Original research

Differentiation of malignant from benign pleural effusions based on artificial intelligence

Sufei Wang,1 Xueyun Tan,1 Piqiang Li,2 Qianqian Fan,3 Hui Xia,1 Shan Tian,4 Feng Pan,3 Na Zhan,3 Rong Yu,6 Liang Zhang,7 Yanran Duan,8 Juanjuan Xu,1 Yanling Ma,1 Wenjuan Chen,1 Yan Li,9 Zilin Zhao,1 Chaoyang Liu,2 Qingjia Bao,2 Lian Yang1

ABSTRACT

Introduction This study aimed to construct artificial intelligence models based on thoracic CT images to perform segmentation and classification of benign pleural effusion (BPE) and malignant pleural effusion (MPE).

Methods A total of 918 patients with pleural effusion were initially included, with 607 randomly selected cases used as the training cohort and the other 311 as the internal testing cohort; another independent external testing cohort with 362 cases was used. We developed a pleural effusion segmentation model (M1) by combining 3D spatially weighted U-Net with 2D classical U-Net. Then, a classification model (M2) was built to identify BPE and MPE using a CT volume and its 3D pleural effusion mask as inputs.

Results The average Dice similarity coefficient, Jaccard coefficient, precision, sensitivity, Hausdorff distance 95% (HD95) and average surface distance indicators in M1 were 87.6±5.0%, 82.2±6.2%, 99.0±1.0%, 83.0±6.6%, 6.9±3.8 and 1.6±1.1, respectively, which were better than those of the 3D U-Net and 3D spatially weighted U-Net. Regarding M2, the area under the receiver operating characteristic curve, sensitivity and specificity obtained with volume concat masks as input were 0.842 (95% CI 0.801 to 0.878), 89.4% (95% CI 84.4% to 93.2%) and 65.1% (95% CI 57.3% to 72.3%) in the external testing cohort. These performance metrics were significantly improved compared with those for the other input patterns.

Conclusions We applied a deep learning model to the segmentation of pleural effusions, and the model showed encouraging performance in the differential diagnosis of BPE and MPE.

INTRODUCTION

Effusions, including pleural effusions, ascites, pericardial effusions and abscesses, are commonly observed in many diseases, such as infections and various cancers. The most common effusions are malignant pleural effusions (MPEs) caused by lung cancer, breast cancer, lymphoma and so on, and benign pleural effusions (BPEs) caused by Mycobacterium tuberculosis infection, heart failure, parapneumonic infections and so on.1–3 The most common conditions leading to ascites are liver disease, cirrhosis and cancer.4 Because pleural effusions are representative effusions, we chose MPE and BPE as our study objects. The gold standard in the diagnosis of MPE and BPE depends on pleural effusion pathogenic/cytological examinations and thoracentesis with pleural biopsy.5–6 However, the low positivity rates for pathogenic diagnosis, the invasiveness and high costs of pleural biopsy, and the risk of complications represent the limitations of these gold-standard techniques, although their high specificity is their most important advantage.7–8 These limitations suggest opportunities for more convenient, highly sensitive and non-invasive methods to improve the diagnostic performance of BPE and MPE.

Thoracic CT is an appropriate method for the further assessment of pleural effusion.9 Because the features extracted from images by radiologists are limited, artificial intelligence (AI) deep learning algorithms are helpful tools for automatically
analysing complex medical images thanks to their strong feature-learning ability. However, no available studies have focused on the differential diagnosis of pleural effusion based on thoracic CT image analysis using deep learning algorithms. U-Net, a convolutional neural network, has become an increasingly important basis for many deep learning models in medical image analysis. It can achieve remarkable generalisation performance when trained with a limited number of images, which makes it especially suitable for our research. In previous studies, U-Net was used for the segmentation of solid organs and lesion regions, such as pancreas segmentation, and 3D cardiac segmentation and automatic ground-glass nodule detection. In our study, we applied U-Net to the segmentation of effusion regions, and specifically pleural effusions. We combined the 3D spatially weighted U-Net with the 2D classical U-Net for pleural effusion segmentation in thoracic CT images to obtain fine masks. The high precision of pleural effusion segmentation identifies predictive features which can be subsequently used to train deep learning models for lesion classification. We thus proposed a deep learning algorithm, based on the global and partial analysis of thoracic CT image features to diagnose BPE and MPE, which can potentially play a critical role in improving patients’ clinical prognosis.

