OPEN

# The diagnostic accuracy of artificial intelligence in thoracic diseases
## A protocol for systematic review and meta-analysis

Yi Yang, MM[a,b], Gang Jin, MM[b], Yao Pang, MM[b], Wenhao Wang, MM[b], Hongyi Zhang, MM[b], Guangxin Tuo, MM[a,b], Peng Wu, MM[a,b], Zequan Wang, MM[b], Zijiang Zhu, PhD[a,b,*]

## Abstract

**Introduction:** Thoracic diseases include a variety of common human primary malignant tumors, among which lung cancer and esophageal cancer are among the top 10 in cancer incidence and mortality. Early diagnosis is an important part of cancer treatment, so artificial intelligence (AI) systems have been developed for the accurate and automated detection and diagnosis of thoracic tumors. However, the complicated AI structure and image processing made the diagnosis result of AI-based system unstable. The purpose of this study is to systematically review published evidence to explore the accuracy of AI systems in diagnosing thoracic cancers.

**Methods and analysis:** We will conduct a systematic review and meta-analysis of the diagnostic accuracy of AI systems for the prediction of thoracic diseases. The primary objective is to assess the diagnostic accuracy of thoracic cancers, including assessing potential biases and calculating combined estimates of sensitivity, specificity, and area under the receiver operating characteristic curve (AUC). The secondary objective is to evaluate the factors associated with different models, classifiers, and radiomics information. We will search databases such as PubMed/MEDLINE, Embase (via OVID), and the Cochrane Library. Two reviewers will independently screen titles and abstracts, perform full article reviews and extract study data. We will report study characteristics and assess methodological quality using the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) tool. RevMan 5.3 and Meta-disc 1.4 software will be used for data synthesis. If pooling is appropriate, we will produce summary receiver operating characteristic (SROC) curves, summary operating points (pooled sensitivity and specificity), and 95% confidence intervals around the summary operating points. Methodological subgroup and sensitivity analyses will be performed to explore heterogeneity.

**PROSPERO registration number:** CRD42019135247

**Abbreviations:** AI = artificial intelligence, AUC = area under the receiver operating characteristic curve, CAD = computer-aided detection, CNN = convolutional neural networks, LDCT = low-dose helical computed tomography, SROC = summary receiver operating characteristic, SVM = support vector machine, VOI = volume of interest.

**Keywords:** artificial intelligence, diagnostic accuracy, meta-analysis, protocol, thoracic cancers

## 1. Introduction

Thoracic diseases include a variety of common human primary malignant tumors, among which lung cancer and esophageal cancer are among the top 10 in cancer incidence and mortality.[1] Though the survival rates of most cancers have improved over the past few decades, lung cancer survival rates have declined, mainly because the cancer is often well advanced, with limited treatment options by the time it is detected.[2] The 5-year survival rate of patients with advanced non-small cell lung cancer is 14% to 15.7%.[3] In contrast, early stages are often curable, and detection during the early stage could drastically improve patient outcomes, minimize overtreatment, and even save lives.[4–6] Although low-dose helical computed tomography (LDCT) can detect more patients with early-stage lung cancer and has few complications, it still requires radiation exposure, has a high false positive rate and may lead to potential overdiagnosis[7]; moreover, 96.4% of lung nodules detected by LDCT screening are not cancerous, and there is no definite method to classify whether these nodules are cancerous or benign, which can lead to a greater economic and psychological burden to patients.[5]

With the increasing awareness of cancer prevention and the vast amount of data generated every day from parallel streams of screening, doctors need to analyze a large number of medical images. Therefore, it is becoming increasingly important to effectively and accurately identify lesions and optimize and simplify the clinical workflow.[8,9] These new challenges have prompted scientists to develop a reliable early detection method that is not a new imaging technology but a way to address challenges in areas such as cancer detection and monitoring based on medical imaging data.[10,11] It is hoped that this technique can simulate the process of traditional imaging diagnosis, which mainly relies on the qualitative characteristics (texture, intensity shape, etc) of the lesions, the outline of the tumor, and anatomic relationships. The focus of this technique is to master these phenotypic descriptions of lesions, known as "semantic characteristics."[12,13] Decades ago, scientists began trying to explore a way to use third-party tools to help clinicians extract semantic features. In recent years, with the development of computer equipment and radiomics, artificial intelligence (AI) has made great progress in medical imaging diagnosis of diseases.[11,14,15]

