



Preliminary study on artificial intelligence diagnosis of pulmonary embolism based on computer in-depth study

Xiang Li¹, Xiang Wang¹, Xin Yang², Yi Lin², Zengfa Huang¹

¹The Central Hospital of Wuhan, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China; ²Huazhong University of Science and Technology, College of Automation and Artificial Intelligence, Wuhan, China

Contributions: (I) Conception and design: X Li, X Yang; (II) Administrative support: X Wang; (III) Provision of study materials or patients: X Li; (IV) Collection and assembly of data: Z Huang; (V) Data analysis and interpretation: X Li, Y Lin; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Xiang Li; Xiang Wang. Wuhan Central Hospital Affiliated to Medical College of Huazhong University of Science and Technology, No.26 Shengli Street, Jiang'an District, Wuhan 430014, China. Email: 11858667@qq.com; 1136611484@qq.com.

Background: Objective to preliminarily verify the feasibility of AI intelligent diagnosis of pulmonary embolism by using a new artificial intelligence (AI) computer-aided diagnosis system (CAD) to localize and quantitatively diagnose pulmonary embolism in pulmonary artery CT angiography (CTA).

Methods: Computed tomography angiography (CTA) data of 85 patients with PE in our hospital from January 2017 to May 2018 were retrospectively collected and randomly allocated to 2 groups: computer depth learning group (n=43) and experimental group (n=42). For the training set (13,144 sheets) and the test set (313 sheets), the auxiliary diagnosis method was obtained and applied to the experimental group.

Results: Among the participants, a good sensitivity of 90.9% and an average false positive of 2.0 were obtained by using the deep learning detection method proposed in this paper, and the detection rate was positively correlated with arterial grade.

Conclusions: The computer-aided diagnostic method proposed in this paper can effectively improve the detection rate of PE, especially for the detection of intra-arterial embolism above grade 3. However, because of the high misdetection rate, more in-depth learning datasets are needed for the detection of embolism below grade 3.

Keywords: Computer-aided diagnosis; deep learning; artificial intelligence; pulmonary embolism (PE)

Submitted Jan 06, 2021. Accepted for publication May 19, 2021.

doi: 10.21037/atm-21-975

View this article at: <http://dx.doi.org/10.21037/atm-21-975>

Introduction

Pulmonary embolism (PE) is one of the 3 most common cardiovascular diseases. It has the characteristics of high morbidity, low clinical diagnosis rate, and high mortality rate. It has become an international medical care problem. The annual incidence of PE in the United States has reached 600,000, and PE accounted for 50,000–200,000 of all deaths. The incidence of PE in cardiovascular diseases is second only to coronary atherosclerotic heart disease and hypertension. In recent years, with the development of the aging population, living standards, and diagnostic techniques, the incidence of PE is on the rise. The clinical

presentation of PE is not specific and is easily ignored, which brings difficulties to clinical diagnosis. To date, computed tomography and angiography (CTA) have become effective and predominant methods for clinical diagnosis of PE. However, CTA detection often involves more than 300 images, and radiologists must visually track to the subbranch of the sixth stage of the pulmonary artery or even smaller to search for bolts. The complexity of the image and its interference factors, such as respiratory movement artifacts (1), flow artifacts, partial volumetric effects, lymph nodes, blood vessel bifurcation, and other false positives (FPs) have seriously interfered with accurate

diagnosis. Therefore, manual interpretation is a very time-consuming, laborious, and complex matter, requiring doctors to have a wealth of clinical experience. Based on the above problems, we performed an automatic analysis of CTA data of the pulmonary artery through a new computer-aided diagnostic method (2), and then studied the huge neural network data in order to obtain a complete detection system. This enabled the diagnostic time of PE by imaging doctors to be greatly shortened, and more accurate diagnosis and treatment information to be provided. However, at present, artificial intelligence has a high false positive rate in the study of pulmonary embolism because of the diversity of embolus morphology and individual differences in vascular conditions. We present the following article in accordance with the STARD reporting checklist (available at <http://dx.doi.org/10.21037/atm-21-975>).

