



Artificial intelligence as a diagnostic tool for lung nodule evaluation

Humberto K. Choi^{1^}, Xiaofeng Wang², Peter J. Mazzone^{1^}

¹Respiratory Institute, Cleveland Clinic, Cleveland, OH, USA; ²Quantitative Health Science, Cleveland Clinic, Cleveland, OH, USA

Contributions: (I) Conception and design: All authors; (II) Administrative support: None; (III) Provision of study materials or patients: None; (IV) Collection and assembly of data: None; (V) Data analysis and interpretation: None; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Humberto K. Choi, MD, FCCP. Cleveland Clinic, 9500 Euclid Avenue Mail Code A90, Cleveland, OH 44195, USA.

Email: choih@ccf.org.

Abstract: The clinical evaluation of lung nodules begins with an assessment of the probability that the nodule is malignant. The probability of malignancy can be estimated based on clinical experience or by using validated prediction models that incorporate known clinical and radiographic features associated with lung cancer. There has been a growing interest in the use of artificial intelligence (AI) to aid clinicians in performing this task. Advances in machine learning techniques have fostered progress in the development of automated systems with the goal to match or exceed the performance of clinicians. In this article we will review how lung nodules are currently evaluated, how AI may be used in lung nodule detection and classification, discuss its limitations, future directions and the necessary steps for its application in clinical practice.

Keywords: Lung nodule; lung cancer; machine learning; deep learning; artificial intelligence (AI)

Received: 03 April 2020; Accepted: 25 September 2020; Published: 30 December 2020.

doi: 10.21037/jmai-20-29

View this article at: <http://dx.doi.org/10.21037/jmai-20-29>

Introduction

Lung nodules are radiographic opacities that are <30 mm in diameter, well-circumscribed, and surrounded completely by normal lung parenchyma. An opacity >30 mm is considered a lung mass (1). Lung nodules are typically detected on computed tomography (CT) scans of the chest performed for symptom evaluation (incidentally) or for lung cancer screening. Their clinical evaluation begins with an assessment of the probability that they are malignant (1). The earlier a lung cancer is detected, the higher the survival rates, however most nodules are not malignant. Thus, the goal of nodule management is to expedite the treatment of malignant nodules while minimizing testing for benign nodules (1,2).

The probability of malignancy can be estimated using

clinical experience and intuition. It can also be assessed by using validated prediction models that incorporate known clinical and radiographic features associated with lung cancer (2). The accuracy of these methods is limited, resulting in difficult clinical decisions about how aggressively a lung nodule should be evaluated (2). Diagnostic molecular biomarkers are being studied as adjuncts to prediction methods, to assist with decision making (3). More recently there has been a growing interest in the use of artificial intelligence (AI) to aid clinicians in detecting and classifying lung nodules.

In this article we will review how lung nodules are currently evaluated, how AI has been studied in lung nodule detection and diagnosis, discuss its limitations, future directions and the necessary steps for its application in

[^] ORCID: Humberto K. Choi: <https://orcid.org/0000-0001-9494-6056>; Peter J. Mazzone: <https://orcid.org/0000-0003-2121-0595>.

Table 1 Clinical and radiographic features included in validated prediction models

Model	Predictors of malignancy	AUC
Mayo Clinic (5)	Age, smoking history, history of extrathoracic cancer ≥ 5 years, nodule diameter, nodule spiculation, upper lobe location	0.80
Veteran's Affair (4)	Age, smoking history, time since quitting smoking, nodule diameter	0.79
Herder (6)	Mayo Clinic model and FDG-PET avidity intensity (none/faint/moderate/intense)	0.92
McWilliams/Tammemagi (Brock) (7)	Age, sex, family history of lung cancer, emphysema, nodule size, nodule type, nodule location, nodule count	0.94
Gurney <i>et al.</i> (8,9)	Nodule spiculation, diameter and cavity wall thickness. Predictors of a benign etiology were volume doubling time >465 days and calcification.	0.87
Bayesian Inference Malignancy Calculator (BIMC) (10)	Age, smoking, history of previous malignancy, nodule diameter, edges, nodule location, volume doubling time, minimum focal density, enhancement at contrast enhanced CT, FDG-PET avidity	0.89
Thoracic Research Evaluation And Treatment (TREAT) (11)	Age, sex, BMI, FEV1, smoking history, hemoptysis, nodule size, nodule growth, spiculation, nodule location, FDG-PET avidity	0.87
Peking University People's Hospital (PKUPH) (12)	Age, nodule diameter, nodule border, nodule calcification, spiculation, family history of cancer	0.87
Cleveland Clinic Model (13)	Age, smoking history, upper lobe location, solid and irregular/speculated edges, emphysema, FDG-PET avidity, history of cancer other than lung	c-index 0.81

AUC, area under the curve; BMI, body mass index; CT, computed tomography; CXR, chest radiograph; FDG-PET, fluorodeoxyglucose-positron emission tomography; FEV1, forced expiratory volume in the first second; LDCT, low-dose CT; NR, not reported; PN, pulmonary nodule.

clinical practice.

