



Clot burden of acute pulmonary thromboembolism: comparison of two deep learning algorithms, Qanadli score, and Mastora score

Hongxia Zhang¹, Yan Cheng², Zhenbo Chen¹, Xinying Cong¹, Han Kang³, Rongguo Zhang³, Xiaojuan Guo⁴, Min Liu⁵

¹Department of Radiology, China Rehabilitation Research Center, Beijing Bo'ai Hospital, Capital Medical University School of Rehabilitation Medicine, Beijing, China; ²Intensive Care Unit, Erlonglu Hospital of Beijing, Beijing, China; ³Institute of AI-Advanced Research, Infervision Medical Technology Co., Ltd., Beijing, China; ⁴Department of Radiology, Beijing Chaoyang Hospital of Capital Medical University, Beijing, China; ⁵Department of Radiology, China-Japan Friendship Hospital, Beijing, China

Contributions: (I) Conception and design: M Liu, R Zhang; (II) Administrative support: M Liu; (III) Provision of study materials or patients: H Zhang, Y Cheng, Z Chen, X Cong, X Guo, M Liu; (IV) Collection and assembly of data: H Zhang, Y Cheng, X Guo; (V) Data analysis and interpretation: H Zhang, Y Cheng, Z Chen, H Kang, R Zhang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Min Liu, MD. Department of Radiology, China-Japan Friendship Hospital, No. 2 Yinghua Dong Street, Hepingli, Chao Yang District, Beijing 100029, China. Email: drradiology@163.com.

Background: The deep learning convolution neural network (DL-CNN) benefits evaluating clot burden of acute pulmonary thromboembolism (APE). Our objective was to compare the performance of the deep learning convolution neural network trained by the fine-tuning [DL-CNN (ft)] and the deep learning convolution neural network trained from the scratch [DL-CNN (fs)] in the quantitative assessment of APE.

Methods: We included the data of 680 cases for training DL-CNN by DL-CNN (ft) and DL-CNN (fs), then retrospectively included 410 patients (137 patients with APE, 203 males, mean age 60.3±11.4 years) for testing the models. The distribution and volume of clots were respectively assessed by DL-CNN(ft) and DL-CNN(fs), and sensitivity, specificity, and area under the curve (AUC) were used to evaluate their performances in detecting clots on a per-patient and clot level. Radiologists evaluated the distribution of clots, Qanadli score, and Mastora score and right ventricular metrics, and the correlation of clot volumes with right ventricular metrics were analyzed with Spearman correlation analysis.

Results: On a per-patient level, the two DL-CNN models had high sensitivities and moderate specificities (DL-CNN (ft): 100% and 77.29%; DL-CNN (fs): 100% and 75.82%), and their AUCs were comparable ($Z=0.30$, $P=0.38$). On a clot level, DL-CNN (ft) and DL-CNN (fs) sensitivities and specificities in detecting central clots were 99.06% and 72.61%, and 100% and 70.63%, respectively. DL-CNN (ft) sensitivities and specificities in detecting peripheral clots were mostly higher than those of DL-CNN (fs), and their AUCs were comparable. Clot volumes measured with the two models were similar ($U=85094.500$, $P=0.741$), and significantly correlated with Qanadli scores [DL-CNN(ft) $r=0.825$, $P<0.001$, DL-CNN(fs) $r=0.827$, $P<0.001$] and Mastora scores [DL-CNN(ft) $r=0.859$, $P<0.001$, DL-CNN(fs) $r=0.864$, $P<0.001$]. Clot volumes were also correlated with right ventricular metrics. Clot burdens were increased in the low-risk, moderate-risk, and high-risk patients. Binary logistic regression revealed that only the ratio of right ventricular area/left ventricular area (RVa/LVa) was an independent predictor of in-hospital death (odds ratio 6.73; 95% CI, 2.7–18.12, $P<0.001$).

Conclusions: Both DL-CNN (ft) and DL-CNN (fs) have high sensitivities and moderate specificities in detecting clots associated with APE, and their performances are comparable. While clot burdens quantitatively calculated by the two DL-CNN models are correlated with right ventricular function and risk stratification, RVa/LVa is an independent prognostic factor of in-hospital death in patients with APE.

Keywords: Deep learning (DL); acute pulmonary embolism (APE); clot burden; computed tomographic pulmonary angiography (CTPA)

Submitted Feb 03, 2021. Accepted for publication Jun 11, 2021.

doi: 10.21037/qims-21-140

View this article at: <https://dx.doi.org/10.21037/qims-21-140>

Introduction

Pulmonary arterial occlusion by fresh emboli can sharply increase pulmonary vascular resistance leading to acute right heart failure and even sudden death in acute pulmonary embolism (APE). Clot burden was an important factor in hemodynamics and right ventricular size (1) and is of great significance for risk evaluation and treatment in APE (1,2). Both the Qanadli score (3) and Mastora score (4) measured on computed tomographic pulmonary angiography (CTPA) have been used to semi-quantitatively assess clot burden, and some studies have indicated that these scores correlate with risk stratification and prognosis in APE (5-7).

The deep learning (DL) algorithm (8-10) trains deep neural networks used in processing large and complex images. It is usually composed of multi-layer simple neural networks with nonlinear input-output mapping characteristics in which convolutional neural networks (CNN) can extract many features from abstracted layers of filters. Currently, there are two approaches (11,12) to train DL-CNN models: the deep learning convolution neural network trained by the fine-tuning [DL-CNN (ft)] and the deep learning convolution neural network trained from the scratch [DL-CNN (fs)]. DL-CNN (fs) is a conventional method that trains a model from scratch using the weights of random initialization and obtains good performance by using large-scale training datasets, while DL-CNN (ft) is the most widely used approach for transfer learning, which starts training with the weights gained from a pre-trained model. Compared with DL-CNN (fs), DL-CNN (ft) can significantly reduce the target labeled data requirements in the field of natural images. A previous study (13) indicated that DL-CNN (ft) was more robust to the size of training sets than DL-CNN (fs) on several distinct medical imaging applications. In our initial study (14), we trained a DL-CNN (fs) model and developed a fully automatic algorithm, end-to-end fully convolutional network based on DL-CNN, U-Net, to auto-segment clots of APE. However, the performance of DL-CNN (ft) in assessing clot burden in APE remains unknown. Thus, this study aimed to compare DL-CNN (ft) and DL-CNN (fs) performances to detect

the clot burden of APE and analyze the correlation of clot burden with risk stratification and short-term prognosis of APE during hospitalization.

