mm solid adenocarcinoma; (R) 19-mm solid other malignancy (epithelioid hemangioendothelioma). ICLR, InferRead CT Lung Research.

Additionally, it appears that the AI has learned that large lesions have a high probability of being malignant. Our results showed that the accuracy of prediction of metastases was proportional to size, whereas accuracy was inversely proportional to size in benign lesions.

The observer study in this report showed that the resident doctor undergoing chest training (R1) had a higher diagnostic accuracy compared to the doctor undergoing general radiology training (R2), although both were more accurate than the AI, especially for metastases. The predictive power was roughly equivalent for subsolid adenocarcinomas and other malignancies.

Usually, a resident's diagnostic level is considered to be relatively low (13,28). However, we believe the residents performed well for several reasons. First, they were not required to distinguish between different histological types of adenocarcinoma (for example, AAH, AIS, MIA, and partial IA), which have similar imaging characteristics and are difficult to distinguish, even by experienced chest radiologists (29,30). This greatly reduced the difficulty of differential diagnosis. The reason for not distinguishing between histological types is that nodule consistency on CT is a more significant prognostic indicator than either pleural invasion or parenchymal invasion (angiolymphatic and/or

Table 4 A comparison of the accuracy of malignancy risk prediction between the ICLR and the two radiologists in different size, consistency, and pathological subgroups

Variable	ICLR		R1		R2		P	P1	P2	P3
	True	False	True	False	True	False		ICLR vs. R1	ICLR vs. R2	R1 vs. R2
Total	394 (81.1)	92 (18.9)	455 (93.6)	31 (6.4)	425 (87.4)	61 (12.6)	<0.001	<0.001	0.001	0.001
Size										
≤10 mm	82 (74.5)	28 (25.5)	99 (90.0)	11 (10.0)	91 (82.7)	19 (17.3)	0.005	0.003	0.256	0.378
10–20 mm	145 (77.5)	42 (22.5)	174 (93.0)	13 (7.0)	160 (85.6)	27 (14.4)	<0.001	<0.001	0.015	0.027
20–30 mm	86 (87.8)	12 (12.2)	94 (95.9)	4 (4.1)	91 (92.9)	7 (7.1)	0.017	0.014	0.231	0.867
>30 mm	81 (89.0)	10 (11.0)	88 (96.7)	3 (3.3)	83 (91.2)	8 (8.8)	0.062	NA	NA	NA
Consistency										
Nonsolid	85 (94.4)	5 (5.6)	86 (95.6)	4 (4.4)	83 (92.2)	7 (7.8)	0.459	NA	NA	NA
Part-solid	71 (94.7)	4 (5.3)	73 (97.3)	2 (2.7)	70 (93.3)	5 (6.7)	0.247	NA	NA	NA
Solid	238 (74.1)	83 (25.9)	296 (92.2)	25 (7.8)	272 (84.7)	49 (15.3)	<0.001	<0.001	<0.001	0.008
Histological type										
Adenocarcinoma	299 (93.4)	21 (6.6)	317 (99.1)	3 (0.9)	301 (94.1)	19 (5.9)	<0.001	0.001	>0.99	0.004
Other malignancy	38 (95.0)	2 (5.0)	40 (100)	0 (0)	37 (92.5)	3 (7.5)	0.247	NA	NA	NA
Metastasis	28 (50.9)	27 (49.1)	54 (98.2)	1 (1.8)	52 (94.5)	3 (5.5)	<0.001	0.001	<0.001	>0.99
Benign	29 (40.8)	42 (59.2)	44 (62.0)	27 (38.0)	35 (49.3)	36 (50.7)	0.012	0.010	0.718	0.233

Each cell indicates the number (percentage) of patients in the corresponding performance of risk prediction (false, true). ICLR, InferRead CT Lung Research; NA, non-available; R1, radiologist 1; R2, radiologist 2.

vascular). Lung adenocarcinomas that appear as subsolid or nonsolid, have a much better prognosis than that of lung adenocarcinomas manifesting as SNs (31-33). Second, doctors had access to clinical information and prior images, which were especially beneficial for diagnosing metastases, but the AI did not have access to this information. Ardila and colleagues also demonstrated the importance of prior images. They proposed a deep learning algorithm to predict lung cancer risk and found that when prior imaging was not available, the AI model outperformed radiologists. In contrast, in cases where prior imaging was available, the model performance was on par with that of the same radiologists (34). Third, a resident doctor's participation in research and review of related literature could strengthen their understanding of pulmonary lesions and improve their ability to diagnose lung lesions.