Lung nodule evaluation

It is a challenge for a clinician to accurately diagnose a lung nodule. The first step is to assess the probability of malignancy (1). The clinical context within which the nodule is found is relevant. For example, a nodule found during the evaluation of a patient with a known cancer raises the suspicion for metastasis, and a nodule found incidentally in a young patient who never smoked cigarettes is likely benign. Clinicians estimate the probability that a nodule is malignant by taking into consideration clinical variables and imaging features that have been shown to be associated with lung cancer (1). The radiographic characteristics that have been independently associated with the likelihood of malignancy include nodule size, density, margin, evidence of growth, the presence of calcification, the presence of fat tissue, and cavitation (1). A clinician will incorporate these imaging features with clinical factors (age, smoking history, history of prior malignancies), and/or use validated clinical prediction models that include these features, to estimate the

probability that the nodule is malignant (2,4). *Table 1* summarizes the clinical and radiographic features included in validated prediction models.

The estimated probability that a nodule is malignant is used to guide management decisions. A probability of malignancy of 5% or less is considered very low risk, between 5% and 65% low to moderate risk and 65% or higher high risk (1). These risk groups are tied to management recommendations (1). Nodules that are classified at very low-risk for malignancy may require no further evaluation or just be followed with serial imaging to assess for growth. Nodules considered to have a high probability of malignancy should be considered for surgical resection, or for diagnostic testing to obtain a pathologic diagnosis. Nodules of low to moderate probability of malignancy are evaluated with additional testing, such as imaging with a fluoro-2-deoxy-D-glucose positron emission tomography (FDG-PET) scan, or by performing a non-surgical biopsy.

There are several limitations to the use of nodule risk prediction models in practice. First, it can be difficult to choose which model to apply as the populations from which the models were derived may not be representative of the

individual to whom they are applied. The accuracy of a prediction model is higher when it is applied in a population that is similar to that used to derive the model. It is not clear how these models perform in demographic minority groups as they are not well represented in the derivation studies. More important than the accuracy of the model outputs is the impact the estimate has on clinical decisions. Expert physician assessment of the probability of malignancy has variably been shown to be more accurate than nodule prediction models. However, physician management decisions may differ from guideline recommendations for the risk group that their patient falls in. One study showed the rate of lung resection in those with low risk nodules was the same as in those with high risk nodules (14).

Artificial intelligence in lung nodule evaluation

In health care, the largest impact of machine learning on big data has been in the field of diagnostic decision support. AI has been studied and implemented for many years in tasks such as computerized electrocardiogram analysis (15). Technological advances have led to growth in research on the use of AI-based systems in medical imaging. AI systems evolved to assist with workflow management, decision support, reader accuracy, or to supplement expertise in underserved areas. Many imaging decision support tasks related to the management of lung nodules could be helped by AI systems.

Radiologists have found AI tools can help improve the efficiency of reading scans. Several studies have shown that nodule detection systems effectively help radiologists identify more nodules (16-18). Brown *et al.* studied the impact of implementing a commercial CT chest based computer-aided detection (CAD) system into the radiology workflow (18). The system provided automated detection and measurement of lung nodules, and it was integrated with a report dictation application. The authors reported reduced reading times by 7–44% compared to conventional manual methods (18).

The sophisticated and rapid analysis of large amounts of data has enabled greater knowledge of populations. The ultimate goal is to reduce cost and improve efficiency and quality of care by customizing population management interventions (19). Lung cancer screening is an area where AI can aid in population management due to challenges such as variability of lung cancer probability estimation between clinicians or radiologists, and high false-positive results (20-22).