AI is a science that applies intelligent machines and systems to imitate the ability of intelligent human activities. The use of AI in the diagnosis of thoracic tumors has come a long way. In 1956, computer-aided detection (CAD)[16] was first proposed to predict the existence of pulmonary nodules. With advances in computer technology, CAD was implemented in the 1980s as a tool for the automatic detection of suspicious pulmonary nodules and their location for radiologists.[17,18] However, due to the difficulty in modeling and low accuracy, CAD has not been well developed. During this period, LeCun designed the first truly convolutional neural network, and the emergence of support vector machine (SVM), random forest, and LSTM also laid the foundation for the subsequent rise of AI. In the last decade, these technologies, together with deep learning, have pushed thoracic cancer diagnosis toward the direction of AI,[19–21] especially for the 3 important functions of disease diagnosis: detection, characterization, and monitoring. AI-based detection tools can be used to reduce oversight, track disease progression, and improve diagnostic accuracy; for example, they can be used to detect and monitor the presence of small nodules in thoracic LDCT and distinguish benign and malignant tumors in thoracic LDCT.[4,22–24] This greatly alleviates the challenges of clinical work.

## 1.1. Background
### 1.1.1. Artificial intelligence.
AI overlaps with computer science and robotics, but it focuses more on empirically processing machine behavior. It is similar to studying animal behavior by combining evolutionary characteristics shaped by physiology, biochemistry, ecology, and the environment.[25,26] AI has shown significant impact and potential value in the following 4 areas related to clinical work: managing population health; as a frontline health worker virtual health assistant; as a patient virtual health assistant; and as physician clinical decision support tools.[4] AI plays important roles in the diagnosis and treatment of tumors, including providing accurate descriptions of changes in disease over time, such as changes in tumor size over time; parallel tracking of multiple lesions; and detecting the association of subtle phenotypic differences within tumors with genotypes[27,28]; thus, AI has the potential to improve clinicians' ability to predict outcomes comprehensively. In simple terms, AI is a classifier that imports the feature semantics required for classification into the classifier and then differentiates the input

images through complex image feature analysis to achieve the purpose of diagnosis.[29–31] In addition, AI can also aggregate multiple streams of data into powerful integrated diagnostic systems, such as combining radiographic images with genomic data and pathological grades to thereby complement clinical decision making. It is worth mentioning that deep learning, as a subset of AI, has the ability of automatic learning by mining data to find links that humans cannot detect.[4,21] However, there have been no studies to evaluate which model is better at diagnosing thoracic tumors than traditional machine learning.

### 1.1.2. Machine learning.
Machine learning algorithms can be divided into 3 types: supervised learning, unsupervised learning, and reinforcement learning. Among them, the development of supervised learning has led to the best results.[32,33] Supervised learning obtains a model by training samples and then uses this model for diagnosis. SVM, random forest, and decision tree are the most widely used in the field of thoracic diseases.[34,35] Between 1980 and 2010, machine learning has eclipsed neural networks. Until recent years, there have been many studies using this type of model. However, the development of unsupervised learning in this field has been slow. Although smooth results have been obtained by clustering algorithms (such as the EM algorithm)[36] and data dimensionality reduction (such as Laplace feature mapping),[37] no representative model has been used for thoracic disease diagnosis.

SVM, a representative machine learning method first proposed by Cortes and Vapnik in 1995, has shown many unique advantages in solving the problems of small sample sizes and nonlinearity as well as in high-dimensional pattern recognition and can be generalized and applied to other machine learning problems such as function fitting. SVM is a binary classification model whose basic model is defined as the largest linear classifier in the feature space, which can be transformed into the solution of a convex quadratic programming problem. SVM is a milestone in the diagnosis of thoracic cancers by AI. By implicitly mapping input vectors to high-dimensional space, SVM can handle nonlinear problems well and improve the efficiency of diagnosis.[38,39]

### 1.1.3. Deep learning.
The development of deep learning is closely related to neural networks. Convolutional neural networks (CNN) and LSTM have been proposed as early as the 1980s.[21] Unfortunately, neural networks have not been used successfully because of their long time calculation,[40] and they are at a disadvantage compared with machine learning algorithms such as SVM. However, with the increase in data flow and the improvement of accuracy requirements, the ability of machine learning technology to process raw data has begun to show deficiencies. With the advent of high-efficiency GPUs, ReLUs, dropout layer, cross-validation, and other devices and technologies, neural networks have once again become remarkable.[41–43] In 2006, the deep belief network proposed by Hinton opened a new era of deep learning. He proposed to use a neural network to reduce the dimensionality of data, which is called sparse coding or automatic coding.[44]