This study had some limitations. First, not all module-specific features were analyzed. However, these features, such as calcification, internal structure, lobulation, spiculation, texture, and subtlety, are important features for

the differential diagnosis of benign and malignant pulmonary nodules. Second, the sample size was unbalanced, with only a small number of non-adenocarcinoma cases. The numbers of the different histological types of adenocarcinomas were also unbalanced. Third, the time required for diagnosis was not assessed and should be included in future studies. Fourth, other currently recognized nodule classification systems, such as LungRADS, were not analyzed, and this may have limited the clinical value of the report. Fifth, the device types and scanning techniques that were not in the model training set will have a certain impact on the model's performance. The generalization of AI models can be improved by including as many types of scanners as possible in the training set. For AI products to be more useful in clinical practice, the training sample size will need to be expanded to optimize the model, and class imbalance issues need to be resolved to make the model more robust. Furthermore, since the best-known CADx schemes distinguish between benign and malignant nodules based on volume doubling time (35,36), embedding clinical information and volume
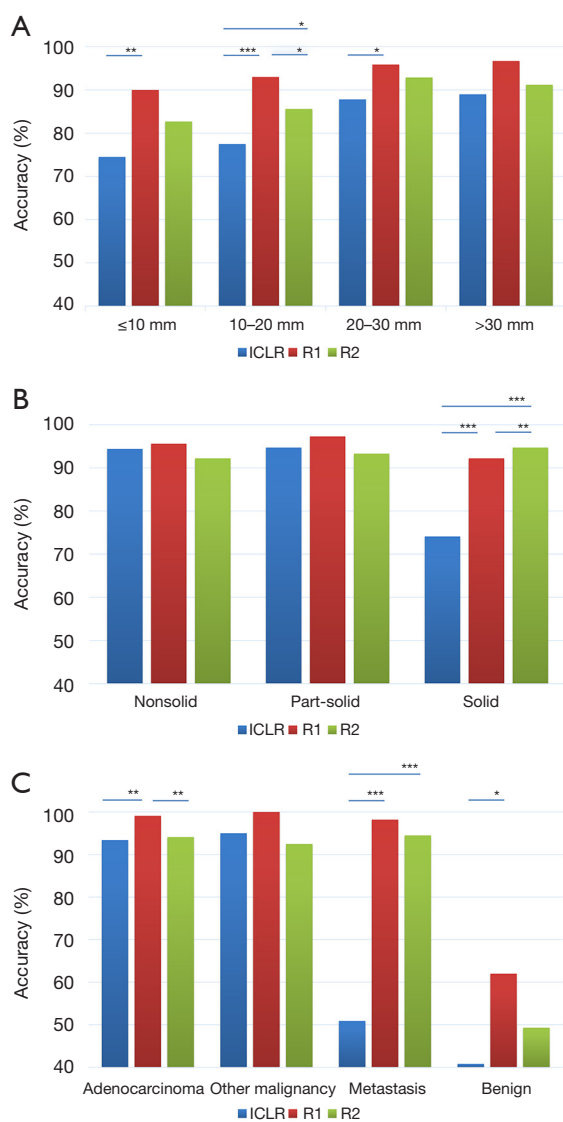
Figure 4 A comparison of the performance of malignancy risk prediction between the ICLR and the two radiologists, stratified by size (A), consistency (B), and pathology (C). *P<0.05, **P<0.01, ***P<0.001. ICLR, InferRead CT Lung Research.

change information in the algorithm in future investigations will be beneficial.

Conclusions

AI malignancy risk prediction for lung nodules and masses with different pathological types is particularly important and useful but complicated and challenging. The ICLR (Infervision, China) had very high accuracy in primary

lung cancer malignancy prediction but poor accuracy in predicting metastases and benign lesions. Further efforts are warranted to augment the number of metastatic and benign lesions in the training dataset to improve the AI products' performance.

Footnote

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/qims-20-1314). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). This study was approved by the ethics committee of the Fifth Affiliated Hospital of Sun Yat-sen University. Informed consent was waived due to the retrospective nature of the study.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: https://creativecommons.org/licenses/by-nc-nd/4.0/.

References

1. Liu S. White paper medical imaging AI in China. Report, Chinese Innovative Alliance of Industry, Education, Research and Application of Artificial Intelligence for Medical Imaging, Beijing, China: 2019:33-4.
2. Fraioli F, Serra G, Passariello R. CAD (computer-aided detection) and CADx (computer aided diagnosis)

systems in identifying and characterizing lung nodules on chest CT: overview of research, developments and new prospects. Radiol Med 2010;115:385-402.