Deep learning techniques applied in nodule evaluation

Deep learning techniques have already reached the field of lung nodule evaluation. The typical deep learning network involves feature extraction from the raw images and an objective function that learns the correlation between the features and the diagnosis. Instead of choosing features a priori, a set of training images and their target labels are provided to a deep learning network. The feature extraction portion usually consists of several layers of nonlinear processing units and transformation functions in addition to using conventional image processing operators such as filters. Convolutional Neural Network (CNN) techniques have become the most common pattern detection, segmentation and classification applications in the medical field. Unlike traditional clinical probability models, deep learning techniques build features from scratch rather than from a pre-selected set of features that rely on the contextual knowledge of the algorithm developer.

Fully automated systems using CNN have shown promising results in lung nodule detection. Huang *et al.* demonstrated a fast and fully automated end-to-end system that can efficiently segment precise lung nodule contours from raw thoracic CT scans (23). The system had four components, including candidate nodule detection with “Faster regional”-CNN, candidate merging, false positive reduction, and nodule segmentation with CNN. The entire system had no human interaction or database specific design, and the average runtime was approximately 16 seconds per scan. The nodule detection accuracy was 91.4% with an average of 1 false positive per scan and 94.6% with an average of 4 false positives per scan. Pehrson *et al.* reviewed machine learning algorithms applied to the Lung Image Database Consortium Image Collection (LIDC-IDRI) database as a tool for the optimization of detecting lung nodules in thoracic CT scans. Their review showed that feature-based machine learning and deep learning algorithms can detect lung nodules with a high level of accuracy. Most feature-based algorithms achieved an accuracy >90%. The deep learning algorithms achieved an AUC in the range of 0.82–0.99 (24).

Recently, a deep learning algorithm with an end-to-end approach with both detection and lung cancer risk categorization tasks was used on cases from the National Lung Cancer Screening Trial (25). The proposed algorithm achieved an overall area under the curve of 94.4%, reduced false-positives by 11% and false-negatives by 5% compared to six radiologists.

Table 2 Selected studies of lung nodule detection systems

Authors	Technique	Database	No of nodules	Sensitivity (%)	Specificity (%)	Average number false-positive/scan
Naqi <i>et al.</i> (26)	Morphological operations, edge detection, bounding box, shape information	LIDC-IDRI	567	98.6	98.2	3.4
Saien <i>et al.</i> (27)	Sparse field level sets	LIDC-IDRI	198	83.9	90.7	3.9
Shaukat <i>et al.</i> (28)	Noise removal, segmentation by optimal thresholding, multiscale dot enhancement filtering, support vector machine	LIDC-IDRI	2,242	98.1	96	2.1
Liu <i>et al.</i> (29)	Fast segmentation method, random forest	LIDC	978	92.4	94.8	4.5
Javaid <i>et al.</i> (30)	K-means clustering, morphological filter	LIDC	133	91.6	96.7	3.1
Manikandan <i>et al.</i> (31)	Fuzzy auto-seed cluster means morphological algorithm	Private	801	100	93	0.38
Krishnamurthy <i>et al.</i> (32)	Morphological region-grow segmentation, edge bridge and fill technique	SPIE, AAPM, IDC	257	88	84	2
de Carvalho Filho <i>et al.</i> (33)	Extraction and reconstruction of the pulmonary parenchyma, nodule segmentation, support vector machine	LIDC-IDRI	182	85.9	97.7	1.8
Magalhaes <i>et al.</i> (34)	Growing neutral gas, 3D distance transform	LIDC	48	85.9	90.7	0.1

AAPM, American Association of Physicists in Medicine; LIDC-IDRI, Lung Image Database Consortium; NR, not reported; SPIE, International Society for Optics and Photonics.

Detection systems

Detection systems concentrate on the identification of suspicious lesions in chest images and alert radiologists to them. These systems are designed to improve sensitivity and accuracy in detection, while improving the efficiency of the radiologist by reducing the exam evaluation time. A good example of the potential application of a lung nodule detection system is in the setting of lung cancer screening, which can be a repetitive and difficult task for radiologists as they are often trying to detect small nodules.

Although they may have different structures, detections systems in general consist of 5 steps: image acquisition, preprocessing, lung segmentation, nodule detection and false positive reduction. Image acquisition is a process of acquiring medical images from imaging modalities. CT scans have been the most studied imaging modality. From a technical point of view, thin-section CT scans should be used when applying nodule detection systems. Lung segmentation improves the efficiency of nodule detection by extracting the lung volume from input CT images and removing the background and other irrelevant components. Many different methods have been used including optimal thresholding, rule-based region growing, global

thresholding, 3-D-adaptive fuzzy thresholding, hybrid segmentation, and connected component labeling. After segmentation, the detection of areas suspicious of being a nodule can be done by multiple gray-level thresholding; shape- or template-matching-based morphological techniques or filtering-based methods. Detection of abnormalities that are not true nodules is common enough that a false-positive reduction step is necessary. This step involves feature extraction and nodule classification of “true nodules” and “non-nodules”.