Methods

Study cohort and design

This study was registered at the Chinese Clinical Trials Registry Center (<http://www.chictr.org/en/>; registration number ChiCTR-OCH-14004929) and was approved by our institutional review board (medical ethics number: 2020-070-1). Following the Declaration of Helsinki (as revised in 2013), the study was conducted, and informed consent from patients was waived for this retrospective study. *Figure 1* demonstrates a flowchart detailing how participants were selected, and research was undertaken. First, data of 680 cases (384 males, mean age =64.3±11.2 years, 510 patients with APE) who attended one of our four hospitals between January 2016 and December 2018 were retrospectively collected for the training model. This dataset was randomly split into a training dataset and a validation dataset. The training dataset included 408 cases with APE (233 males, mean age =65.8±15.7 years) and 136 cases without pulmonary embolism (PE) (74 males, mean age =52.6±14.3 years) for training DL-CNN. The validation dataset included 102 cases with APE (57 males, mean age =63.1±7.3 years) and 34 cases without PE (20 males, mean age =60.5±18.2 years) for optimizing model parameters. A total of 410 cases (203 males, mean age =60.3±11.4 years) who attended one of our four hospitals between January 2019 and June 2020 were then retrospectively collected as the testing dataset, which included 137 cases with APE and 273 cases without PE. The diagnosis of APE followed the guidelines of the European Society of Cardiology (ESC) and the European Respiratory Society (ERS) in 2019 (15). Patients (n=83) with incomplete clinical data or unsatisfactory CTPA image quality were excluded, as were those who were diagnosed with chronic pulmonary embolism (CPE) (n=103), chronic thromboembolic pulmonary hypertension (CTEPH) (n=32),

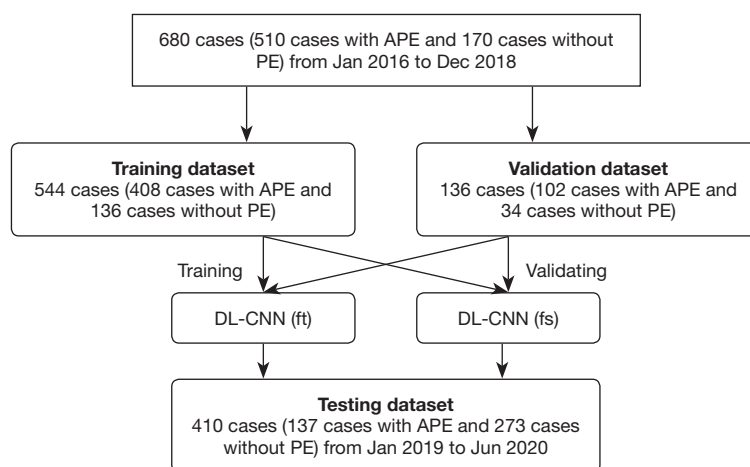


Figure 1 Experiment flow chart. APE, acute pulmonary thromboembolism; PE, pulmonary thromboembolism; DL-CNN (ft), the deep learning convolution neural network trained by fine-tuning; DL-CNN (fs), the deep learning convolution neural network trained from scratch.

pulmonary arterial tumor (n=17), pulmonary vasculitis (n=21), mediastinal fibrosis (n=34), Patients who were lost in follow-up (n=88) and patients with malignancy (n=24) were also excluded. The risk stratification of APE patients was recorded from personal medical charts following the guidelines of the European Society of Cardiology (ESC) and the European Respiratory Society (ERS) in 2019 (15) and was classified according to clinical, imaging, and laboratory indicators. APE patients were divided into low-risk, intermediate-risk, and high-risk groups.

CTPA examination

CTPA was performed in the craniocaudal direction with multidetector CT scanners (Optima CT660, GE Healthcare; Lightspeed VCT/64, GE Healthcare; Toshiba aquilion one tsx-301c/320; Philips ICT/256; Siemens sensation/16, SOMATOM definition dual-source CT) by using a standard CT pulmonary angiography protocol. All images were acquired with the patient in the supine position and breath-holding, and the scanning range was from the thoracic entrance to the supradiaphragmatic level. Scan parameters were as follows: tube voltage 100–120 kV, tube current 100–300 mAs, gantry rotation time 0.8 s, speed of CT table 39.37 mm/s, and a contrast agent (Ultravist, 370 mgI/mL, Schering Bayer) at an injection rate of 4.5 mL/s and an amount of up to 70 mL (bolus tracking technique). A soft tissue reconstruction kernel was used, the reconstructed section thickness was 0.625 mm to 1 mm, and the reconstructed

section interval ranged from 1 to 1.25 mm.

DL-CNN (ft) and DL-CNN (fs)

Similar to our previous study (14), to reduce the diversity of data sources to DL-CNN, all CTPA images were pre-processed by several operations before feeding into the network. These included windowing operation with a window width of 620 and window level of 160, mapping each pixel value into [0, 1], and copying each imaging to form a three-channel image.

We selected 544 cases as a training dataset to train DL-CNN by fine-tuning and from scratch, respectively. Training DL-CNN by fine-tuning was that the weights of the segmentation network were all trained with pre-training, which were initialized using the network parameters of our previous study (14). In training DL-CNN from scratch, the weights of the segmentation network were all trained without pre-training and initialized using the Xavier method (16). The U-Net trained by fine-tuning and from scratch were DL-CNN (ft) and DL-CNN (fs) respectively. In the validation step, the remaining 136 cases were used as a validation dataset for obtaining the optimal network parameters of DL-CNN (ft) and DL-CNN (fs).

Once in use, given a sequence of CTPA slices from a patient, our trained model for clot segmentation would output the probability of each pixel for the foreground or the clot via sigmoid function. Based on our previous findings (14), the probability threshold was set to 0.1.

Following the automatically extracted lung lobes and segments, all segmented clots were then mapped into the central pulmonary artery or peripheral pulmonary artery, and clot volumes were obtained according to automated segmentation results. Based on the calculated total clot volume, it was then possible to identify whether the given patient suffered from APE. For patients who were identified as APE, it could also be determined whether a clot was located in the central pulmonary artery or the peripheral pulmonary artery according to the mapped results of all segmented clots.

Evaluation of APE by radiologists

CTPA images were randomly assigned to two chest radiologists with 15- and 13-year's experience, respectively, who independently located the clots and calculated the computer tomographic pulmonary artery obstruction index (CTPAOI), including the Qanadli and Mastora scores. The clots were located into the central pulmonary artery (main pulmonary artery, left and right pulmonary artery, inter lobar artery, or five lobar arteries) and the peripheral pulmonary artery (18 segmental pulmonary arteries) including the apex segment of right superior lobe (R1), anterior segment of right superior lobe (R2), posterior segment of right superior lobe (R3), medial segment of right middle lobe (R4), lateral segment of right middle lobe (R5), posterior segment of right inferior lobe (R6), medial basal segment of right inferior lobe (R7), anterior basal segment of right inferior lobe (R8), lateral basal segment of right inferior lobe (R9), posterior basal segment of right inferior lobe (R10), apicoposterior segment of left superior lobe (L1.3), anterior segment of left superior lobe (L2), superior lingular segment of left superior lobe (L4), inferior lingular segment of left superior lobe (L5), posterior segment of left inferior lobe (L6), anteromedial basal segment of left inferior lobe (L7.8), lateral basal segment of left inferior lobe (L9), and posterior basal segment of left inferior lobe (L10).