Methods

General information

Using CTA image collection of the lungs from November 2016 to May 2017, 85 cases of PE were diagnosed at the Wuhan Central Hospital attached to Tongji Medical College of Huazhong University of Science and Technology. Among them, 38 were males and 47 were females, with an average age of 52.5 years (27–78 years). The participants were randomly divided into a computer in-depth learning group (n=43) and an experimental group (n=42).

All participants were permitted to consume a small amount of liquid diet before examination, liver and kidney function was normal, allergy test was negative, they had received no other special treatment, and were imaged at our hospital with Philips 256-layer spiral CT machine (Philips, Best, Netherlands) and Ante double tube high-pressure syringe (Shenzhen Antegoko Industrial Co., China). The contrast agent used was iodoethylene alcohol (370 mg/mL), injected at a flow rate of 4.0–4.5 mL per second to a total of 60–70 mL, followed by 30 mL normal saline water injected at the same flow rate using a Surestart intelligent trigger (trigger threshold was 90–100 Hu). A region of interest (ROI) was placed in the trunk of the pulmonary artery and a pulmonary arterial phase scan ranging from the entire lung field to the lower edge of the bilateral diaphragm was performed. The scan parameters were as follows: detector 0.5×128 layer, tube voltage 120 kV, tube current 300–350 mAs, scan time 6.5–8.0 S, field of view 250 mm × 250 mm, collection matrix 512×512, layer thickness

0.5–1.0 mm.

The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013) and approved by Ethics Committee of Wuhan Central Hospital (No. ZRMS2019000122) and written informed consent was obtained from all patients.

Study methods

CT workstation post-processing

The obtained CT data were processed through Siemens' own workstation, uploaded in Digital Imaging and Communications in Medicine (DICOM) format, and 3-dimensional (3D) virtual reality (VR) images were obtained through surface imaging technology. Then, vascular probe technology was used to explore the pulmonary artery trunk, left and right pulmonary arteries, and branches at all levels, respectively, to obtain various blood vessels, multiplanar reformation (MPR) and maximum intensity projection (MIP) images. The results of each vessel embolization were recorded.

Computer preprocessing of CTA images of the lungs

Lung area segmentation, according to the principle of CT imaging and the congenital good contrast of the lungs (–1,024, –400) HU, and the CT value of each organization was usually greater than –100 HU, so the threshold was set to –100 HU. The CT images could be binarized in 3 dimensions, followed by 3D continuous domain analysis to obtain the corresponding lung area and extract the ROI area. The project computer research team passed the large rate method. In order to determine the optimal segmentation area, the results of bivalent 3D lung parenchyma division were obtained, but the trachea and branches remained (3), and the left and right lungs had adhesion.

The CT-value range of the trachea and bronchus is similar to that of lungs, and it was difficult to separate through threshold segmentation. The computer research team designed a 3D automatic region growth algorithm to achieve trachea and bronchus segmentation by automatically selecting initial seed points.

This project utilized a watershed segmentation algorithm based on markup control. By specifying the markup point information of each segment in advance (the pixels of these markers were known in the area where they are located), the auxiliary watershed algorithm was used to separate the left and right lungs.

Automatic division of lung blood vessels: enhanced

training data based on generation versus learning. In recent years, the use of Generational Advanced Network (GAN) to capture the distribution of sample data and generate corresponding data samples based on this distribution model has attracted the attention of many researchers (4). In the field of medical image analysis, GAN network is used to simulate the data distribution of sparse sample space, thus producing more abundant training samples, which is a very promising solution to solve the shortage of medical image data and diversity. The noise signal distribution of actual pulmonary blood vessel samples was analyzed. By calculating and analyzing the reflection of real samples to noise (real samples \rightarrow noise), we can understand the correlation between the distribution of lung blood vessels and noise signals, and obtain noise input that is more in line with clinical significance.

By combining the segment loss function and the pixel-wise loss function, the importance of “thick blood vessels” and “fine blood vessels” in the loss calculation can be balanced during the training process, so that the model can better learn accurate vascular segmentation characteristics. Instead of directly representing all the pixels in the search range as real points, bone similarity was defined to measure the structural similarity and thickness consistency between reference vascular segmentation and source vascular segmentation. On the basis of the bone similarity measure, Se, Sp, and ACC were redefined as based on the characteristics of pixel matching:

$$Se = \frac{TP}{TP + FN}, Sp = \frac{TN}{TN + FP}, Acc = \frac{TP + TN}{N} \quad [1]$$

Among them, TP refers to the correctly detected pixels, FN is an error classified as a non-vascular pixel for vascular pixels, TN refers to a correctly classified non-vascular pixel, and FP refers to a non-vascular pixel that is incorrectly classified as a vascular pixel.