Table 2 summarizes selected studies of lung nodule detection systems that reported sensitivity, specificity and false-positive rates. The overall sensitivity and specificity are high. They ranged from 85.9–100%, and 84–98.7%, respectively. The many different morphologies of nodules impact the false-positive rates. The detection systems need to identify nodules of different sizes and textures (solid, part solid, ground glass) in different locations (isolated, juxtaleural, perifissural or juxtavascular).

Nodule classification and cancer characterization

There are many examples of the successful use of machine

Table 3 Selection of studies that used machine learning techniques for lung nodule classification

Authors	Database	Features	Classifier	AUC
Xie <i>et al.</i> (35)	LIDC-IDRI	Pre-selected, deep features	ANN	0.96
Xie <i>et al.</i> (36)	LIDC-IDRI	Pre-selected, deep features, 3D	CNN	0.95
Zhao <i>et al.</i> (37)	LIDC-IDRI	Deep features	CNN	0.87
Causey <i>et al.</i> (38)	LIDC-IDRI	Pre-selected, deep features, 3D	RF	0.99
Tajbakhsh <i>et al.</i> (39)	Private	Deep features	MTANN, CNN	0.88
Shen <i>et al.</i> (40)	LIDC-IDRI	Deep features, 3D	CNN	0.93

ANN, artificial neural network; AUC, area under the curve; CNN, convolutional neural network; LIDC-IDRI, Lung Image Database Consortium and Image Database Resource Initiative; MTANN, massive-training artificial neural networks; RF, random forest; SVM, support vector machines.

learning techniques to classify and characterize lung nodules. *Table 3* summarizes a selection of studies that used machine learning techniques for lung nodule classification. These studies applied defined feature-based algorithms as well as deep learning techniques; and different types of classifiers including CNN, fuzzy system, and support vector machine learning. They all show high accuracy with AUCs ranging from 87–99%. There is currently no agreement about the most efficient and effective methods (*Table 1*). Recent studies have compared AI systems with existing clinical prediction models and demonstrated that 2 different CNN models were able to classify lung nodule more accurately (41,42). Massion *et al.* trained a CNN model using scans from the National Lung Screening Trial, and its accuracy was compared with the Mayo and Brock clinical prediction models (5,7). The study showed that the deep learning system was more accurate in classifying low and high risk categories compared to the clinical models (41).

Hawkins *et al.* evaluated a cohort of patients with screen-detected lung cancer matched with a cohort of screening subjects with benign nodules. The authors identified 23 stable radiomic features that could predict nodules that would be found to be malignant 1 and 2 years later with accuracies of 80% and 79%, respectively (43). A few studies have demonstrated an association between radiomic features and the invasiveness of adenocarcinomas. Chae *et al.* differentiated part-solid nodules in pre-invasive and invasive adenocarcinoma using texture analysis with high accuracy (44). Maldonado *et al.* developed a computer-aided nodule assessment and risk yield (CANARY) system that builds a unique nodule radiomic signature that correlated well with the degree of tissue invasion and postsurgical patient outcomes (45). There is also evidence that radiomic features

can be helpful in predicting gene expression profiling of non-small cell lung cancer (NSCLC). Liu *et al.* identified 11 radiomic features in peripheral lung adenocarcinomas that distinguished epidermal growth factor receptor (EGFR) mutant from wild types groups (46).

Advances in targeted therapies and immunotherapies have changed the paradigm for the diagnosis of lung cancer (47). The histologic confirmation of malignancy, or the simple differentiation of non-small cell carcinomas and small cell carcinomas, is no longer sufficient. Histologic subtyping, and molecular and genetic characterization are now standard of care as they help determine the most effective treatment (48). The identification of molecular and genetic markers requires invasive procedures to obtain an appropriate amount of tissue for testing. This process can be challenging, influenced by the general health of a patient and tumor location (48). It may prolong the time from diagnosis to treatment. Non-invasive methods to predict the presence of specific molecular and genetic markers may impact patient care. Since CT is routinely performed during lung cancer evaluation, there is an opportunity to assess imaging derived deep learning models as a non-invasive method for molecular characterization of a tumor. Wang *et al.* proposed a deep learning model to predict EGFR mutation status in lung adenocarcinoma using CT scans (49). An end-to-end deep learning model was used to predict the EGFR mutation status by CT scanning in 844 patients with lung adenocarcinoma. This model demonstrated an ability to differentiate EGFR-mutant and EGFR-wild type tumors (AUC 0.85 in the training cohort). Furthermore, this deep learning model showed higher accuracy than machine learning models based on hand-crafted CT features or clinical characteristics.