3. Al Mohammad B, Brennan PC, Mello-Thoms C. A review of lung cancer screening and the role of computer-aided detection. Clin Radiol 2017;72:433-42.

4. Wang Y, Wu B, Zhang N, Liu J, Ren F, Zhao L. Research progress of computer aided diagnosis system for pulmonary nodules in CT images. J Xray Sci Technol 2020;28:1-16.

5. Li D, Mikela Vilmun B, Frederik Carlsen J, Albrecht-Beste E, Ammitzbøl Lauridsen C, Bachmann Nielsen M, Lindskov Hansen K. 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. Diagnostics (Basel) 2019;9:E207.

6. Infervision. Available online: https://www.infervision.com/product/5/, Accessed 19 March 2021.

7. Iwasawa T, Matsumoto S, Aoki T, Okada F, Nishimura Y, Yamagata H, Ohno Y. A comparison of axial versus coronal image viewing in computer-aided detection of lung nodules on CT. Jpn J Radiol 2015;33:76-83.

8. Zhao Y, de Bock GH, Vliegenthart R, van Klaveren RJ, Wang Y, Bogoni L, de Jong PA, Mali WP, van Ooijen PM, Oudkerk M. Performance of computer-aided detection of pulmonary nodules in low-dose CT: comparison with double reading by nodule volume. Eur Radiol 2012;22:2076-84.

9. Das M, Mühlenbruch G, Heinen S, Mahnken AH, Salganicoff M, Stanzel S, Günther RW, Wildberger JE. Performance evaluation of a computer-aided detection algorithm for solid pulmonary nodules in low-dose and standard-dose MDCT chest examinations and its influence on radiologists. Br J Radiol 2008;81:841-7.

10. Huang P, Park S, Yan R, Lee J, Chu LC, Lin CT, Hussien A, Rathmell J, Thomas B, Chen C, Hales R, Ettinger DS, Brock M, Hu P, Fishman EK, Gabrielson E, Lam S. Added Value of Computer-aided CT image features for early lung cancer diagnosis with small pulmonary nodules: a matched case-control study. Radiology 2018;286:286-95.

11. Liang M, Tang W, Xu DM, Jirapatnakul AC, Reeves AP, Henschke CI, Yankelevitz D. Low-Dose CT Screening for Lung Cancer: Computer-aided Detection of Missed Lung Cancers. Radiology 2016;281:279-88.

12. Vassallo L, Traverso A, Agnello M, Bracco C, Campanella D, Chiara G, Fantacci ME, Lopez Torres E, Manca A, Saletta M, Giannini V, Mazzetti S, Stasi M, Cerello P,

Regge D. A cloud-based computer-aided detection system improves identification of lung nodules on computed tomography scans of patients with extra-thoracic malignancies. Eur Radiol 2019;29:144-52.

13. Yamada Y, Shiomi E, Hashimoto M, Abe T, Matsusako M, Saida Y, Ogawa K. Value of a computer-aided detection system based on chest tomosynthesis imaging for the detection of pulmonary nodules. Radiology 2018;287:333-9.

14. Liu Y, Balagurunathan Y, Atwater T, Antic S, Li Q, Walker RC, Smith GT, Massion PP, Schabath MB, Gillies RJ. Radiological image traits predictive of cancer status in pulmonary nodules. Clin Cancer Res 2017;23:1442-9.

15. Shen W, Zhou M, Yang F, Yang C, Tian J. Multi-scale convolutional neural networks for lung nodule classification. Inf Process Med Imaging 2015;24:588-99.

16. Liu C, Hu SC, Wang C, Lafata K, Yin FF. Automatic detection of pulmonary nodules on CT images with YOLOv3: development and evaluation using simulated and patient data. Quant Imaging Med Surg 2020;10:1917-29.

17. Hussein S, Cao K, Song Q, Bagci U. Risk stratification of lung nodules using 3D CNN-based multi-task learning. Information Processing in Medical Imaging. Cham, Switzerland: Spring, 2017:249-60.

18. Mao L, Chen H, Liang M, Li K, Gao J, Qin P, Ding X, Li X, Liu X. Quantitative radiomic model for predicting malignancy of small solid pulmonary nodules detected by low-dose CT screening. Quant Imaging Med Surg 2019;9:263-72.

19. Chen H, Liang M, Li X, Wu T, Zhang L, Liu X. An individualised radiomics composite model predicting prognosis of stage I solid lung adenocarcinoma. Clin Radiol 2020;75:562.e11-562.e19.