Arteriovenous vascular separation. The anatomy, morphology, and topology information of the pulmonary structure were combined to separate the pulmonary arteriovenous. Because the pulmonary artery is parallel to the bronchial artery, the Db value of the pulmonary artery branch (i.e., the distance from the bronchial region to the vascular segment) is smaller than the vein. The DVD value of the pulmonary vein branch (distance between the leaves closest to the blood vessel) is relatively small because it is close to the leaves of the lung segment. Also, because the pulmonary artery is relatively bright compared to the vein, the central light reflection phenomenon is obvious, so the

separation of arteriovenous vessels is achieved based on the different characteristics of the arteriovenous vein in the lung image.

Statistical analysis

To evaluate the validity of the test, we used two observation indexes: computational sensitivity and false positive rate (FP).

Sensitivity: in the gold standard to determine the sick (positive) population, the probability of AI detection positive, true positive (The ability to detect disease).

False positive rate: the positive result is obtained, but the positive result is false, that is, the probability of positive detection in people who are judged to be disease-free (negative) in manual interpretation (no disease, but the test results said that there was a disease), as the misdiagnosis rate.

This study used sensitivity and FP rate to evaluate the performance of the test. *Sensitivity* and *false positive rate* are expressed as follows:

$$Sensitivity = TP / P \quad [2]$$

$$False\ Positive\ per\ scan = TP / num_scan \quad [3]$$

It indicates that the number of positive results in all predictions, the number of negative results in all predictions, and the number of CT images tested. The higher the sensitivity and the lower the corresponding FP rate, the better the performance of the detection system.

Results

We divided 43 study groups of CTA 512×512 pictures into training sets (1,992 sheets) and test sets (192 sheets). The 101×101 embolism distribution block diagram was divided into a training set (13,144 sheets) and a test set (313 sheets), as shown in *Figure 1A,B*. Through the deep learning model of data learning and analysis, the final training was to obtain a prediction model. The test set was then tested. The results of the experiment are shown in *Table 1*.

When the prediction model analyzes the input test image, it will output a value between 0 and 1, which represents the probability that the test image contains pulmonary embolism segments. By setting a threshold, we can distinguish whether the test image contains pulmonary embolism or not. At the same threshold, the ROC curves of sensitivity and false positive are shown in *Figure 2*. The overall sensitivity was 0.990 (99.0%). When the average number of false positives per case was 2, the sensitivity was 0.823.

This computer-aided diagnostic system was applied to 40 cases in the experimental group. The results are shown in