Challenges with the use of AI

Although AI has shown potential in lung nodule detection and classification, there are challenges that are worth discussing:

- ❖ False-positives in nodule detection: a major concern with the use of AI in lung nodule detection is the high prevalence of false-positive nodules due to their large variations in size, shape and location on CT scans. False-positive reduction techniques are critical to distinguish true lung nodules from other structures that may look like nodules such as blood vessels, pleura and atelectasis. False-positives may be a burden to the radiologist, lead to inaccurate judgments, and be a barrier to widespread use of detection systems. They may also cause emotional stress to patients, lead to unnecessary tests or procedures that carry their own risks, and increase the cost of care. Improvements in false-positive reduction techniques and the use of large public databases have helped mitigate this limitation (50).
- ❖ Lack of reasoning: the loss of “explainability” is one of the possible consequences of the use of AI for lung nodule classification. The malignancy prediction models that are currently used are mostly based on risk factors that are known to be associated with lung cancer and are connected to a biologic process so the reasoning is understandable. When AI is used, the patterns recognized are not necessarily associated with a biological process and no reasoning is provided for the output. It can also be challenging to change or remove an association that has already been made. This has several implications. For example, it may be difficult to explain to a patient who inquires why a nodule was classified as high risk by AI if a clinical prediction model calculator would classify it as low risk. A challenging question is whether clinicians and patients would follow the recommendation of a computer system when they cannot understand its reasoning.
- ❖ Comprehensive evaluation of scans beyond nodules: AI systems are typically focused on detecting and classifying nodules in the lung parenchyma. They do not include extra-pulmonary features that may be present in other locations such as the mediastinum and abdominal organs. For example, a lung nodule can be correctly classified

as high probability for malignancy and the output recommendation could be a transthoracic needle biopsy or lung resection. However the presence of mediastinal adenopathy would indicate consideration for a bronchoscopy with endobronchial ultrasound, or a concomitant liver lesion would indicate liver biopsy as the next step instead of a lung nodule biopsy.

- ❖ Liability issues: there are legal and ethical questions surrounding AI. Some of the questions related to the use of AI for lung nodule evaluation include who is ultimately responsible for its results and potential clinical recommendations, and whether the use of AI would require informed consent. AI in health care has not faced any major legal challenge but further research and understanding of the legal and ethical implications certainly need to be part of the process of bringing this technology to clinical practice.
- ❖ The scarcity of labelled data: deep neural network architectures need to be trained on the entire image or larger sections of images. Creating ground truth for medical images consumes time and requires training. One large publicly accessible database of lung nodules was completed is the LIDC-IDRI which was a collaboration of seven academic centers and eight medical imaging companies (51). Each case required two-phase image annotation from four experienced thoracic radiologists. Even when the imaging data is adequate, finding an early stage malignant nodule in the CT scan is like finding a needle in the haystack. To develop a deep neural network that can find the needle requires many good signals in the training data set. This requires very large data sets, of which few are available.
- ❖ Confidentiality of data: an ideal deep learning model for lung nodule characterization would incorporate patient demographic, clinical, biomarker and radiographic features, as well as learn when it receives new data. To reach this point a large amount of patient data will be needed, bringing with it data storage security and patient privacy concerns.

Steps to implementation of AI systems in clinical practice

In April 2019, the Food and Drug Administration (FDA)

released a proposed regulatory framework for medical devices that use AI algorithms (52). The framework was issued for discussion purposes and it was not intended to be a guideline. The FDA is seeking industry feedback on the proposal with the goal of issuing a draft guideline in the future. The approach would allow the FDA's regulatory oversight to take into consideration the nature of AI products while ensuring that its standards for safety and effectiveness are maintained. Last year, the FDA approved the first AI-based medical devices. One device is used for detecting retinopathy and the other for alerting providers of a potential stroke. However, the AI based products that the FDA cleared use "locked" algorithms that do not "learn" in real-time. The locked algorithms are modified by manufacturers on a periodic basis, which involves training the algorithm using new data and manual verification and validation of the updated algorithm. In contrast, "adaptive" algorithms do not need manual modification to learn and adapt. These algorithms learn from new user data through continuous use.