At the same time, the CNN, a particular type of deep, feedforward network, has had many practical successes, even during the period when neural networks were out of favor. However, it was not until 2012, when Alex Net won the ImageNet classification contest, that CNN again attracted attention. CNN is a multilayer structure that is composed of a

convolutional layer, nonlinear processing unit, and subsampling layer. The convolutional layer is the core of the CNN and is used to extract the corresponding features of images. Its functions are similar to those of filters (or similar to the retina of the eye). Then, the result obtained by the convolution layer is activated by nonlinear mapping (ReLU). The next pooling layer will compress the input characteristic graph, reduce the spatial dimension of the input data, and simplify the network computing complexity, including maximum pooling and average pooling. To avoid overfitting in this process, the dropout layer regularizes the information, then reduces the dimensions of the output characteristic graph, changes the 2D matrix to 1D, multiplies the weight, and finally passes the output value to the classifier (such as the softmax classifier).[45,46]

### 1.1.4. Radiomics.
Radiomics information, which comes from the high-throughput mining of large-scale image features from medical images, enables data to be quantified and applied to the process of AI systems, which significantly affects the accuracy of diagnosis. Radiomics typically consists of 5 components: data selection, medical imaging, feature extraction, exploratory analysis, and modeling. First, we need to determine 3 points: the imaging protocol, the volume of interest (VOI), and a prediction target. To avoid unnecessary confounding variability, a standard imaging protocol should be selected. Then, there is image segmentation, where VOIs are obtained manually or (semi-) automatically. This stage usually determines which voxels in the image are analyzed, so the process is prone to introducing biases in evaluating derived radiological features, which may also be a factor in the differences between different diagnostic models. Next comes the core of radiomics, which is the high-throughput extraction of quantitative image features to characterize VOIs. The feature eigenvalues depend on factors that can include image preprocessing (e.g., filtering or intensity) and reconstruction (e.g., iterative reconstruction). In addition, it is also influenced by tag naming, algorithms, and software implementation of the feature extraction algorithm. Next, the features from the previous step are combined with the prediction target to create a single dataset. Finally, the previously generated single dataset is combined with the AI model to form the "semantic feature" set, and the images that are later used for classification are also quantified according to the above steps.[15,47,48]

### 1.2. Objective

However, the complicated AI structure and image processing made the diagnosis result of AI-based system unstable. Therefore, to establish a more comprehensive and extensive summary of the AI diagnostic models for thoracic cancers, we will conduct a systematic review and meta-analysis. Our main objective is to assess the accuracy of the diagnosis of thoracic diseases, including assessing potential biases and calculating combined estimates of sensitivity, specificity, and area under the receiver operating characteristic curve (AUC). The second objective is to evaluate factors related to diagnostic accuracy and different models (machine learning vs deep learning), classifiers (including CNN, SVM, random forests, LSTM, etc), and radiomics information (image segmentation methods, feature extraction methods).

## 2. Methods

### 2.1. Search strategy

This review adopts recommendations based on the conduct and reporting of systematic reviews and meta-analyses outlined by the Preferred Reporting Items for Systematic Reviews and Meta-analyses statement. We will search the following databases for relevant studies: PubMed/MEDLINE, Embase (via OVID), and the Cochrane Library, for reports on the diagnostic accuracy of AI in thoracic diseases published between 1946 and December 2019. There will be no language restrictions. We developed a search strategy using a combination of keywords and medical subject headings (MeSH)/EMTREE terms, and the following expressions will be used:(Endoscope OR gastroscope OR "Spiral Cone-Beam Computed Tomography" OR CT OR CAT) AND (((nodule OR cancer OR tumor) AND (lung OR pulmonary OR thoracic)) OR (("squamous cell carcinoma" OR carcinoma OR cancer OR neoplasm OR "adenosquamous carcinoma" OR "basosquamous carcinoma" OR Barrett's) AND (esophageal) OR esophagus))) AND (Computer aided OR Machine learning OR Deep learning OR Algorithms OR Artificial Intelligence).