According to the location of clots, APE was divided into four types: Type I: Central APE (clot only located in the central pulmonary artery); Type II: Peripheral APE (clot only located in the peripheral pulmonary artery); Type III: Mixed APE (clot located in the central and peripheral pulmonary artery); Type IV: negative. According to the previously described methods (17,18), right ventricular and arterial metrics on CTPA images (*Figure 2*) were measured independently by one resident with 5-year experience. RVd Right Ventricular transverse diameter (RVd), Left Ventricular transverse diameter (LVd), Right Ventricle Area

(RVa), and Left Ventricle Area (LVa) were measured at the four-chamber level. RVd or LVd is the maximum vertical distance from the free wall of the bilateral ventricle to the interventricular septum, and RVa or LVa is the maximum cross-section of the bilateral ventricle shown on the cross-sectional image, which is calculated automatically by computer after drawing lines along the inner edge of the ventricular wall. Main Pulmonary Artery diameter (MPAd) is the largest diameter on the transverse view, and Ascending Aortic diameter (AAd) was measured at the same level. Right Ventricular Anterior Wall Thickness (RVAWT) is the maximum thickness of the middle part of the free wall, and Interventricular Septal Thickness (IVST) is the maximum thickness of the middle part of the free wall. Superior Vena Cava maximal diameter (SVCD) was measured at the level of the Azygos vein crossing the right main bronchus. RVd/LVd, RVa/LVa, and MPAd/AAd were calculated. Measurement of the Ventricular Septal Angle (SVSA) took place as follows (*Figure 2*): the maximum plane of the left and right ventricles was selected on transverse image obtained at mediastinal window settings [level, 50 Hounsfield Units (HU); width, 350 HU], and the angle of the line between the xiphoid process and spinous process of the vertebral body and interventricular septum was included.

Statistical analysis

Statistical analyses were performed using SPSS (version 26.0, IBM Corp.) and MedCalc statistical software (version 15.6.1, Ostend, Belgium). Quantitative data were expressed as mean \pm SD or median (interquartile interval, IQR), M-W U test was utilized for the nonparametric test of non-normal quantitative data, and chi-square test was utilized for qualitative data comparison. The performances of DL-CNN (ft) and DL-CNN (fs) were compared with AUC of receiver operating characteristic (ROC) analysis, together with specificity and sensitivity. Correlations between clot volume and Qanadli score, Mastora score, and right ventricular metrics were evaluated by the Spearman rank test. Prognostic factors were analyzed with binary logistic regression, and $P < 0.05$ was considered the level for statistical significance.

Results

Baseline characteristics

A total of 137 cases with APE (64 males, mean age

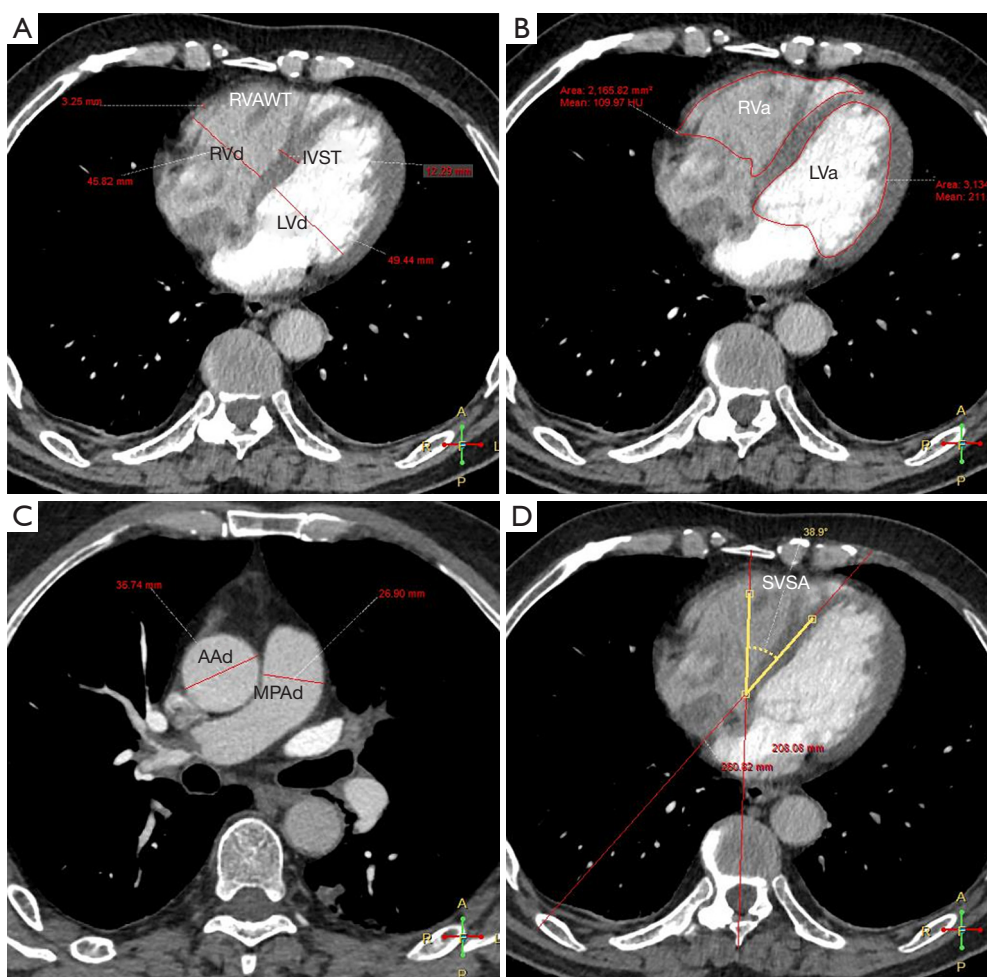


Figure 2 Right ventricular metrics measured on transversal computed tomographic pulmonary angiography image. (A) RVd, LVd, RVAWT, and IVST. (B) RVa and LVa. (C) MPAd and AAd. (D) SVSA (Spinal Ventricular Septal Angle -angle between interventricular septum and chest midline). RVd, Right Ventricular transverse diameter; LVd, Left Ventricular transverse diameter; RVAWT, Right Ventricular Anterior Wall Thickness; IVST, Interventricular Septal Thickness; RVa, Right Ventricle Area; LVa, Left Ventricle Area; MPAd, Main Pulmonary Artery diameter; AAd, Ascending Aortic diameter; SVSA, Spinal Ventricular Septal Angle.

=62.5±16.8 years) and 273 cases without PE (139 males, mean age =61.5±12.4 years) were included to test DL-CNN (ft) and DL-CNN (fs). According to risk stratification, there were 77 low-risk cases, 43 intermediate-risk cases, and 17 high-risk cases among APE patients, while six cases died during hospitalization. According to radiology evaluation, the 137 patients with APE included two cases with type I, 29 cases with type II, and 106 cases with type III. The radiologists reported a total of 1,364 clots. The median value of the Qanadli score and Mastora scores were 14 (IQR, 8–20) and 35 (IQR, 11.5–76.5).