As any software is a medical device, products that use AI-based algorithms are expected to be assessed for their analytical and clinical validity. Analytical validation demonstrates that a specific device is suitable for its intended use by showing similar output when provided with similar input at separate times. Clinical validation demonstrates diagnostic accuracy, a measure of how well a test detects or predicts a clinical outcome. Despite analytical and clinical validation, clinical utility requires that clinicians and patients accept the AI-based algorithms for diagnosis and decision making, and that decisions based on the algorithms lead to more benefit than harm. The FDA clearance process is focused on the demonstration of safety and effectiveness. Although demonstration of clinical utility is not required in the proposed regulatory framework, it will be necessary for this technology to achieve wide acceptance in clinical practice. Successful demonstration of clinical utility involves developing sufficient evidence to demonstrate that a diagnostic test results in an improvement in patient outcomes. For example, a clinically useful AI system for lung nodule evaluation should help improve the balance of benefit to harm of diagnostic tests by leading to fewer lung cancer deaths, help expedite the treatment for early lung cancer, and help avoid unnecessary invasive procedures for benign nodules without delaying the diagnosis of malignant nodules. Other potential outcomes that would be considered clinically useful include improvement in the efficiency of care, compliance with care

recommendations and cost-effectiveness.

Conclusions

AI systems may complement or augment traditional lung nodule evaluation. Deep learning algorithms are promising and expected to impact how lung nodules will be managed in the future. The application of these algorithms may assist with population management of nodules detected incidentally or by screening. Improvements in nodule characterization will aid in decision making about the need for additional diagnostic testing. Clinical utility remains to be proven—will patients and clinicians accept the AI conclusions without understanding how the conclusions were achieved, and will decisions be changed to the benefit of patients. Future advances need to be focused on perfecting the techniques, mitigating false-positive results, assessment of clinical utility, and how to best translate the results to patients.

Acknowledgments

Funding: None.

Footnote

Peer Review File: Available at <http://dx.doi.org/10.21037/jmai-20-29>

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <http://dx.doi.org/10.21037/jmai-20-29>). Dr. PJM reports grants from Veracyte, grants from Oncocyte, grants from SEER, grants from Exact Sciences, grants from Grail, grants from MagArray, grants from Tencent, outside the submitted work. The other 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.