### 2.2. Eligibility criteria and exclusion criteria
#### 2.2.1. Eligibility criteria.
All the articles will be assessed independently by 2 researchers. The inclusion criteria are as follows: assessment of the diagnostic accuracy of AI for diagnosing thoracic cancer (including early lung cancer and early esophageal cancer); medical imaging data including CT images and white light endoscopic images; and inclusion of sensitivity, specificity, or sufficient information to construct the $2 \times 2$ tables (outcome).

#### 2.2.2. Exclusion criteria.
We will exclude studies from which we cannot obtain or calculate the data from the text, appendices, or after contacting the main authors; moreover, duplicates or subcohorts of already published cohorts, reviews, case reports, conference abstracts, animal experiments, and meta-analyses will be excluded from the study.

### 2.3. Study selection and data extraction

We will use EndNote software (Thomson Reuters, Toronto, Ontario, Canada) to store and remove duplicate references. The articles will be screened independently by the 2 authors (GJ, GXT) based on the titles and abstracts, and those not meeting the criteria will be eliminated. Disagreements will be solved by discussion and submission to the third author (WHW) if necessary. After this initial phase, the full texts of all remaining articles will be independently reviewed by the 2 authors (GJ, GXT) for inclusion or exclusion in the final study. Conflicts will be resolved in the same way as when they were initially screened.

Two authors (YP, PW) will extract the study characteristics from each included study, and the extraction of data from the selected articles will be independently performed by 2 investigators. Disagreements will be resolved through discussion and consensus or by consulting a third member (YY) of the review team. The data to be collected will include the following: author, year of publication, number needed to treat, race, model, algorithm, image source, diagnostic gold standard, sensitivity (SEN), specificity (SPE), true positive (TP) rate, false positive (FP) rate, false negative (FN) rate, true negative (TN) rate, nodule segmentation, feature extraction and selection, nodule classification, and AI for diagnosing thoracic disease.

### 2.4. Quality evaluation

Two reviewers (HYZ, ZQW) will appraise all the included studies by using a checklist based on the Quality Assessment of

Diagnostic Accuracy Studies 2 (QUADAS-2) guidelines. The QUADAS-2 includes 4 parts regarding patient selection, index test, reference standard, and flow and timing risk of bias. The risk of bias is classified as "low," "high," or "unclear."[49]

### 2.5. Data analysis

We will first extract the 2 × 2 contingency table (TP, FP, TN, FN). Some primary studies do not directly provide all the data in a 2 × 2 table. We will calculate the missing data based on the existing data in the paper, such as SEN, SPE, and the number of cases, through the calculator in Review Manager 5.3. For the meta-analysis, the pooled SEN and SPE with 95% confidence intervals for diagnosis by AI will be calculated. The descriptive forest plot and summary receiver operating characteristic (SROC) curves will be derived by Meta-disc 1.4. We will plot the 95% CI and prediction regions around the averaged accuracy estimates in the SROC space, and the AUC will be calculated. SROC curves are defined by sensitivity ($y$-axis) and specificity ($x$-axis), and AUC is the final comparison indicator. The criteria for AUC classification are 0.90 to 1 (excellence), 0.80 to 0.90 (good), 0.70 to 0.80 (fair), 0.60 to 0.70 (poor), and 0.50 to 0.60 (failure).[50]

### 2.6. Assessment of heterogeneity

Initially, we will visually inspect the forest plots of each study's sensitivities and specificities as well as SROC curves related to the individual study results to examine heterogeneity. Statistical heterogeneity will be evaluated informally from the forest plots of the study estimates and more formally using the $I^2$ statistic ($I^2 > 50\% =$ significant heterogeneity). In addition, different diagnostic thresholds of the included studies may lead to heterogeneity; we will use the Spearman correlation coefficients to test whether there is a threshold effect. When there is a threshold effect, sensitivity and specificity will be negatively correlated, and the results will present a "shoulder-arm" point distribution on the SROC curve.

### 2.7. Publication policy

Publication bias will be explored with Begg and Egger funnel plots.

## 3. Discussion

In recent years, with the emergence of deep learning technology, many AI models have been used for the diagnosis of thoracic diseases, but the accuracy of these models varies greatly. Although there is evidence that AI can improve the accuracy of thoracic cancer diagnosis, the evidence is limited, and there has been no systematic evaluation. We will use appropriate methods and quality assessment tools to systematically evaluate the models used to diagnose thoracic diseases and provide evidence for clinical practice. When establishing a diagnostic strategy that can accurately reflect the current evidence, greater scientific rigor is needed. We will also conduct subgroup analysis based on the model used, the type of classifier, and radiomics information.