20. Feng B, Chen X, Chen Y, Lu S, Liu K, Li K, Liu Z, Hao Y, Li Z, Zhu Z, Yao N, Liang G, Zhang J, Long W, Liu X. Solitary solid pulmonary nodules: a CT-based deep learning nomogram helps differentiate tuberculosis granulomas from lung adenocarcinomas. Eur Radiol 2020;30:6497-507.

21. Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Trans Pattern Anal Mach Intell 2017;39:1137-49.

22. Jiang H, Learned-Miller E. Face Detection with the Faster R-CNN. IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017). IEEE 2017;650-7.

23. Wu H, Tang W, Wu C, Deng Y, Zhang R. Highly Robust Prediction of Lung Nodule Malignancy by Deep Learning Model: A Multiracial, Multinational Study. medRxiv 2020.11.24.20237354. Available online: https://doi.org/10.1101/2020.11.24.20237354

24. Henschke CI, Li K, Yip R, Salvatore M, Yankelevitz DF. The importance of the regimen of screening in maximizing the benefit and minimizing the harms. Ann Transl Med 2016;4:153.

25. Travis WD, Brambilla E, Noguchi M, Nicholson AG, Geisinger KR, Yatabe Y, et al. International association for the study of lung cancer/American thoracic society/European respiratory society international multidisciplinary classification of lung adenocarcinoma. J Thorac Oncol 2011;6:244-85.

26. Travis WD, Brambilla E, Nicholson AG, Yatabe Y, Austin JHM, Beasley MB, Chirieac LR, Dacic S, Duhig E, Flieder DB, Geisinger K, Hirsch FR, Ishikawa Y, Kerr KM, Noguchi M, Pelosi G, Powell CA, Tsao MS, Wistuba I; WHO Panel. The 2015 World Health Organization classification of lung tumors: impact of genetic, clinical and radiologic advances since the 2004 classification. J Thorac Oncol 2015;10:1243-60.

27. Zheng B, Wang X, Lederman D, Tan J, Gur D. Computer-aided detection; the effect of training databases on detection of subtle breast masses. Acad Radiol 2010;17:1401-8.

28. Asplund S, Johnsson AA, Vikgren J, Svalkvist A, Boijsen M, Fisichella V, Flinck A, Wiksell A, Ivarsson J, Rystedt H, Månsson LG, Kheddache S, Båth M. Learning aspects and potential pitfalls regarding detection of pulmonary nodules in chest tomosynthesis and proposed related quality criteria. Acta Radiol 2011;52:503-12.

29. Gong J, Liu J, Hao W, Nie S, Wang S, Peng W. Computer-aided diagnosis of ground-glass opacity pulmonary nodules using radiomic features analysis. Phys Med Biol 2019;64:135015.

30. Zhao W, Yang J, Sun Y, Li C, Wu W, Jin L, Yang Z, Ni B, Gao P, Wang P, Hua Y, Li M. 3D deep learning from CT scans predicts tumor invasiveness of subcentimeter pulmonary adenocarcinomas. Cancer Res 2018;78:6881-9.

31. Yip R, Ma T, Flores RM, Yankelevitz D, Henschke CI; International Early Lung Cancer Action Program Investigators. Survival with Parenchymal and Pleural Invasion of Non-small-cell Lung Cancers less than 30 mm. J Thorac Oncol 2019;14:890-902.

32. Yip R, Yankelevitz DF, Hu M, Li K, Xu DM, Jirapatnakul A, Henschke CI. Lung cancer deaths in the National Lung Screening Trial attributed to nonsolid nodules. Radiology 2016;281:589-96.

33. Yip R, Wolf A, Tam K, Taioli E, Olkin I, Flores RM, Yankelevitz DF, Henschke CI. Outcomes of lung cancers manifesting as nonsolid nodules. Lung Cancer 2016;97:35-42.

34. Ardila D, Kiraly AP, Bharadwaj S, Choi B, Reicher JJ, Peng L, Tse D, Etemadi M, Ye W, Corrado G, Naidich DP, Shetty S. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nat Med 2019;25:954-61.

35. Revel MP, Merlin A, Peyrard S, Triki R, Couchon S, Chatellier G, Frija G. Software volumetric evaluation of doubling times for differentiating benign versus malignant pulmonary nodules. AJR Am J Roentgenol 2006;187:135-42.

36. Alahmari SS, Cherezov D, Goldgof D, Hall L, Gillies RJ, Schabath MB. Delta radiomics improves pulmonary nodule malignancy prediction in lung cancer screening. IEEE Access 2018;6:77796-806.