DL-CNN (ft) and DL-CNN (fs) for evaluation of APE

In the 137 patients with APE, no type I locations were correctly detected by DL-CNN (ft) or DL-CNN (fs), as all were wrongly detected as type III. Two cases and one case with type II were correctly detected by DL-CNN (ft) and DL-CNN (fs), respectively. Except for one case, which was wrongly detected as type I by DL-CNN (ft), the others with type II locations were wrongly detected as type III. Both DL-CNN (ft) and DL-CNN (fs) correctly classified 106 cases as type III. In 273 patients without PE, DL-CNN (ft) misclassified 11 cases into type I, five cases into type II, and

Table 1 AUCs, sensitivity, and specificity of the two DL-CNN models on a per patient level

DL-CNN model	AUC of ROC	Sensitivity	Specificity
DL-CNN (ft)	0.886±0.016 (95% CI, 0.855–0.918)	100%	77.29%
DL-CNN (fs)	0.879±0.017 (95% CI, 0.846–0.912)	100%	75.82%

DL-CNN, the deep learning convolution neural network; DL-CNN (ft), the deep learning convolution neural network trained by fine-tuning; DL-CNN (fs), the deep learning convolution neural network trained from scratch; AUC, area under the curve; ROC, receiver operating characteristic curve.

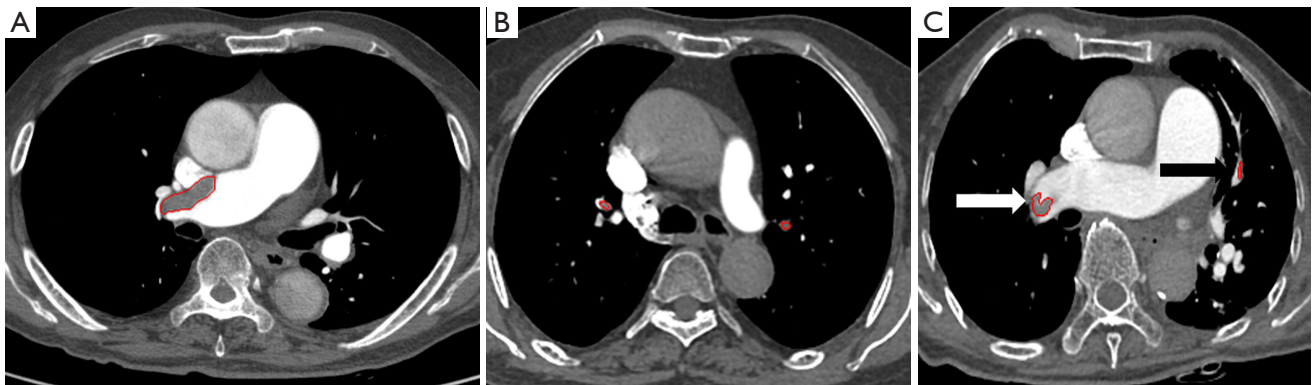


Figure 3 The correct identification and labeling of clots by the DL-CNN model. (A) The correct identification and labeling of a central pulmonary artery clot by DL-CNN (ft) in an 82-year-old male. (B) The correct identification and labeling of a peripheral pulmonary artery clot by DL-CNN (ft) in a 78-year-old female. (C) The correct identification and labeling of central and peripheral pulmonary artery clots by DL-CNN (fs) in a 74-year-old female (white arrow pointing to the central pulmonary artery clot; black arrow pointing to the peripheral pulmonary artery clot). DL-CNN, the deep learning convolution neural network; DL-CNN (ft), the deep learning convolution neural network trained by fine-tuning; DL-CNN (fs), the deep learning convolution neural network trained from scratch.

46 cases into type III, and DL-CNN (fs) misclassified five cases into type I, five cases into type II, and 56 cases into type III.

Table 1 shows the sensitivity and specificity of DL-CNN (ft) in the diagnosis of APE, respectively were 100% and 77.29%, on a per-patient level, and the sensitivity and specificity of DL-CNN (fs) were 100% and 75.82%, respectively. The AUCs of DL-CNN (ft) [0.886±0.016 (95% CI, 0.855–0.918)] and DL-CNN (fs) [0.879±0.017 (95% CI, 0.846–0.912)] in the diagnosis of APE were comparable ($Z=0.30$, $P=0.38$).

Among 1,738 clots labeled by DL-CNN (ft), 1071 were true clots (Figure 3), while 667 were false-positive clots (Figures 4,5). However, 293 clots were missed by DL-CNN (ft) (Figure 6). Among 1,747 clots labeled by DL-CNN (fs), 1,061 were true clots (Figure 3), 686 clots were false-positive clots (Figures 4,5), and 303 clots were missed by DL-CNN (fs) (Figure 6). Table 2 shows the sensitivities and specificities of DL-CNN (ft) and DL-CNN (fs) for detecting central

clots were 99.06% and 100%, and 72.61% and 70.63%, respectively, while the AUCs of DL-CNN (ft) [0.858±0.018 (95% CI, 0.823–0.894)] and DL-CNN (fs) [0.848±0.018 (95% CI, 0.818–0.888)] in the detection of central clots were comparable ($Z=0.20$, $P=0.42$). Table 3 shows the sensitivities and specificities of DL-CNN (ft) in detecting clots of each peripheral pulmonary artery were slightly higher than those of DL-CNN (fs) (except R10), although the AUCs of DL-CNN (ft) and DL-CNN (fs) for detecting clots in each peripheral pulmonary artery were comparable.

Correlations between clot burden and right ventricular metrics

Clot volumes measured with DL-CNN (ft) and DL-CNN (fs) in APE patients were 3.1 mL (IQR, 0.5–10.9) and 3.2 mL (IQR, 0.4–10.3), respectively, and there was no significant difference between them ($U=85094.500$, $P=0.741$). Table 4 shows that clot volumes measured with