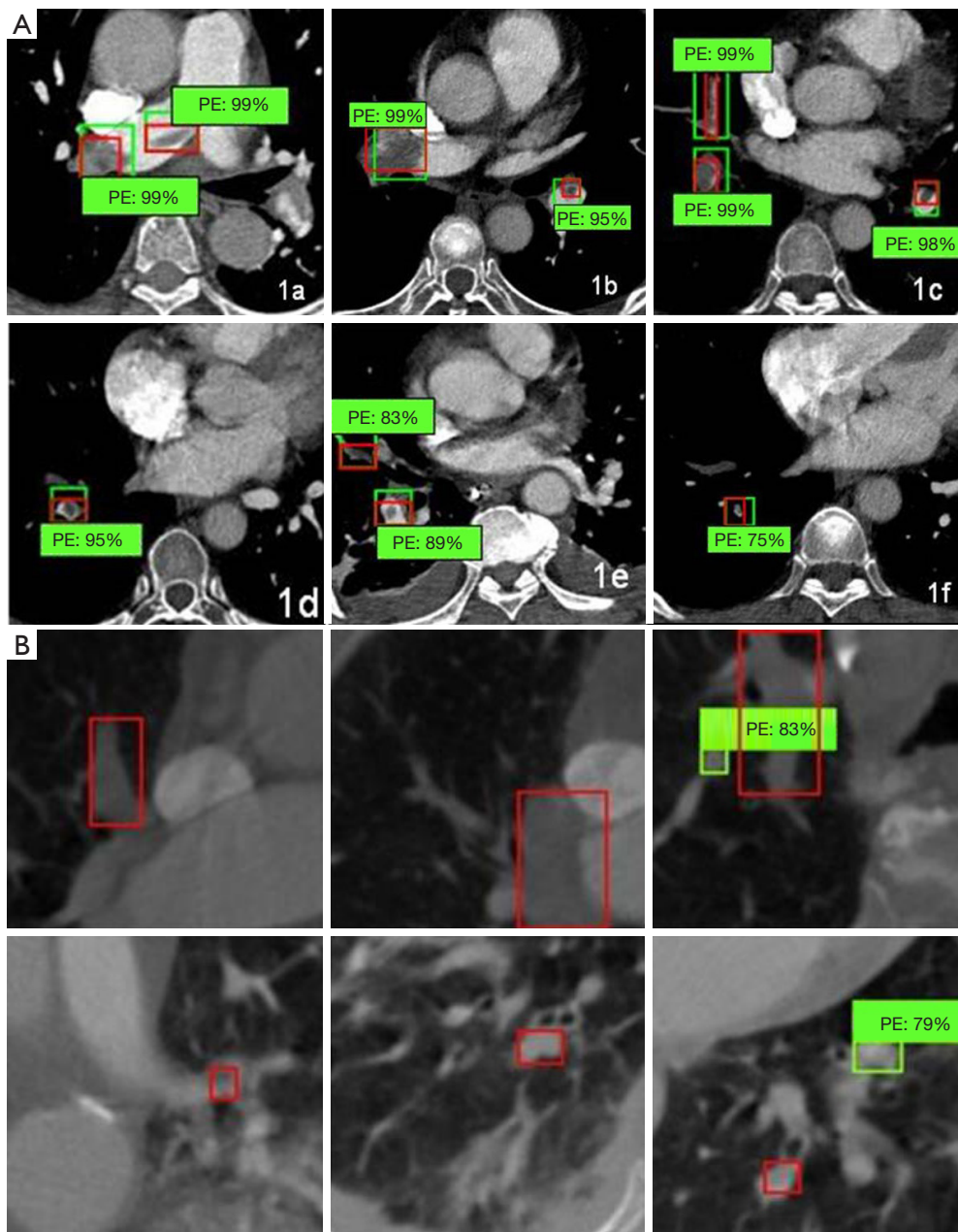


Figure 1 Sensitivity evaluation of PE test set in computer in-depth learning group. (A): 1a–1f: different branch emboli have different AI recognition accuracy. 1a–1c shows that the recognition accuracy of grade 1–2 branches is as high as 99%, while that of grade 3 branches is decreased, but there is still 98%; 1e shows the accuracy rate of the emboli in the four branches was about 83–89%; 1f shows the accuracy of 1f five level branch is only 75%. (B) AI is prone to errors, such as artifacts caused by poor breath holding, heart beating artifacts, poor filling of contrast media, infection and contamination of venous filling. PE, pulmonary embolism.

Table 1 Computer learning group test results (graded by PE affected artery)

Classification of embolization arterial	Total embolization	Number of missed or mistaken inspections	Number of checkups	Susceptibility
Pulmonary artery trunk	13	0	13	100%
Left/right pulmonary artery	29	1	29	99%
Pulmonary artery secondary branch	147	13	140	95%
Pulmonary artery tertiary branch	78	25	69	88%
Pulmonary artery level 4 branch	35	21	28	80%
Pulmonary artery branch 5 and below	11	17	8	75%
Total	313	77	287	92%

PE, pulmonary embolism.

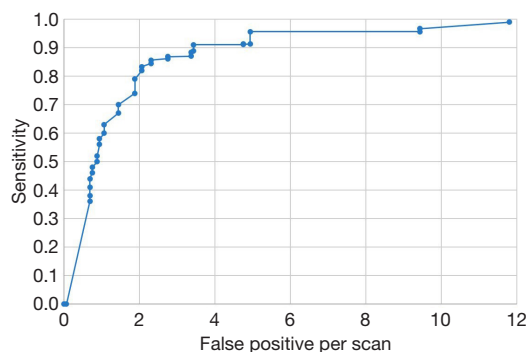


Figure 2 ROC curve of sensitivity and mean FP. ROC, receiver operating characteristic; FP, false positive.