MATERIALS AND METHODS
Patients and study design
In this consecutive study, 918 pleural effusion cases retrospectively collected from Wuhan Union Hospital between January 2016 and December 2021 were enrolled, with 311 cases randomly selected as the internal testing cohort and the other 607 as the training cohort. Another independent cohort including 362 patients with pleural effusion collected from Renmin Hospital of Wuhan University between January 2020 and May 2022 was used as the external testing cohort. Patients who met the following inclusion criteria were enrolled: (1) diagnosed with pleural effusion by CT scan of the chest and (2) underwent pleural effusion pathogenic/cytological examinations and diagnostic thoracentesis with or without pleural biopsy. The exclusion criteria were (1) pleural effusion whose cause could not be determined, (2) under the age of 18 and (3) unavailable clinical information. The diagnostic criteria for MPE and BPE adopted in this study are based on our previous studies, and the criteria are further described in online supplemental methods.

Two professional physicians, HX and WC, collected clinical information, including demographic characteristics, radiological features and laboratory testing results of the enrolled patients from electronic medical records. The volume of pleural effusion was classified as mild (<500 mL), moderate (500–1000 mL) or severe (>1000 mL). The parameters of the CT scanner are presented in online supplemental methods.

Architecture of the pleural effusion segmentation model (M1)
The pleural effusion segmentation model (M1) is a cascaded two-step deep-learning model (figure 1A). The initial coarse segmentation results are obtained from the spatial attention information based on the 3D spatially weighted U-Net. We used a 3D spatial attention mechanism to capture large-scale contextual features of unrelated image regions, we first extract the results of the pleural effusion segmentation models (the coarse segmentation model based on the 3D spatially weighted U-Net and the fine segmentation model based on the 2D classical U-Net). Differential diagnosis is then achieved by the pleural effusion classification model based on the residual blocks and SE blocks. (A) Schematic flowchart of the proposed algorithm for pleural effusion segmentation. (B) Schematic flowchart of the proposed algorithm for predicting the probability of pleural effusion being MPE. (C) Architecture of the 3D attention layer in the segmentation model. (D) Architecture of the SE block in the classification model. BPE, benign pleural effusion; MPE, malignant pleural effusion; SE, squeeze-and-excitation.
information, thus enhancing the representative ability of the model (figure 1C). However, owing to the huge amount of information per sample for 3D U-Net, the region of interest needs to be cropped into small image patches to be used as inputs. Using this method, natural contour information may be lost to some degree. Therefore, the model’s learning of natural contour information was enhanced through the 2D classical U-Net. The concatenation of 3D spatially weighted U-Net and 2D classical U-Net helps obtain a fine segmentation of pleural effusion. Details about the algorithms of the coarse and fine segmentation models are shown in online supplemental methods.

**Architecture of the pleural effusion classification model (M2)**

We developed a 3D deep convolutional neural network to identify patients with MPE from thoracic CT volumes. As shown in figure 1B, this classification model (M2) uses CT volume and its 3D pleural effusion masks as inputs (details in online supplemental methods). The 3D pleural effusion fine masks are obtained by assembling 2D fine masks generated by the fine segmentation model. This component takes advantage of both stacked bottleneck blocks and squeeze-and-excitation (SE) blocks. The bottleneck block is introduced to extract deeper features from the CT volumes and to solve the problem of degradation in the network training process. The SE block is introduced to improve the representational power of the network by enabling it to perform dynamic channel-wise feature recalibration (figure 1D; online supplemental methods). Inputting the 3D fine masks of pleural effusion according to the 3D thoracic CT volume helps reduce the effects of background information and improves the classification of BPE and MPE. Details about the training process of the pleural effusion segmentation and classification model are shown in online supplemental methods.

**Quantitative assessment indicators**

For the pleural effusion segmentation model, the Dice similarity coefficient (DSC) and Jaccard coefficient were used to evaluate the spatial overlap between the model-generated contour (M) and the ground truth contour (G). In our study, G means the sets defined by these boundaries of pleural effusion area generated by the AI model. Precision and sensitivity measure the detection capability for identifying the correct regions. The Hausdorff distance 95% (HD95) and average surface distance (ASD) measure the boundary similarity between the model-generated contour and the ground truth contour. Details about the above indicators are described in online supplemental methods.