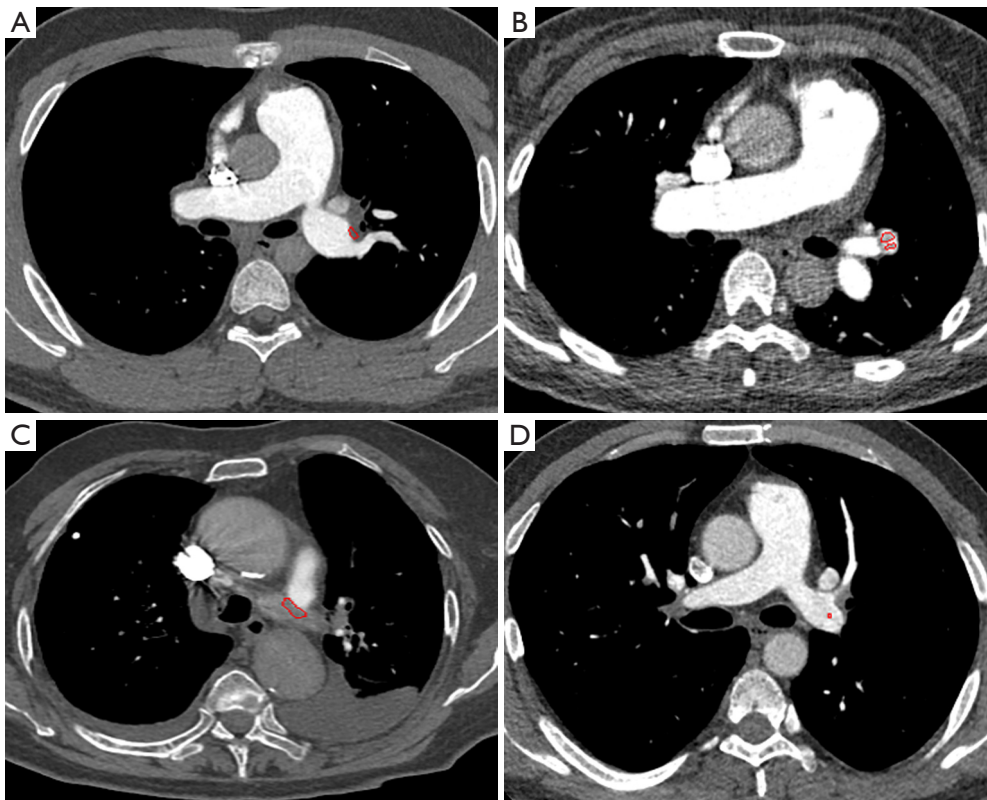


Figure 4 False-positive clot in the central pulmonary artery labeled by the deep learning convolution neural network (DL-CNN) model, caused by (A) surrounding soft tissue, (B) adjacent vein, (C) lymph node, (D) inhomogeneous velocity artifact.

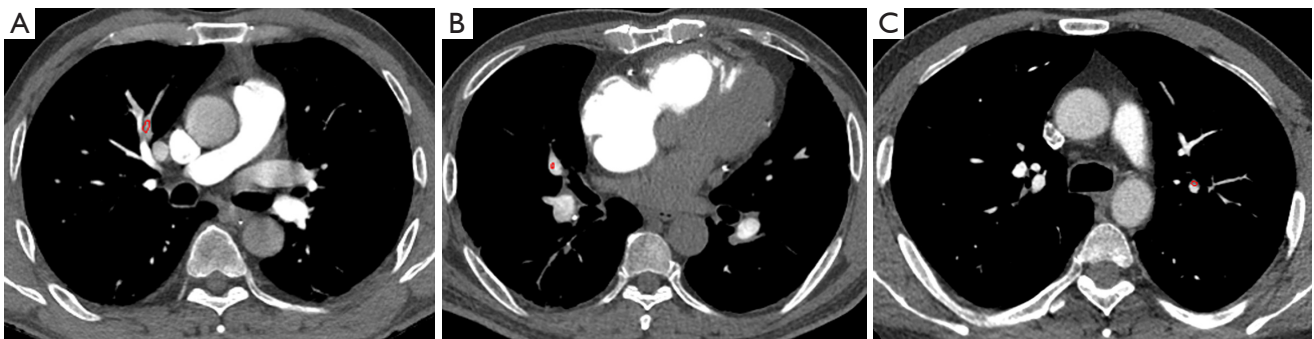


Figure 5 False-positive clot in the peripheral pulmonary artery labeled by the deep learning convolution neural network (DL-CNN) model, caused by (A) adjacent vein, (B) inhomogeneous velocity artifact, (C) surrounding soft tissue.

DL-CNN (ft) and DL-CNN (fs) in APE patients were significantly correlated with Qanadli and Mastora scores. Moreover, clot volumes measured with DL-CNN (ft) and DL-CNN (fs) in APE patients were significantly correlated with RVd/LVd, RVa/LVa, MPAd, MPAd/AAd, and SVSA ($P < 0.05$), and there were no significant correlations between

clot volumes and SVCd, RVAWT, IVST, and AAd ($P > 0.05$).

Clot volumes with risk stratification and in-hospital mortality

Figure 7 shows clot burden measured with DL-CNN (ft),



Figure 6 False-negative clots in the DL-CNN model. (A) DL-CNN (ft) missed clot in a 73-year-old male. (B) DL-CNN (fs) missed clot in a 70-year-old male. (C) R10 clot was identified and labeled as R9 by DL-CNN (fs) in a 65-year-old female. DL-CNN, the deep learning convolution neural network; DL-CNN (ft), the deep learning convolution neural network trained by fine-tuning; DL-CNN (fs), the deep learning convolution neural network trained from scratch.

Table 2 AUCs, sensitivity, and specificity of the two DL-CNN models in detecting clots of the central pulmonary artery

DL-CNN model	AUC of ROC	Sensitivity	Specificity
DL-CNN(ft)	0.858±0.018 (95% CI, 0.823–0.894)	99.06%	72.61%
DL-CNN(fs)	0.848±0.018 (95% CI, 0.818–0.888)	100.00%	70.63%

DL-CNN, the deep learning convolution neural network; DL-CNN (ft), the deep learning convolution neural network trained by fine-tuning; DL-CNN (fs), the deep learning convolution neural network trained from scratch; AUC, area under the curve; ROC, receiver operating characteristic curve.

Table 3 AUCs, sensitivity, and specificity of the two DL-CNN models in detecting clots of the peripheral pulmonary artery

Items	DL-CNN (ft)		DL-CNN (fs)		AUC of DL-CNN (ft)	AUC of DL-CNN (fs)
	Sensitivity	Specificity	Sensitivity	Specificity		
R1	87.84%	91.67%	86.49%	90.17%	0.898±0.024 (95% CI, 0.851–0.944)	0.883±0.025 (95% CI, 0.835–0.932)
R2	80.72%	88.68%	77.11%	89.30%	0.796±0.030 (95% CI, 0.738–0.854)	0.912±0.021 (95% CI, 0.870–0.953)
R3	84.29%	91.47%	82.86%	91.47%	0.879±0.027 (95% CI, 0.826–0.931)	0.872±0.028 (95% CI, 0.817–0.926)
R4	81.25%	88.18%	87.50%	86.97%	0.847±0.027 (95% CI, 0.794–0.901)	0.872±0.024 (95% CI, 0.825–0.919)
R5	87.32%	90.86%	84.51%	90.27%	0.891±0.025 (95% CI, 0.843–0.939)	0.847±0.025 (95% CI, 0.822–0.926)
R6	89.23%	88.70%	86.15%	90.14%	0.890±0.024 (95% CI, 0.842–0.937)	0.881±0.026 (95% CI, 0.830–0.933)

Table 3 (continued)

Table 3 (continued)