Table 2. The detection rate of branches at all levels obtained by this method was similar to that of the test group, and the FP rate was also increased as the arterial classification became finer. The number of errors increased significantly, sensitivity reduced, and detection sensitivity was positively correlated with arterial classification.

Discussion

Computer-aided detection of PE is a challenging area in computer vision. The detection of PE is more challenging than that of pulmonary nodules due to the large network structure and size changes of pulmonary arteries. In 2002, the earliest computer aided method developed by Masutani for PE detection based on volumetric image analysis. Das *et al.* (5). They performance of the PE detection system scanned by CT arterial imaging (CTPA) was evaluated using data from 186 segment PE and 120 segment PE from 33 cases. The sensitivity of the system to segmented

PEI and sub-segmented PE is 88% and 78%, respectively. Digumarthy *et al.* (6) conducted a similar study using CTPA to examine 36 patients to evaluate the same system. The 2 radiologists read the film together, with a third for resolving disagreements: 23 cases contained 130 segmented PEs, 107 sub-segmented PEIs, and the remaining 13 cases were negative. The system detected 92% of segmented PE and 90% of sub-segmented PE, with a FP rate of 4.8 per case. In a study by Zhou *et al.* (7,8), an automatic pulmonary vascular division and PE detection system was developed for CTA images, and some studies were conducted to evaluate the performance of the system. In the preliminary study, they used 14 case data sets, of which 8 had large-scale pulmonary parenchyma or pleural diseases. A total of 163 embolisms were identified, of which 94 were adjacent to the subsegment artery and 64 were located in the subsegment artery. The results showed that the visibility of PE was greater than 2, and the proportion of blocked blood vessels was 20–80%. The CAD system detected an average of 64% of sub-segmented PE and 84% of PE, and the average pseudopositive rate was 14.4 per case (9). Bouma *et al.* studied and simulated the automatic detection of PE in CTA images (10). Data sets for 32 positive cases were used as training sets, of which 202 were marked and 19 positive cases were used for evaluation, of which 116 were marked. These patient data were selected to represent the clinical situation, including moving artifacts, lung lesions, different embolism levels, and vascular enhancement. The results (11) showed that the sensitivity of PE was increased by 22% in clinical radiologists (12,13). Although the application of the above studies to PE has developed greatly, it has not led to a more effective method of division of pulmonary arteries. However, the distribution of blood vessels in the

Table 2 Experimental group results (graded by PE-affected artery)

Classification of embolization arterial	Total embolization	Number of missed or mistaken inspections	Number of checkups	Susceptibility
Pulmonary artery trunk	2	0	2	100%
Left/right pulmonary artery	3	1	3	99.9%
Pulmonary artery secondary branch	30	7	28	93.3%
Pulmonary artery tertiary branch	15	9	11	73%
Pulmonary artery level 4 branch	9	8	6	66.7%
Pulmonary artery branch 5 and below	2	5	1	50%
Total	61	24	51	83.6%

PE, pulmonary embolism.

CT image of the lungs is key to the diagnosis of diseases and the development of surgical plans (14). The correct division of lung blood vessels is of great significance to provide reference for radiologists, thus greatly shortening the diagnosis time, improving the accuracy of diagnosis, and reducing the FP rate. There is a very high social and market demand for such improvements, which is a key scientific and technological problem that needs urgent resolution. With the development of artificial intelligence and deep learning theory, deep learning has made breakthrough progress in the field of image classification and target detection, and has been verified in many applications in the field of computer vision.

In this study, by means of computer in-depth learning, a certain amount of embolism data was applied to pre-study and then tested in the test group. The results of detection of branches above level 3 were still quite satisfactory, but the errors and omissions of branches below level 3 were significantly improved, which affected the overall detection rate. Further research is required involving a larger sample size and learning of more embolism images. However, the artificial intelligence extraction of the branches below three levels is a higher challenge to the artificial interpretation, diagnosis and fine labeling of the previous samples. In summary, in-depth study of methods to improve the accuracy of early detection is of great significance to further optimize social medical resources, achieve large-scale early screening, and improve people's living standards. In combination with practical clinical problems, the theoretical methods and clinical applications of in-depth learning will be innovated. Applying the powerful understanding ability of deep convolution neural networks to the detection of PE will certainly improve performance and provide practical

and reliable supporting information for clinical diagnosis. Based on the analysis of research status and development in China and globally, we found that the combination of artificial intelligence, deep learning, and medical imaging assisted diagnosis is the inevitable trend of scientific research.