The area under the receiver operator characteristic (ROC) curve (AUC), sensitivity and specificity were used to evaluate the predictive performance of M2 as a pleural effusion classification model.

**Statistical analysis**

Implementation details are described thoroughly in online supplemental methods. Statistical analyses were performed using SPSS Statistics (V.22). Comparisons were performed using the Mann-Whitney U test for continuous variables and χ² or Fisher’s exact test for categorical variables, as appropriate. ROC curves were generated to evaluate the classification performance. Statistical significance was defined as a two-sided p value <0.05.

**RESULTS**

**Baseline characteristics of patients**

The baseline characteristics of the enrolled patients are presented in table 1. We compared the distribution of sex, age, volume of pleural effusion (mild/moderate/severe) and unilateral/bilateral pleural effusion between the two groups (BPE vs MPE). There was no significant difference between the two groups in any of the three cohorts in terms of age and volume of pleural effusion. However, the distribution of gender in patients with MPE and BPE was significantly distinct (p<0.001) in all three cohorts. Lung cancer was the leading cause of MPE (24.2% in the training cohort, 24.7% in the internal testing cohort, 40.3% in the external testing cohort), while parapneumonics was the leading cause of BPE (18.3% in the training cohort, 19.3% in the internal testing cohort, 24.3% in the external testing cohort; table 2).

**3D spatially weighted attention mechanism in coarse segmentation for pleural effusion area discovery**

To highlight the importance of the attention mechanism inserted in 3D U-Net, figure 2 depicts the pleural effusion areas delineated

---

**Table 1** Patient characteristics in the training, internal testing and external testing cohorts

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Training cohort</th>
<th>Internal testing cohort</th>
<th>External testing cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MPE (n=341)</td>
<td>BPE (n=266)</td>
<td>MPE (n=179)</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>177 (51.9%)</td>
<td>81 (30.5%)</td>
<td>101 (56.4%)</td>
</tr>
<tr>
<td>Male</td>
<td>164 (48.1%)</td>
<td>185 (69.5%)</td>
<td>78 (43.6%)</td>
</tr>
<tr>
<td>Age, median (IQR), years</td>
<td>60.0 (50.0–68.0)</td>
<td>61.0 (48.0–71.0)</td>
<td>0.694</td>
</tr>
<tr>
<td>Unilateral pleural effusion</td>
<td>179 (52.5%)</td>
<td>127 (47.7%)</td>
<td>0.246</td>
</tr>
<tr>
<td>Bilateral pleural effusion</td>
<td>162 (47.5%)</td>
<td>139 (52.3%)</td>
<td>0.750</td>
</tr>
<tr>
<td>Volume of pleural effusion</td>
<td>236 (69.2%)</td>
<td>186 (69.9%)</td>
<td>124 (69.3%)</td>
</tr>
</tbody>
</table>

Data shown are the number and percentage of patients, with the exception of age (median and IQR). Statistical comparisons were performed for each clinical variable between the two groups (MPE vs BPE). P values were computed using the Mann-Whitney U test for age as a continuous variable and the χ² test or Fisher’s exact test for categorical variables, as appropriate.

BPE, benign pleural effusion; MPE, malignant pleural effusion.

---


3
by 3D U-Net and 3D spatially weighted U-Net, respectively, and displays heatmaps to indicate the importance of each part of the pleural effusion areas. The cut-off value used to acquire the high-response area was 0.5. It can be clearly observed that the high-response areas were mainly concentrated at the pleural effusion boundary when using 3D spatially weighted U-Net, while they were gathered in the pleural effusion inner part when using 3D U-Net, showing that the attention mechanism significantly improved the accuracy of the pleural effusion area segmentation. Compared with 3D U-Net segmentation, the results of 3D spatially weighted U-Net segmentation better fit the ground truth.

Comparison among 3D U-Net, 3D spatially weighted U-Net and segmentation deep learning model (M1)

For visual demonstration, representative pleural effusion area segmentation results of M1 (two-step method: 3D spatially weighted U-Net and 2D classical U-Net) are compared with the results of 3D spatially weighted U-Net (one-step method).