One major limitation is that the proportion of malignant tumors in the original study sample is much higher than that in the normal population, which may lead to overfitting and more optimistic results. In addition, the original studies are mainly retrospective, and the quality of the original studies will affect the quality of the systematic evaluation.

## Author contributions

Conceptualization: Zijiang Zhu.
Investigation: Yi Yang, Gang Jin, Yao Pang, Hongyi Zhang.
Methodology: Yi Yang, Gang Jin.
Project administration: Yi Yang, Gang Jin, Wenhao Wang, Guangxin Tuo, Peng Wu, Zequan Wang.
Supervision: Zijiang Zhu, Yi Yang.
Writing – original draft: Yi Yang, Gang Jin.
Writing – review & editing: Yi Yang, Gang Jin, Yao Pang, Hongyi Zhang.
Yi Yang orcid: 0000-0001-5801-0170.

## References

[1] Miller KD, Goding Sauer A, Ortiz AP, et al. Cancer Statistics for Hispanics/Latinos, 2018. CA Cancer J Clin 2018;68:425–45.
[2] Aberle DR, Adams AM, Berg CD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med 2011;365:395–409.
[3] Reck M, Heigener DF, Mok T, et al. Management of non-small-cell lung cancer: recent developments. Lancet 2013;382:709–19.
[4] Bi WL, Hosny A, Schabath MB, et al. Artificial intelligence in cancer imaging: clinical challenges and applications. CA Cancer J Clin 2019;69:127–57.
[5] Wang Z, Xin J, Sun P, et al. Improved lung nodule diagnosis accuracy using lung CT images with uncertain class. Comput Methods Programs Biomed 2018;162:197–209.
[6] Aberle DR, Abtin F, Brown K. Computed tomography screening for lung cancer: has it finally arrived? Implications of the national lung screening trial. J Clin Oncol 2013;31:1002–8.
[7] McKee BJ, Regis S, Borondy-Kitts AK, et al. NCCN guidelines as a model of extended criteria for lung cancer screening. J Natl Compr Canc Netw 2018;16:444–9.
[8] Bridge JA, Greenhouse JB, Ruch D, et al. Association between the Release of Netflix's 13 reasons why and suicide rates in the United States: an interrupted time series analysis. J Am Acad Child Adolesc Psychiatry 2019;S0890-8567(19)30288-6.
[9] Tang A, Tam R, Cadrin-Chenevert A, et al. Canadian Association of radiologists white paper on artificial intelligence in radiology. Can Assoc Radiol J 2018;69:120–35.
[10] Shen H. Low-dose CT for lung cancer screening: opportunities and challenges. Front Med 2018;12:116–21.
[11] Wilson R, Devaraj A. Radiomics of pulmonary nodules and lung cancer. Transl Lung Cancer Res 2017;6:86–91.
[12] Li Q, Kim J, Balagurunathan Y, et al. CT imaging features associated with recurrence in non-small cell lung cancer patients after stereotactic body radiotherapy. Radiat Oncol 2017;12:158.
[13] Yip SSF, Liu Y, Parmar C, et al. Associations between radiologist-defined semantic and automatically computed radiomic features in non-small cell lung cancer. Sci Rep 2017;7:3519.
[14] Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol 2017;14:749–62.
[15] Yip SS, Aerts HJ. Applications and limitations of radiomics. Phys Med Biol 2016;61:R150–66.
[16] Kononenko I. Machine learning for medical diagnosis: history, state of the art and perspective. Artif Intell Med 2001;23:89–109.
[17] Gurcan MN, Sahiner B, Petrick N, et al. Lung nodule detection on thoracic computed tomography images: preliminary evaluation of a computer-aided diagnosis system. Med Phys 2002;29:2552–8.
[18] Doi K. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. Comput Med Imaging Graph 2007;31:198–211.
[19] Bengio Y, LeCun Y, Nohl C, et al. LeRec: a NN/HMM hybrid for on-line handwriting recognition. Neural Comput 1995;7:1289–303.
[20] Wu XD, Kumar V, Quinlan JR, et al. Top 10 algorithms in data mining. Knowled Informat Syst 2008;14:1–37.
[21] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521: 436–44.
[22] Ather S, Kadir T, Gleeson F. Artificial intelligence and radiomics in pulmonary nodule management: current status and future applications. Clin Radiol 2020;75:13–9.