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

- Gould MK, Donington J, Lynch WR, et al. Evaluation of individuals with pulmonary nodules: when is it lung cancer? Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest* 2013;143:e93S-e120S.
- Choi HK, Ghobrial M, Mazzone PJ. Models to Estimate the Probability of Malignancy in Patients with Pulmonary Nodules. *Ann Am Thorac Soc* 2018;15:1117-26.
- Seijo LM, Peled N, Ajona D, et al. Biomarkers in Lung Cancer Screening: Achievements, Promises, and Challenges. *J Thorac Oncol* 2019;14:343-57.
- Gould MK, Ananth L, Barnett PG, et al. A clinical model to estimate the pretest probability of lung cancer in patients with solitary pulmonary nodules. *Chest* 2007;131:383-8.
- Swensen SJ, Silverstein MD, Ilstrup DM, et al. The probability of malignancy in solitary pulmonary nodules. Application to small radiologically indeterminate nodules. *Arch Intern Med* 1997;157:849-55.
- Herder GJ, van Tinteren H, Golding RP, et al. Clinical prediction model to characterize pulmonary nodules: validation and added value of 18F-fluorodeoxyglucose positron emission tomography. *Chest* 2005;128:2490-6.
- McWilliams A, Tammemagi MC, Mayo JR, et al. Probability of cancer in pulmonary nodules detected on first screening CT. *N Engl J Med* 2013;369:910-9.
- Gurney JW. Determining the likelihood of malignancy in solitary pulmonary nodules with Bayesian analysis. Part I. Theory. *Radiology* 1993;186:405-13.
- Gurney JW, Lyddon DM, McKay JA. Determining the likelihood of malignancy in solitary pulmonary nodules with Bayesian analysis. Part II. Application. *Radiology* 1993;186:415-22.
- Soardi GA, Perandini S, Motton M, et al. Assessing probability of malignancy in solid solitary pulmonary nodules with a new Bayesian calculator: improving diagnostic accuracy by means of expanded and updated features. *Eur Radiol* 2015;25:155-62.
- Deppen SA, Blume JD, Aldrich MC, et al. Predicting lung cancer prior to surgical resection in patients with lung nodules. *J Thorac Oncol* 2014;9:1477-84.
- Li Y, Wang J. A mathematical model for predicting malignancy of solitary pulmonary nodules. *World J Surg* 2012;36:830-5.
- Reid M, Choi HK, Han X, et al. Development of a Risk Prediction Model to Estimate the Probability of Malignancy in Pulmonary Nodules Being Considered for Biopsy. *Chest* 2019;156:367-75.
- Tanner NT, Aggarwal J, Gould MK, et al. Management of Pulmonary Nodules by Community Pulmonologists: A Multicenter Observational Study. *Chest* 2015;148:1405-14.
- Heden B, Ohlin H, Rittner R, et al. Acute myocardial infarction detected in the 12-lead ECG by artificial neural networks. *Circulation* 1997;96:1798-802.
- Awai K, Murao K, Ozawa A, et al. Pulmonary nodules at chest CT: effect of computer-aided diagnosis on radiologists' detection performance. *Radiology* 2004;230:347-52.
- Beigelman-Aubry C, Raffy P, Yang W, et al. Computer-aided detection of solid lung nodules on follow-up MDCT screening: evaluation of detection, tracking, and reading time. *AJR Am J Roentgenol* 2007;189:948-55.
- Brown M, Browning P, Wahi-Anwar MW, et al. Integration of Chest CT CAD into the Clinical Workflow and Impact on Radiologist Efficiency. *Acad Radiol* 2019;26:626-31.
- Anoushiravani AA, Patton J, Sayeed Z, et al. Big Data, Big Research: Implementing Population Health-Based Research Models and Integrating Care to Reduce Cost and Improve Outcomes. *Orthop Clin North Am* 2016;47:717-24.
- van Riel SJ, Jacobs C, Scholten ET, et al. Observer variability for Lung-RADS categorisation of lung cancer screening CTs: impact on patient management. *Eur Radiol* 2019;29:924-31.
- Martin MD, Kanne JP, Broderick LS, et al. Lung-RADS: Pushing the Limits. *Radiographics* 2017;37:1975-93.
- Mehta HJ, Mohammed TL, Jantz MA. The American College of Radiology Lung Imaging Reporting and Data System: Potential Drawbacks and Need for Revision. *Chest* 2017;151:539-43.
- Huang X, Sun W, Tseng TB, et al. Fast and fully-automated detection and segmentation of pulmonary nodules in thoracic CT scans using deep convolutional neural networks. *Comput Med Imaging Graph* 2019;74:25-36.
- Pehrson LM, Nielsen MB, Ammitzbol Lauridsen C. Automatic Pulmonary Nodule Detection Applying Deep Learning or Machine Learning Algorithms to the LIDC-