Table 2 Causes of pleural effusion in the training, internal testing and external testing cohorts

<table>
<thead>
<tr>
<th>Causes</th>
<th>Training cohort (n=607)</th>
<th>Internal testing cohort (n=311)</th>
<th>External testing cohort (n=362)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Malignant effusions</td>
<td>341 (56.1%)</td>
<td>179 (57.6%)</td>
<td>204 (56.3%)</td>
</tr>
<tr>
<td>Lung cancer</td>
<td>147 (24.2%)</td>
<td>77 (24.7%)</td>
<td>146 (40.3%)</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>91 (15.0%)</td>
<td>46 (14.8%)</td>
<td>20 (5.5%)</td>
</tr>
<tr>
<td>Lymphoma</td>
<td>64 (10.5%)</td>
<td>36 (11.6%)</td>
<td>2 (0.6%)</td>
</tr>
<tr>
<td>Gastrointestinal cancer</td>
<td>28 (4.6%)</td>
<td>13 (4.2%)</td>
<td>24 (6.6%)</td>
</tr>
<tr>
<td>Ovarian</td>
<td>4 (0.6%)</td>
<td>4 (1.3%)</td>
<td>7 (1.9%)</td>
</tr>
<tr>
<td>Others</td>
<td>7 (1.2%)</td>
<td>3 (1.0%)</td>
<td>5 (1.4%)</td>
</tr>
<tr>
<td>Benign effusions</td>
<td>266 (43.9%)</td>
<td>132 (42.4%)</td>
<td>158 (43.7%)</td>
</tr>
<tr>
<td>Parapneumonics</td>
<td>111 (18.3%)</td>
<td>60 (19.3%)</td>
<td>88 (24.3%)</td>
</tr>
<tr>
<td>Heart failure</td>
<td>59 (9.7%)</td>
<td>25 (8.0%)</td>
<td>11 (3.1%)</td>
</tr>
<tr>
<td>Tuberculous pleuritis</td>
<td>83 (13.7%)</td>
<td>43 (13.8%)</td>
<td>36 (10.0%)</td>
</tr>
<tr>
<td>Pulmonary embolism</td>
<td>3 (0.5%)</td>
<td>0 (0.0%)</td>
<td>3 (0.8%)</td>
</tr>
<tr>
<td>Pericardial diseases</td>
<td>1 (0.2%)</td>
<td>1 (0.3%)</td>
<td>3 (0.8%)</td>
</tr>
<tr>
<td>Others</td>
<td>9 (1.5%)</td>
<td>3 (1.0%)</td>
<td>17 (4.7%)</td>
</tr>
</tbody>
</table>

Data shown are the number and percentage of patients.

Figure 2 Pleural effusion area discovery. The cut-off value used to acquire the high-response area in heatmaps was 0.5. The second and third rows show the heatmap of 3D U-Net and the 3D spatially weighted U-Net, respectively. The fourth and fifth rows show the pleural effusion area discovered by 3D U-Net and the 3D spatially weighted U-Net, respectively.

Figure 3 Comparison of the segmentation results of a patient using the ground truth contour, 3D spatially weighted U-Net (one-step method) and M1 (two-step method: 3D spatially weighted U-Net and 2D classical U-Net) in 2D and 3D views. The first column shows the images in three different CT slices (A–C) and the 3D (D) view of a patient’s thoracic CT scan. The second column (a1, b1, c1 and d1) shows the corresponding experts’ ground truth contour (pleural effusion outline highlighted in red). The third column (a2, b2, c2 and d2) shows the pleural effusion segmentation results (pleural effusion outline highlighted in yellow) using 3D spatially weighted U-Net. The fourth column (a3, b3, c3 and d3) shows the pleural effusion segmentation results (pleural effusion outline highlighted in blue) using M1.
Table 3 Comparison of the proposed segmentation model (M1) with the 3D U-Net and 3D spatially weighted U-Net in terms of similarity, difference and positioning performance

<table>
<thead>
<tr>
<th>Methods</th>
<th>DSC (%)</th>
<th>Jaccard (%)</th>
<th>Precision (%)</th>
<th>Sensitivity (%)</th>
<th>HD95</th>
<th>ASD</th>
</tr>
</thead>
<tbody>
<tr>
<td>3D U-Net</td>
<td>30.2±4.9</td>
<td>26.7±6.3</td>
<td>76.2±4.1</td>
<td>29.4±7.8</td>
<td>197.6±14.7</td>
<td>4.3±0.4</td>
</tr>
<tr>
<td>3D spatially weighted U-Net</td>
<td>67.1±2.4</td>
<td>69.8±2.7</td>
<td>81.0±3.3</td>
<td>83.6±5.3</td>
<td>72.7±11.2</td>
<td>4.7±0.8</td>
</tr>
<tr>
<td>M1</td>
<td>87.6±5.0</td>
<td>82.2±6.2</td>
<td>99.0±1.0</td>
<td>83.0±6.6</td>
<td>6.9±3.8</td>
<td>1.6±1.1</td>
</tr>
</tbody>
</table>