Items	DL-CNN (ft)		DL-CNN (fs)		AUC of DL-CNN (ft)	AUC of DL-CNN (fs)
	Sensitivity	Specificity	Sensitivity	Specificity		
R7	96.82%	88.47%	95.24%	88.47%	0.926±0.017 (95% CI, 0.894–0.959)	0.919±0.019 (95% CI, 0.882–0.955)
R8	67.57%	93.75%	60.81%	94.64%	0.807±0.034 (95% CI, 0.740–0.873)	0.777±0.036 (95% CI, 0.707–0.847)
R9	90.00%	90.59%	87.14%	90.00%	0.903±0.023 (95% CI, 0.859–0.947)	0.886±0.025 (95% CI, 0.837–0.935)
R10	0	100%	0	100%	0.500±0.036 (95% CI, 0.429–0.571)	0.500±0.036 (95% CI, 0.429–0.571)
L1.3	76.47%	90.94%	75.00%	91.23%	0.837±0.032 (95% CI, 0.775–0.899)	0.831±0.032 (95% CI, 0.768–0.894)
L2	82.26%	87.64%	83.87%	88.51%	0.850±0.030 (95% CI, 0.791–0.908)	0.862±0.029 (95% CI, 0.805–0.918)
L4	66.15%	92.17%	64.62%	92.17%	0.792±0.036 (95% CI, 0.720–0.863)	0.784±0.037 (95% CI, 0.712–0.856)
L5	95.08%	85.10%	93.44%	83.67%	0.901±0.020 (95% CI, 0.862–0.940)	0.886±0.022 (95% CI, 0.843–0.929)
L6	77.19%	91.22%	75.44%	90.08%	0.842±0.034 (95% CI, 0.776–0.908)	0.828±0.035 (95% CI, 0.760–0.896)
L7.8	72.94%	92.00%	77.65%	89.85%	0.825±0.030 (95% CI, 0.766–0.883)	0.837±0.028 (95% CI, 0.782–0.893)
L9	91.67%	91.71%	75.00%	92.57%	0.867±0.030 (95% CI, 0.807–0.926)	0.838±0.034 (95% CI, 0.771–0.905)
L10	84.29%	87.65%	85.71%	85.29%	0.860±0.027 (95% CI, 0.803–0.907)	0.855±0.027 (95% CI, 0.803–0.907)

R1, apex segment of right superior lobe. R2, anterior segment of right superior lobe. R3, posterior segment of right superior lobe. R4, medial segment of right middle lobe. R5, lateral segment of right middle lobe. R6, posterior segment of right inferior lobe. R7, medial basal segment of right inferior lobe. R8, anterior basal segment of right inferior lobe. R9, lateral basal segment of right inferior lobe. R10, posterior basal segment of right inferior lobe. L1.3, apicoposterior segment of left superior lobe. L2, anterior segment of left superior lobe. L4, superior lingular segment of left superior lobe. L5, inferior lingular segment of left superior lobe. L6, posterior segment of left inferior lobe. L7.8, anteromedial basal segment of left inferior lobe. L9, lateral basal segment of left inferior lobe. L10, posterior basal segment of left inferior lobe. DL-CNN, the deep learning convolution neural network; DL-CNN (ft), the deep learning convolution neural network trained by fine-tuning; DL-CNN(fs), the deep learning convolution neural network trained from scratch; AUC, area under the curve; ROC, receiver operating characteristic curve.

Table 4 Correlations between right ventricular metrics and clot volumes measured with two DL-CNN models

CT metrics	DL-CNN (ft)		DL-CNN (fs)	
	r	P	r	P
Qanadli score	0.825	P<0.001	0.827	P<0.001
Mastora score	0.859	P<0.001	0.864	P<0.001
RVd /LVd	0.476	P<0.001	0.460	P<0.001
RVa /LVa	0.523	P<0.001	0.504	P<0.001
MPAd/(mm)	0.219	0.010	0.218	0.011
MPAd/AAd	0.315	P<0.001	0.311	P<0.001
SVSA	0.279	0.001	0.287	0.001
SVCd (mm)	0.168	0.051	0.155	0.071
RVAWT (mm)	0.039	0.648	0.025	0.768
IVST (mm)	-0.099	0.250	-0.085	0.323
AAd (mm)	-0.122	0.157	-0.114	0.185

RVd, Right Ventricular transverse diameter; LVd, Left Ventricular transverse diameter; RVa, Right Ventricle Area; LVa, Left Ventricle Area; MPAd, Main Pulmonary Artery diameter; AAd, Ascending Aortic diameter; RVAWT, Right Ventricular Anterior Wall Thickness; IVST, Interventricular Septal Thickness; SVCd, Superior Vena Cava maximal diameter; SVSA, Spinal Ventricular Septal Angle. DL-CNN, the deep learning convolution neural network; DL-CNN (ft), the deep learning convolution neural network trained by fine-tuning; DL-CNN (fs), the deep learning convolution neural network trained from scratch.

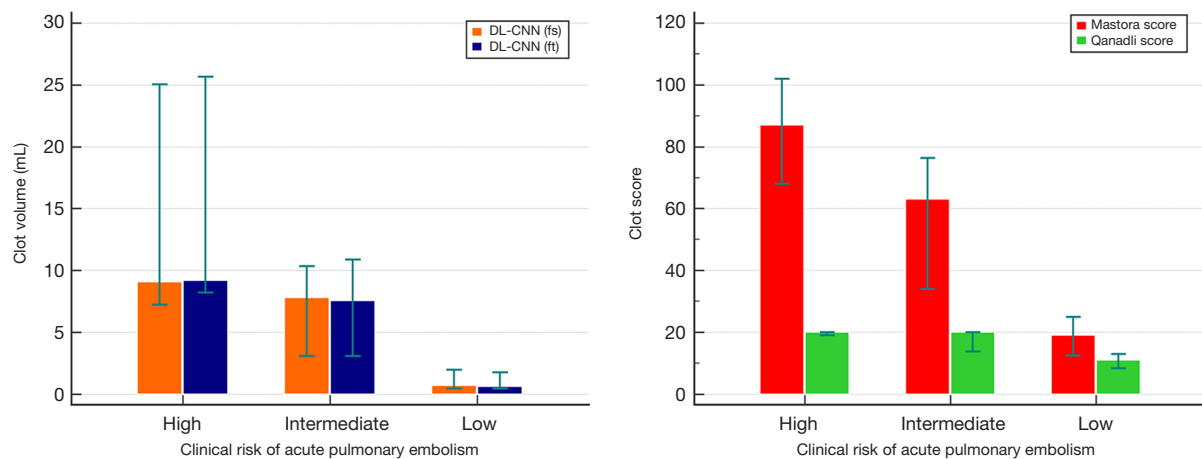


Figure 7 Clot burden in low-risk, intermediate-risk, and high-risk APE patients. (A) Clot burden (Y-axis: clot volume) evaluated by DL-CNN (fs) and DL-CNN (ft). (B) Clot burden (Y-axis: clot score) evaluated by Mastora score and Qanadli score. DL-CNN (ft), the deep learning convolution neural network trained by fine-tuning; DL-CNN (fs), the deep learning convolution neural network trained from scratch.

DL-CNN (fs), and Mastora score increased in low-risk, moderate-risk, and high-risk patients, while the Qanadli score in intermediate-risk and high-risk patients were comparable. Binary logistic regression indicated that RVa/LVa was the independent prognostic factor of in-hospital

death (odds ratio =6.73; 95% CI, 2.7–18.12; P<0.001).

Discussion

DL-CNN (ft) and DL-CNN (fs) detected clots on a per-

patient level and in the central pulmonary artery with high sensitivities and moderate specificities, while both methods detected clots in the peripheral pulmonary artery with moderate to high sensitivities and specificities. Clot volumes from DL-CNN (ft) were comparable to DL-CNN (fs) and closely correlated with right ventricular metrics. Further, although clot burdens measured with DL-CNN (ft) and DL-CNN (fs) were increased in low-risk, moderate-risk, and high-risk patients, RVa/LVa was the independent prognostic factor for in-hospital death.