- IDRI Database: A Systematic Review. *Diagnostics* (Basel) 2019;9:29.
25. 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. *Nat Med* 2019;25:954-61.
 26. Naqi SM, Sharif M, Yasmin M. Multistage segmentation model and SVM-ensemble for precise lung nodule detection. *Int J Comput Assist Radiol Surg* 2018;13:1083-95.
 27. Saïen S, Moghaddam HA, Fathian M. A unified methodology based on sparse field level sets and boosting algorithms for false positives reduction in lung nodules detection. *Int J Comput Assist Radiol Surg* 2018;13:397-409.
 28. Shaukat F, Raja G, Gooya A, et al. Fully automatic detection of lung nodules in CT images using a hybrid feature set. *Med Phys* 2017;44:3615-29.
 29. Liu JK, Jiang HY, Gao MD, et al. An Assisted Diagnosis System for Detection of Early Pulmonary Nodule in Computed Tomography Images. *J Med Syst* 2017;41:30.
 30. Javaid M, Javid M, Rehman MZ, et al. A novel approach to CAD system for the detection of lung nodules in CT images. *Comput Methods Programs Biomed* 2016;135:125-39.
 31. Manikandan T, Bharathi N. Lung Cancer Detection Using Fuzzy Auto-Seed Cluster Means Morphological Segmentation and SVM Classifier. *J Med Syst* 2016;40:181.
 32. Krishnamurthy S, Narasimhan G, Rengasamy U. Three-dimensional lung nodule segmentation and shape variance analysis to detect lung cancer with reduced false positives. *Proc Inst Mech Eng H* 2016;230:58-70.
 33. de Carvalho Filho AO, de Sampaio WB, Silva AC, et al. Automatic detection of solitary lung nodules using quality threshold clustering, genetic algorithm and diversity index. *Artif Intell Med* 2014;60:165-77.
 34. Magalhaes Barros Netto S, Correa Silva A, Acatauassu Nunes R, et al. Automatic segmentation of lung nodules with growing neural gas and support vector machine. *Comput Biol Med* 2012;42:1110-21.
 35. Xie Y, Zhang J, Xia Y, et al. Fusing texture, shape and deep model-learned information at decision level for automated classification of lung nodules on chest CT. *Inf Fusion* 2018;42:102-10.
 36. Xie Y, Xia Y, Zhang J, et al. Knowledge-based Collaborative Deep Learning for Benign-Malignant Lung Nodule Classification on Chest CT. *IEEE Trans Med Imaging* 2019;38:991-1004.
 37. Zhao X, Liu L, Qi S, et al. Agile convolutional neural network for pulmonary nodule classification using CT images. *Int J Comput Assist Radiol Surg* 2018;13:585-95.
 38. Causey JL, Zhang J, Ma S, et al. Highly accurate model for prediction of lung nodule malignancy with CT scans. *Sci Rep* 2018;8:9286.
 39. Tajbakhsh N, Suzuki K. Comparing two classes of end-to-end machine-learning models in lung nodule detection and classification: MTANNs vs. CNNs. *Pattern Recognition* 2017;63:476-86.
 40. Shen W, Zhou M, Yang F, et al. Multi-crop Convolutional Neural Networks for lung nodule malignancy suspiciousness classification. *Pattern Recognition* 2017;61:663-73.
 41. Massion PP, Antic S, Ather S, et al. Assessing the Accuracy of a Deep Learning Method to Risk Stratify Indeterminate Pulmonary Nodules. *Am J Respir Crit Care Med* 2020;202:241-9.
 42. Baldwin DR, Gustafson J, Pickup L, et al. External validation of a convolutional neural network artificial intelligence tool to predict malignancy in pulmonary nodules. *Thorax* 2020;75:306-12.
 43. Hawkins S, Wang H, Liu Y, et al. Predicting Malignant Nodules from Screening CT Scans. *J Thorac Oncol* 2016;11:2120-8.
 44. Chae HD, Park CM, Park SJ, et al. Computerized texture analysis of persistent part-solid ground-glass nodules: differentiation of preinvasive lesions from invasive pulmonary adenocarcinomas. *Radiology* 2014;273:285-93.
 45. Maldonado F, Boland JM, Raghunath S, et al. Noninvasive characterization of the histopathologic features of pulmonary nodules of the lung adenocarcinoma spectrum using computer-aided nodule assessment and risk yield (CANARY)--a pilot study. *J Thorac Oncol* 2013;8:452-60.
 46. Liu Y, Kim J, Balagurunathan Y, et al. Radiomic Features Are Associated With EGFR Mutation Status in Lung Adenocarcinomas. *Clin Lung Cancer* 2016;17:441-8.e6.
 47. Hirsch FR, Scagliotti GV, Mulshine JL, et al. Lung cancer: current therapies and new targeted treatments. *Lancet* 2017;389:299-311.
 48. Rodriguez-Canales J, Parra-Cuentas E, Wistuba, II. Diagnosis and Molecular Classification of Lung Cancer. *Cancer Treat Res* 2016;170:25-46.
 49. Wang S, Shi J, Ye Z, et al. Predicting EGFR mutation status in lung adenocarcinoma on computed tomography image using deep learning. *Eur Respir J* 2019;53:1800986.
 50. Zhang G, Jiang S, Yang Z, et al. Automatic nodule detection for lung cancer in CT images: A review. *Comput*

- Biol Med 2018;103:287-300.
51. Armato SG 3rd, 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. *Med Phys* 2011;38:915-31.
 52. Proposed regulatory framework for modifications

doi: 10.21037/jmai-20-29

Cite this article as: Choi HK, Wang X, Mazzone PJ. Artificial intelligence as a diagnostic tool for lung nodule evaluation. *J Med Artif Intell* 2020;3:16.

to artificial intelligence/machine learning (AI/NK)-based software as a medical device (SaMD). 2019. Available online: <https://www.regulations.gov/contentStreamer?documentId=FDA-2019-N-1185-0001&attachmentNumber=1&contentType=pdf>. Accessed April 22nd 2019.