Acknowledgments

Funding: General project of science and Technology Department of Hubei Province in 2019, fund No. 2019CFB652 “artificial intelligence diagnosis of pulmonary embolism based on CT angiography”.

Footnote

Reporting Checklist: The authors have completed the STARD reporting checklist. Available at <http://dx.doi.org/10.21037/atm-21-975>

Data Sharing Statement: Available at <http://dx.doi.org/10.21037/atm-21-975>

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <http://dx.doi.org/10.21037/atm-21-975>). 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 approved by Ethics Committee of Wuhan Central Hospital

(No. ZRMS2019000122) and written informed consent was obtained from all patients. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

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

- Schoepf UJ, Costello P. CT angiography for diagnosis of pulmonary embolism: State of the Art1. *Radiology* 2004;230:329-37.
- Ko JP, Naidich DP. Computer-aided diagnosis and the evaluation of lung disease. *J Thorac Imaging* 2004;19:136-55.
- Chan HP, Hadjiiski L, Zhou C, et al. Computer-aided diagnosis of lung cancer and pulmonary embolism in computed tomography—A review. *Acad Radiol* 2008;15:535-55.
- Masutani Y, MacMahon H, Doi K. Computerized detection of pulmonary embolism in spiral CT angiography based on volumetric image analysis. *IEEE Trans Med Imaging* 2002;21:1517-23.
- Das M., Schneider AC., Schoepf UO, et al. RSNA 2003. Chicago, IL: 2003. Computer-aided diagnosis of peripheral pulmonary emboli. Chicago, November 30-December 5, 2003RSNA Program Book. 2003;351-2.
- Digumarthy S, Kagay C, Legasto A, et al. RSNA. Chicago, IL: 2006. Computer-aided detection (CAD) of acute pulmonary emboli: Evaluation in patients without significant pulmonary disease. *RSNA Program Book* 2006;255-61.
- Zhou C, Chan H, Patel S, et al. Preliminary Investigation of Computer-aided Detection of Pulmonary Embolism in Three-dimensional Computed Tomography Pulmonary Angiography Images. *Acad Radiol* 2005;12:782-92.
- Zhou C, Chan H, Sahiner B, et al. Automatic multi-scale enhancement and segmentation of pulmonary vessels in CT pulmonary angiography images for CAD applications. *Med Phys* 2007;34:4567-77.
- Jeudy J, Flukinger T, White C. RSNA. Chicago, IL: 2006. Evaluation of pulmonary embolism using an automated computer-aided detection tool. *RSNA Program Book* 2006;255-60.
- Schoepf UJ, Schneider AC, Das M, et al. Pulmonary embolism: Computer-aided detection at multi-detector row spiral computed tomography. *J Thorac Imaging* 2007;22:319-23.
- Maizlin ZV, Vos PM, Godoy MB, et al. Computer-aided detection of pulmonary embolism on CT angiography: Initial experience. *J Thorac Imaging* 2007;22:324-9.
- Das M, Salganicoff M, Bakai A, et al. Computer-aided detection of pulmonary embolism: Assessment of sensitivity with regard to vessel segments. *RSNA Program Book* 2006;487-92.
- Das M, Muhlenbruch G, Helm A, et al. Computer-aided detection of pulmonary embolism: influence on radiologists' detection performance with respect to vessel segments. *Eur Radiol* 2008;18:1350-5.
- Buhmann S, Herzog P, Liang J, et al. Clinical evaluation of a computer-aided diagnosis (CAD) prototype for the detection of pulmonary embolism. *Acad Radiol* 2007;14:651-8.

(English Language Editor: J. Jones)

Cite this article as: Li X, Wang X, Yang X, Lin Y, Huang Z. Preliminary study on artificial intelligence diagnosis of pulmonary embolism based on computer in-depth study. *Ann Transl Med* 2021;9(10):838. doi: 10.21037/atm-21-975