While in our initial study, only DL-CNN (fs) was tested (14), in the present study, we compared DL-CNN (ft) and DL-CNN (fs) and found that on a per-patient level, both had high sensitivities and moderate specificities in the diagnosis of APE. Weikert *et al.* (19) reported that DL-CNN had a sensitivity of 92.7% and specificity of 95.5% in detecting PE. We speculate that the difference between these results and ours is because of differences in the proportion of positive and negative subjects.

Both DL-CNN (ft) and DL-CNN (fs) also showed high sensitivities and moderate specificities in detecting central clots on a clot level. Soft tissue around the central pulmonary artery and adjacent vein, hilar soft tissue and hilar lymph nodes, inhomogeneous velocity artifact, and radiation artifact of the superior vena cava contrast media were sometimes wrongly detected as central clots. DL-CNN (ft) and DL-CNN (fs) showed moderate to high sensitivities and specificities in detecting peripheral clots. Breathing artifact, the pulmonary vein, inhomogeneous velocity artifact, and soft tissue around the peripheral pulmonary artery were the main causes of false-positive peripheral clots, while clots in the occluded peripheral pulmonary artery were easily missed, causing false negatives. As the performances of DL-CNN (ft) and DL-CNN (fs) were comparable in detecting both central and peripheral clots, both can be used to screen APE. However, if there is a positive result, this should be manually checked, while if there is a negative result, it is suggested that APE can be excluded.

Qanadli and Mastora scores were used to assess the clot burden of APE semi-quantitatively. Our previous research (14) showed that clot volumes from DL-CNN (fs) were correlated with Qanadli and Mastora scores, the sensitivity and specificity of DL-CNN (fs) in detecting clots were 94.6% and 76.5%, respectively, and the AUC was 0.926 (95% CI, 0.884–0.968). In the present study, clot volumes from DL-CNN (ft) were comparable to DL-CNN (fs), and clot volumes from both correlated with Qanadli and

Mastora scores. The sensitivity and specificity of DL-CNN (ft) in the diagnosis of APE were 100% and 77.29%, respectively, while the sensitivity and specificity of DL-CNN (fs) were 100%, 75.82%, respectively. AUCs of DL-CNN (ft) and DL-CNN (fs) in the diagnosis of APE were comparable. While our previous research (14) was mainly on a per-patient level, in the present study, we further analyzed the performance of DL-CNN (ft) and DL-CNN (fs) in measuring clot distribution and volume on the level of lobe and segment arteries, which is more conducive for clinicians to provide better-targeted treatment.

The Mastora index was shown to reflect clot burden and change of right heart function (20), and the Qanadli index was shown to be a strong independent predictor of right ventricular dysfunction in APE (21), correlating linearly with different variables associated with higher morbidity and mortality (22). Abdelwahab *et al.* (23) found a significant correlation between clot volume and parameters of RV dysfunction assessed by CTPA, but there was no significant correlation between the two using echocardiography. Rodrigues *et al.* (24) also found that most usual echocardiographic parameters evaluating RV fail to demonstrate a good correlation with clot burden, but Bach *et al.* (25) found there was no association of global obstruction and prognosis. Ghuysen *et al.* (26) also found that the pulmonary obstruction index could not predict patient outcome. In the current study, clot volumes from DL-CNN (ft) and DL-CNN (fs) were correlated with RV function metrics such as RVa/LVa and RVd/LVd. This confirmed the effect of clot burdens of APE on right ventricular size and function. Clot burdens measured with DL-CNN (ft) and DL-CNN (fs) increased in low-risk, moderate-risk, and high-risk patients, suggesting clot volume is useful for risk stratification. Shen *et al.* (5) showed clot volume measured with both CADe and RVd/LVd were two independent factors of high-risk in APE patients, and a prospective multicenter cohort study in 457 patients showed a right-to-left ventricular dimensional ratio ≥ 0.9 was an independent predictor of adverse in-hospital outcomes (27). Our results showed that only RVa/LVa was an independent prognostic factor of death during hospitalization, consistent with those of previous studies (28–30). This also suggests that clot burden may be related to risk stratification but may not determine the short-term prognosis of APE.

Study limitations

The strength of our study is its use of different CT

scanners to achieve a higher external validity and better robustness of the DL-CNN model. However, there were several limitations in the study. Firstly, we only segmented the lung to the level of lobes and segments and not to the specific segmental arteries. Secondly, we only analyzed the correlation of metrics on CTPA and the short-term prognosis during hospitalization. Only patients with APE and negative cases were included, which limited DL-CNN in evaluating chronic PE. Although clot burden correlated with RV function and risk-stratification, RVa/LVa were the independent predictors of in-hospital death. Future studies should further refine pulmonary artery segmentation and concentrate on automatically segmenting clot and ventricular size by DL-CNN.

Conclusions

Our study confirmed that both DL-CNN (ft) and DL-CNN (fs) have high sensitivities and moderate specificities in detecting clots, and their performances are comparable. While clot volumes measured with the two DL-CNN models correlated with right ventricular function and clinical risk stratification, RVa/LVa was the independent prognostic factor of in-hospital death.

Acknowledgments

Funding: This work was supported by the Beijing Nature Science Foundation (Grant No. 7182149), National Natural Science Foundation of China (Grant No. 81871328), Youth Talents project of Chinese Academy of Medical Science (Grant No. 2018RC320013), and Beijing Science and Technology Commission Pharmaceutical and Technology Innovation Project (Grant No. Z18110000 1918034).

Footnote

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://dx.doi.org/10.21037/qims-21-140>). 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 the medical ethics committee of China Rehabilitation Research Center (IRB No. 2020-070-1). The

study was conducted in accordance with the Declaration of Helsinki (as revised in 2013), and informed consent from patients was waived for this retrospective study.

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

References

1. Tuzovic M, Adigopula S, Amsallem M, Kobayashi Y, Boulate D, Krishnan G, Liang D, Schnittger I, McConnell MV, Haddad F. Abstract 10293: regional right ventricular dysfunction in acute pulmonary embolism associated with increased clot burden and greater RV dysfunction. *Circulation* 2015;132:A10293.
2. El-Menyar A, Nabir S, Ahmed N, Asim M, Jabbour G, Al-Thani H. Diagnostic implications of computed tomography pulmonary angiography in patients with pulmonary embolism. *Ann Thorac Med* 2016;11:269-76.
3. Qanadli SD, El Hajjam M, Vieillard-Baron A, Joseph T, Mesurole B, Oliva VL, Barré O, Bruckert F, Dubourg O, Lacombe P. New CT index to quantify arterial obstruction in pulmonary embolism: comparison with angiographic index and echocardiography. *AJR Am J Roentgenol* 2001;176:1415-20.
4. Mastora I, Remy-Jardin M, Masson P, Galland E, Delannoy V, Bauchart JJ, Remy J. Severity of acute pulmonary embolism: evaluation of a new spiral CT angiographic score in correlation with echocardiographic data. *Eur Radiol* 2003;13:29-35.
5. Shen C, Yu N, Wen L, Zhou S, Dong F, Liu M, Guo Y. Risk stratification of acute pulmonary embolism based on the clot volume and right ventricular dysfunction on CT pulmonary angiography. *Clin Respir J* 2019;13:674-82.
6. Furlan A, Aghayev A, Chang CC, Patil A, Jeon KN, Park B, Fetzer DT, Saul M, Roberts MS, Bae KT. Short-term mortality in acute pulmonary embolism: clot burden and signs of right heart dysfunction at CT pulmonary angiography. *Radiology* 2012;265:283-93.
7. Ouriel K, Ouriel RL, Lim YJ, Piazza G, Goldhaber SZ. Computed tomography angiography with pulmonary