Data shown are means±SD.
ASD, average surface distance; DSC, Dice similarity coefficient; HD95, Hausdorff distance 95%; M1, two-step method (3D spatially weighted U-Net and 2D classical U-Net).

Diagnosis of pleural effusions.

The proposed M2 model with volume concat mask as input consistently achieved the highest accuracy across the internal and external testing cohorts (figure 4A,B). In addition, the classification score indicated a notable distinction between BPE and MPE with different input patterns in the internal and external testing cohorts (all p<0.001). The input with volume concat mask bore the most significant distinction between BPE and MPE in both the internal and external testing cohorts, as revealed by the violin plots (figure 4C,D).

AUC, sensitivity and specificity were used as the main indicators for evaluating the diagnostic performance of M2. The three indicators for input with volume concat mask were 0.883 (95% CI 0.841 to 0.916), 78.4% (95% CI 71.6% to 84.2%) and 86.2% (95% CI 79.0% to 91.6%) in the internal testing cohort, and 0.842 (95% CI 0.801 to 0.878), 89.4% (95% CI 84.4% to 93.2%) and 65.1% (95% CI 57.3% to 72.3%) in the external testing cohort, which were significantly improved compared with those for input with only volume and input with volume mask multiply (table 4). The similar AUC values of the internal and external testing cohorts suggested an encouraging level of generalisability of M2 for diagnosing BPE and MPE in new patients. The input with the volume concat mask significantly improved the classification performance of M2, while, notably, the decrease in the speed of the network compared with the other two input patterns was negligible.

Comparison of the heatmaps between typical MPE and BPE

Comparison of the activation heatmaps generated by M2 between two randomly selected patients with MPE (one with lung cancer, one with breast cancer) and two randomly selected patients with BPE (one with tuberculous pleuritis, one with heart failure) is shown in figure 5. The activation heatmaps indicated the importance of different parts of the pleural effusion regions and suggested that different areas drew the attention of M2 to various degrees. The important areas found by M2, which were considered closely associated with the nature of pleural effusion (BPE or MPE), varied in different patients. The difference in features between high-importance pleural effusion areas and other pleural effusion areas requires further research.

DISCUSSION

In this study, we proposed a new architecture for the differential diagnosis of BPE and MPE based on pleural effusion segmentation of thoracic CT images. This deep learning architecture was trained using 607 CT images, and its performance was validated in an internal testing cohort (311 pleural effusion cases) and an external testing cohort (362 pleural effusion cases) from Wuhan Union Hospital and Renmin Hospital of Wuhan University. The encouraging diagnostic performance of the deep learning model was shown in both the internal (AUC 0.883, 95% CI 0.841 to 0.916) and external (AUC 0.842, 95% CI 0.801 to 0.878) testing cohorts. In addition, we combined this AI model with some clinical data, including gender, age, unilateral/bilateral pleural effusion and volume of pleural effusion, to predict BPE and MPE. The results showed that combining clinical indicators could improve the AUC in all three cohort (training cohort: 0.903 vs 0.896, internal testing cohort: 0.895 vs 0.882, external testing cohort: 0.868 vs 0.842) (online supplemental figure S1). This deep learning model discovered suspect pleural effusion areas and produced fine segmentations in the first step, then identified BPE and MPE by holistically and partially analysing thoracic CT image features, revealing that the features of thoracic CT images were closely related to the nature of pleural effusions. Our study provides an alternative, easy-to-use method to achieve non-invasive and efficient diagnosis of BPE and MPE from original CT images without human assistance.