[23] Hamet P, Tremblay J. Artificial intelligence in medicine. Metabolism 2017;69S:S36–40.

[24] Xue P, Qiao YL, Jiang Y. [Application of artificial intelligence in diagnosis of medical endoscope]. Zhonghua Zhong Liu Za Zhi 2018;40:890–3.

[25] Sun D. Development of new diagnostic techniques - machine learning. Adv Exp Med Biol 2017;1010:203–15.

[26] Deo RC. Machine learning in medicine. Circulation 2015;132:1920–30.

[27] Zhu X, Dong D, Chen Z, et al. Radiomic signature as a diagnostic factor for histologic subtype classification of non-small cell lung cancer. Eur Radiol 2018;28:2772–8.

[28] Coudray N, Ocampo PS, Sakellaropoulos T, et al. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. Nat Med 2018;24:1559–67.

[29] Lee A, Taylor P, Kalpathy-Cramer J, et al. Machine learning has arrived!. Ophthalmology 2017;124:1726–8.

[30] Erickson BJ, Korfiatis P, Akkus Z, et al. Machine learning for medical imaging. Radiographics 2017;37:505–15.

[31] Giger ML. Machine learning in medical imaging. J Am Coll Radiol 2018;15(3 pt B):512–20.

[32] Eppenhof KAJ, Pluim JPW. Pulmonary CT registration through supervised learning with convolutional neural networks. IEEE Trans Med Imaging 2019;38:1097–105.

[33] Wang S, Li D, Petrick N, et al. Optimizing area under the ROC curve using semi-supervised learning. Pattern Recognit 2015;48:276–87.

[34] Bashir U, Kawa B, Siddique M, et al. Non-invasive classification of non-small cell lung cancer: a comparison between random forest models utilising radiomic and semantic features. Br J Radiol 2019;92:20190159.

[35] Bajre MK, Pennington M, Woznitza N, et al. Expanding the role of radiographers in reporting suspected lung cancer: a cost-effectiveness analysis using a decision tree model. Radiography (Lond) 2017;23:273–8.

[36] Meng XL. The EM algorithm and medical studies: a historical link. Stat Methods Med Res 1997;6:3–23.

[37] Shi Y, Lai R, Kern K, et al. Harmonic surface mapping with Laplace-Beltrami eigenmaps. Med Image Comput Comput Assist Interv 2008;11 (pt 2):147–54.

[38] Choi W, Oh JH, Riyahi S, et al. Radiomics analysis of pulmonary nodules in low-dose CT for early detection of lung cancer. Med Phys 2018;45:1537–49.

[39] 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.

[40] Price CA. The CNN vs. CNN debate continues. Nephrol Nurs J 2000;27:456.

[41] Pouladi F, Salehinejad H, Gilani AM, Ieee. Recurrent neural networks for sequential phenotype prediction in genomics. In: Jumelly DA, Hussain A, Tawfik H, MacDermott A, Lempereur B, eds. Proceedings 2015 International Conference on Developments in Esystems Engineering Dese 2015: 225–230.

[42] Poernomo A, Kang DK. Biased dropout and crossmap dropout: learning towards effective dropout regularization in convolutional neural network. Neural Netw 2018;104:60–7.

[43] Biswas M, Kuppili V, Saba L, et al. State-of-the-art review on deep learning in medical imaging. Front Biosci (Landmark Ed) 2019;24:392–426.

[44] Hinton GE, Osindero S, Teh YW. A fast learning algorithm for deep belief nets. Neural Comput 2006;18:1527–54.

[45] Jin KH, McCann MT, Froustey E, et al. Deep convolutional neural network for inverse problems in imaging. IEEE Trans Imaging Process 2017;26:4509–22.

[46] Paoletti ME, Haut JM, Plaza J, et al. A new deep convolutional neural network for fast hyperspectral image classification. ISPRS J Photogram Remote Sens 2018;145:120–47.

[47] Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. Radiology 2016;278:563–77.

[48] Kumar V, Gu Y, Basu S, et al. Radiomics: the process and the challenges. Magn Reson Imaging 2012;30:1234–48.

[49] Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155:529–36.

[50] Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. J Clin Epidemiol 2005;58:882–93.