- artery thrombus burden and right-to-left ventricular diameter ratio after pulmonary embolism. *Vascular* 2017;25:54-62.
8. Cui FZ, Gong TT, Liu JH, Mu XG. Research progress of artificial intelligence in diagnosis of chest diseases. *Chinese Medical Equipment* 2019;34:164-7.
 9. Zhou T, Tan T, Pan X, Tang H, Li J. Fully automatic deep learning trained on limited data for carotid artery segmentation from large image volumes. *Quant Imaging Med Surg* 2021;11:67-83.
 10. Xia Y, Lu S, Wen L, Eberl S, Fulham M, Feng DD. Automated identification of dementia using FDG-PET imaging. *Biomed Res Int* 2014;2014:421743.
 11. Girshick R, Donahue J, Darrell T, Malik J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. *IEEE Conference on Computer Vision and Pattern Recognition* 2014:580-7.
 12. Yosinski J, Clune J, Bengio Y, Lipson H. How transferable are features in deep neural networks? *Adv Neural Inf Process Syst* 2014:3320-28.
 13. Tajbakhsh N, Shin JY, Gurudu SR, Hurst RT, Kendall CB, Gotway MB, Jianming Liang. Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning? *IEEE Trans Med Imaging* 2016;35:1299-312.
 14. Liu W, Liu M, Guo X, Zhang P, Zhang L, Zhang R, Kang H, Zhai Z, Tao X, Wan J, Xie S. Evaluation of acute pulmonary embolism and clot burden on CTPA with deep learning. *Eur Radiol* 2020;30:3567-75.
 15. Konstantinides SV, Meyer G, Becattini C, Bueno H, Geersing GJ, Harjola VP, et al. 2019 ESC Guidelines for the diagnosis and management of acute pulmonary embolism developed in collaboration with the European Respiratory Society (ERS): The Task Force for the diagnosis and management of acute pulmonary embolism of the European Society of Cardiology (ESC). *Eur Respir J* 2019;54:1901647.
 16. Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. *J Mach Learn Res* 2010; 9:249-56.
 17. Liu M, Ma Z, Guo X, Zhang H, Yang Y, Wang C. Computed tomographic pulmonary angiography in the assessment of severity of chronic thromboembolic pulmonary hypertension and right ventricular dysfunction. *Eur J Radiol* 2011;80:e462-9.
 18. Liu M, Guo X, Zhu L, Zhang H, Hou Q, Guo Y, Yang Y, Jiang T. Computed Tomographic Pulmonary Angiographic Findings Can Predict Short-Term Mortality of Saddle Pulmonary Embolism: A Retrospective Multicenter Study. *J Comput Assist Tomogr* 2016;40:327-34.
 19. Weikert T, Winkel DJ, Bremerich J, Stieltjes B, Parmar V, Sauter AW, Sommer G. Automated detection of pulmonary embolism in CT pulmonary angiograms using an AI-powered algorithm. *Eur Radiol* 2020;30:6545-53.
 20. Chen S, Cheng R, Zhang G. Comparison of value of Qanadli versus Mastora pulmonary embolism index in evaluating straddle-type pulmonary embolism. *Zhonghua Yi Xue Za Zhi* 2014;94:3629-32.
 21. Rodrigues B, Correia H, Figueiredo A, Delgado A, Moreira D, Ferreira Dos Santos L, Correia E, Pipa J, Beirão I, Santos O. Clot burden score in the evaluation of right ventricular dysfunction in acute pulmonary embolism: quantifying the cause and clarifying the consequences. *Rev Port Cardiol* 2012;31:687-95.
 22. Praveen Kumar BS, Rajasekhar D, Vanajakshamma V. Study of clinical, radiological and echocardiographic features and correlation of Qanadli CT index with RV dysfunction and outcomes in pulmonary embolism. *Indian Heart J* 2014;66:629-34.
 23. Abdelwahab HW, Arafa S, Bondok K, Batouty N, Bakeer M. Relationship between clot burden in pulmonary computed tomography angiography and different parameters of right cardiac dysfunction in acute pulmonary embolism. *Cardiovasc J Afr* 2020;31:21-4.
 24. Rodrigues AC, Guimaraes L, Guimaraes JF, Monaco C, Cordovil A, Lira E, Vieira ML, Fischer CH, Nomura C, Morhy S. Relationship of clot burden and echocardiographic severity of right ventricular dysfunction after acute pulmonary embolism. *Int J Cardiovasc Imaging* 2015;31:509-15.
 25. Bach AG, Nansalmaa B, Kranz J, Taute BM, Wienke A, Schramm D, Surov A. CT pulmonary angiography findings that predict 30-day mortality in patients with acute pulmonary embolism. *Eur J Radiol* 2015;84:332-7.
 26. Ghuysen A, Ghaye B, Willems V, Lambermont B, Gerard P, Dondelinger RF, D'Orio V. Computed tomographic pulmonary angiography and prognostic significance in patients with acute pulmonary embolism. *Thorax* 2005;60:956-61.
 27. Becattini C, Agnelli G, Vedovati MC, Pruszczyk P, Casazza F, Grifoni S, Salvi A, Bianchi M, Douma R, Konstantinides S, Lankeit M, Duranti M. Multidetector computed tomography for acute pulmonary embolism: diagnosis and risk stratification in a single test. *Eur Heart J* 2011;32:1657-63.
 28. Ghaye B, Ghuysen A, Willems V, Lambermont B, Gerard P, D'Orio V, Gevenois PA, Dondelinger RF. Severe

- pulmonary embolism: pulmonary artery clot load scores and cardiovascular parameters as predictors of mortality. *Radiology* 2006;239:884-91.
29. Jia D, Zhou XM, Hou G. Estimation of right ventricular dysfunction by computed tomography pulmonary angiography: a valuable adjunct for evaluating the severity of acute pulmonary embolism. *J Thromb Thrombolysis* 2017;43:271-8.
30. Becattini C, Agnelli G, Germini F, Vedovati MC. Computed tomography to assess risk of death in acute pulmonary embolism: a meta-analysis. *Eur Respir J* 2014;43:1678-90.

Cite this article as: Zhang H, Cheng Y, Chen Z, Cong X, Kang H, Zhang R, Guo X, Liu M. Clot burden of acute pulmonary thromboembolism: comparison of two deep learning algorithms, Qanadli score, and Mastora score. *Quant Imaging Med Surg* 2022;12(1):66-79. doi: 10.21037/qims-21-140