Previous studies have demonstrated that thoracic CT image features, such as fluid loculation, pleural lesions, pleural nodules and extrapleural fat, can help discriminate MPEs from BPEs.18 19 Pleural nodules and nodular pleural thickening were reported to be associated to MPE, while circumferential pleural thickening was more common in tuberculous pleural effusion (TPE).18 20 Zhang et al revealed that spectral CT imaging features combined with patient age and disease history could differentiate BPEs from MPEs with a
Pleural disease

sensitivity of 100% and a specificity of 71.4%, as well as an AUC of 0.933.21

Plural effusions can be divided into transudates and exudates. Although no CT feature can accurately distinguish transudates from exudates, Abramowitz et al indicated that fluid loculation and pleural thickening were more common in exudates than in transudates.22 Discrimination of a pleural effusion as transudate or exudate is important for further evaluation and treatment. Some causes of BPE, such as heart failure and cirrhosis, generate transudates. However, some causes of BPE, such as infections and pulmonary embolism, generate exudates, as does MPE.3 23

The immunological microenvironment and inflammatory responses in MPE are two important factors that lead to the production of different components. In addition to neoplastic cells, cytokines and chemokines produced by immune cells, signalling molecules generated by tumour-associated macrophages, and fibroblasts are the main components of the surviving environment of tumour cells in pleural effusions.24 In early-stage TPE, lymphocyte predominance characterises a large proportion of the fluid; in the meantime, a higher mycobacterial burden appears in effusions that have localisations.25 Li et al identified different peptide profiles between BPE and MPE through proteomic analysis and established a model to discriminate between BPE and MPE.26 The different pleural effusion components for BPE and MPE may be a crucial cause of different thoracic CT image features, making it feasible to classify BPE and MPE using a deep learning model based on thoracic CT image features.

In the present study, we proposed an AI model to implement the segmentation of pleural effusion regions. The deep learning model for segmentation proposed in our study successfully integrated 3D spatially weighted U-Net and 2D classical U-Net. Our results showed that the cascaded segmentation architecture combining 3D spatially weighted U-Net with 2D classical U-Net (M1) was superior to the other two segmentation methods (only 3D U-Net and only 3D spatially weighted U-Net). Applying the spatial attention mechanism to 3D U-Net not only focuses the deep learning model on the regions of interest for input thoracic CT images, avoiding the interference of background information, but also extracts both shallow-level and deep-level attention information, which can improve the feature extraction ability of the model.15 27 However, in order to reduce the cubically growing number of network parameters caused by 3D convolution, using patches (crop region of interest into small image patches) as input may lose some natural contour information. 3D spatially weighted U-Net cascaded by a 2D classical U-Net can be conducive to supplementing natural contour information and excluding most error information about the pleural effusion region. In addition, in the deep learning model for pleural effusion classification, we input the holistic thoracic CT image and the fine segmentation region of pleural effusion generated by M1 at the same time. On the one hand, this approach stresses the features within the pleural effusion areas; while on the other hand, it does not neglect the related information within the areas outside the pleural effusion.

It has been reported that the primary tumour cannot be found in approximately 10% of MPEs.28 Therefore, it is of vital importance to identify MPEs of unknown origin in a timely and non-invasive manner. It would be worthwhile to apply these segmentation and classification deep learning models to other sorts of effusions. Since the causes of ascites and abscess vary depending on the type of tumour and pathogen infection, a single AI model able to determine which type of cancer or bacteria is the reason for effusion production would represent remarkable progress. Further research and efforts are required to achieve this goal.

Although the proposed deep learning model of pleural effusion segmentation and classification showed encouraging performance, our study has several limitations. First, the data source only derived from two hospitals which may have limited the generalisability and robustness of the deep learning model. Second, high model

<table>
<thead>
<tr>
<th>Input</th>
<th>Testing cohort</th>
<th>AUC (95% CI)</th>
<th>Sensitivity % (95% CI)</th>
<th>Specificity % (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Only volume</td>
<td>Internal</td>
<td>0.830 (0.783 to 0.871)</td>
<td>67.1 (59.6 to 73.9)</td>
<td>85.4 (78.1 to 91.0)</td>
</tr>
<tr>
<td></td>
<td>External</td>
<td>0.766 (0.719 to 0.808)</td>
<td>58.0 (50.9 to 64.8)</td>
<td>82.5 (75.9 to 88.0)</td>
</tr>
<tr>
<td>Volume multiply mask</td>
<td>Internal</td>
<td>0.825 (0.778 to 0.866)</td>
<td>83.5 (77.2 to 88.7)</td>
<td>73.1 (64.6 to 80.5)</td>
</tr>
<tr>
<td></td>
<td>External</td>
<td>0.766 (0.720 to 0.808)</td>
<td>91.8 (87.2 to 95.1)</td>
<td>50.6 (42.7 to 58.4)</td>
</tr>
<tr>
<td>Volume concat mask</td>
<td>Internal</td>
<td>0.883 (0.841 to 0.916)</td>
<td>78.4 (71.6 to 84.2)</td>
<td>86.2 (79.0 to 91.6)</td>
</tr>
<tr>
<td></td>
<td>External</td>
<td>0.842 (0.801 to 0.878)</td>
<td>89.4 (84.4 to 93.2)</td>
<td>65.1 (57.3 to 72.3)</td>
</tr>
</tbody>
</table>

Data are presented as % (95% CI).

AUC, area under the receiver operating characteristic curve.

Figure 5 Comparison of the activation heatmaps generated by the pleural effusion classification model between four randomly selected patients with lung cancer, breast cancer, tuberculous pleuritis and heart failure, respectively.
interpretability of deep learning networks is considered valuable, but the association between the imaging representations and the nature feature of pleural effusions cannot be fully understood in our study because of the end-to-end learning strategy. Third, despite the advantages of the proposed model, which uses exclusively thoracic CT images, in terms of convenience and time-saving, the predictive performance may be improved by combining this model with other clinical models; however, this point was not clarified in this study. Future large-scale external validations from multiple centres are necessary to provide convincing evidence of the generalisability of the deep learning model proposed in this study.

In conclusion, our research proposed an original deep learning model: a combination of 3D spatially weighted U-Net and 2D classical U-Net were used for the segmentation of pleural effusion. Subsequently, a deep learning model was established for the differential diagnosis of BPE and MPE based on thoracic CT images with masks. The non-invasiveness and high efficiency of the segmentation and classification models suggest their potential clinical utility. Our work shows the potential of AI to assist radiologists in identifying malignant disease and thereby improving patient care.

Author affiliations
1Department of Respiratory and Critical Care Medicine, NHRC Key Laboratory of Pulmonary Diseases, Wuhan Union Hospital, Wuhan, Hubei, China
2State Key Laboratory of Magnetic Resonance and Atomic and Molecular Physics, Chinese Academy of Sciences Wuhan Institute of Physics and Mathematics, Wuhan, Hubei, China
3Department of Radiology, Wuhan Union Hospital, Wuhan, Hubei, China
4Department of Infectious Diseases, Wuhan Union Hospital, Wuhan, Hubei, China
5Department of Pathology, Renmin Hospital of Wuhan University, Wuhan, Hubei, China
6Department of Gastroenterology, Renmin Hospital of Wuhan University, Wuhan, Hubei, China
7Department of Radiology, Renmin Hospital of Wuhan University, Wuhan, Hubei, China
8School of Public Health, Tongji Medical College of Huazhong University of Science and Technology, Wuhan, Hubei, China
9Department of Pathology, Wuhan Union Hospital, Wuhan, Hubei, China

Acknowledgements We are grateful to Dr Liazheng Wang (Philips Healthcare), Dr Peng Sun (Philips Healthcare) and Mr Yuyang Chen (Putnam Science Academy, USA) for many useful discussions through the formation of this work. We would like to thank Editage (www.editage.cn) for English language editing.

Contributors SW: conceptualisation, project administration, writing—original draft. XT: investigation, methodology, writing—original draft. PL: formal analysis, methodology, software, writing—original draft. QF: methodology, data curation. YR: data curation, methodology, LX: data curation, methodology, YD: data curation, methodology, investigation, methodology, JX: funding acquisition, supervision. YM: data curation, supervision. WC: data curation, methodology. YL: investigation, supervision. ZZ: methodology. CT: methodology, supervision. QB: methodology, supervision, writing—review and editing. YT: data curation, methodology, supervision. YJ: conceptualisation, funding acquisition, supervision, writing—review and editing. YJ is the guarantor of this publication.

Funding This work was supported by the National Natural Science Foundation of China (Nos. 81770096; No. 82102496; No. 82172034).

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval The institutional ethics committees of Wuhan Union Hospital (No. S1041) and Renmin Hospital of Wuhan University (No.WDRY2019-K014) reviewed and approved this study protocol and waived the need for informed consent due to the study design.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iD
Lian Yang http://orcid.org/0000-0003-4664-2474

REFERENCES