Supplementary Methods

Diagnostic criteria for malignant pleural effusion (MPE) and benign pleural effusion (BPE)

According to the criteria adopted in our previous study[1, 2], if malignant cells were detected in the pleural effusion based on cytologic examination or pleural biopsy, the effusion was classified as malignant. BPE was identified by a known aetiology, such as tuberculous pleural effusion (TPE) or parapneumonic effusion, without any signs of cancer. TPE was diagnosed if acid-fast stains or Lowenstein–Jensen cultures of pleural effusion, sputum, bronchoalveolar lavage fluid, or pleural biopsy specimens were positive.

Computed tomography (CT) imaging protocols

For the establishment of the training and internal testing dataset, non-enhanced chest CT data were obtained from Wuhan Union Hospital. All these chest CT examinations were performed on two commercial multi-detector CT scanners: Philips Ingenuity Core128 (Philips Medical Systems) (n=410) and SOMATOM Definition AS (Siemens Healthineers) (n=508) with the routine mediastinal window reconstruction: axial images with a matrix size of 512 × 512, slice thickness of 5 mm; and mediastinal kernels of iDose5 in Philips Ingenuity Core128 and b30f in SOMATOM Definition AS. Before the scanning, patients were instructed on breath-holding in order to minimize motion artifacts. Afterwards, CT images were acquired during a single breath-hold. For external validation, additional chest CT data were obtained from a third-party hospital (Renmin Hospital of Wuhan University). These routine non-contrast chest CT scans were performed on four commercial multi-detector CT scanners: GE Optima CT680 CT (n=155), 64-slice LightSpeed VCT (n=95), BrightSpeed Elite CT (n=89) and Revolution CT (n=23; GE Medical Systems). The standard reconstruction of axial mediastinal-window images was implemented following similar commercial scanning protocols from the CT manufacturers[3].

Model training

The specific training process is as follows:

First, we train the coarse segmentation component on the dataset to generate coarse pleural effusion masks. Then, we concatenate the CT images with the corresponding coarse pleural effusion masks generated by the coarse segmentation component as the input to the fine segmentation component. We train the fine segmentation component to generate the fine pleural effusion masks. Finally, we fuse the CT images and the corresponding fine pleural effusion masks refined by the fine segmentation component and feed them to the classification component (M2) to train the classification component to deliver more accurate diagnosis results.

Mathematical description of the pleural effusion classification model

SE block

SE block is derived from SENet, which mainly learns the correlation between channels and filters out the attention for the channels, and slightly increases the computational effort, but the result of it is better. By processing the feature map from the convolution, a one-dimensional vector with the same number of channels is obtained as the evaluation score of each channel, and then the score is applied to the corresponding channel separately to get its result. It contains three main operations as follows:

Squeeze: Compressing the features along the spatial dimension, turning each two-dimensional feature channel into a real number, which somehow has a global perceptual field, and the output dimension matches the number of input feature channels.

Excitation: Based on the correlation between the feature channels, a weight is generated for each feature channel, which represents the importance of the feature channels.

Reweight: The weights from Excitation are considered as the importance of each feature channel, and then weighted to the previous features by multiplying them channel by channel to complete the rescaling of the original features in the channel dimension.

Details of the pleural effusion segmentation model

Coarse segmentation model

The coarse segmentation model plays a major role in the generation of coarse pleural effusion segmentation masks. The architecture of this component is based on a 3D spatially weighted U-Net. The 3D U-Net is commonly used for medical image segmentation tasks because of its multiscale feature fusion capability. At the same time, we utilized a 3D spatial attention mechanism to capture large-scale contextual information, thus enhancing the representative ability of the model.

The 3D attention layer is placed before and after the concatenation operation to fully exploit the spatial contextual information on the intra-plane level and leverage it for volumetric spatial weighting on the inter-plane level. This emphasizes the regions of interest in volumetric feature maps.

Under the 3D convolutional network architecture, we assume that the volumetric feature tensors Fi input to the 3D attention layer in the i-th layer is of size I×J×K×P, where I, J, and K are the length, width, and height of a feature tensor, respectively, and Fi contains P channels. To acquire the spatial statistical information representation on a chosen plane, we applied global average pooling (GAP) to each plane of the 3D space. The three resulting orthogonal vectors compress the statistical information in the entire slice along each plane. Moreover, we also adopt fully connected (FC) layers, and rectified linear unit (ReLU) and sigmoid activation functions, to introduce additional nonlinearity in the generation of the weight vectors. The structure is similar to a bottleneck architecture with two fully connected layers. The first fully connected layer is used for dimensionality reduction, with a reduction ratio of 1/4 to limit capacity and improve model generalization. With the weighting vectors for each plane, we weighted the three dimensions of the feature tensors Fi. Mathematically, if the feature tensor is weighted along the three planes, this is equivalent to a tensor product of the orthogonal weighting vectors.

Fine segmentation model

The input of this component contains cropped CT image patches and coarse pleural effusion masks. As aforementioned, 3D spatially weighted U-Nets can capture large-scale contextual information. However, because of GPU memory limits, the input

cannot cover the complete CT images, and the coarse segmentation model loses contour information during the learning process. Thus, we make use of a 2D classical U-Net to enhance the learning of natural contour information so that the precision of fine segmentation can be improved.

Details of the pleural effusion classification model

The classification model is similar to the 3D ResNet but with several modifications. In this model, a 3D convolutional layer, a 3D pooling layer and a 3D SE block is defined as a group. Skip connection is added to input and output of each group. Two stacks of groups with a batch normalization layer is defined as a SE-ResBlock. First, the input will pass through a 3D convolutional layer, a 3D BN layer and a 3D pooling layer, then through four stacks of SE-ResBlocks for feature extraction, and finally, through the GAP and FC layers to predict the probability of MPE.

Details of the quantitative assessment indicators

For the pleural effusion segmentation model, the Dice similarity coefficient (DSC) (1) and Jaccard coefficient (2) were used to evaluate the spatial overlap between the model-generated contour (M) and the ground truth contour (G):

$$DSC(M,G) = \frac{2|M \cap G|}{|M|+|G|}, \quad (1)$$

$$Jaccard = \frac{|M \cap G|}{|M \cup G|}, \quad (2)$$

Precision (3) and sensitivity (4) measure the detection capability for identifying the correct regions.

$$Precision = \frac{TP}{TP+FP}, \quad (3)$$

$$Sensitivity = \frac{TP}{TP+FN}, \quad (4)$$

where true positives (TP) are defined as the regions of the segmentation results consistent with the ground truth, and false positives (FP) as the regions of the segmentation results that are not consistent with the ground truth. False negatives (FN) are defined as the ground truth regions that are not included in the segmentation results.

The Hausdorff distance 95% (HD95) (5) and average surface distance (ASD) (6) measure the boundary similarity between the model-generated contour and the ground truth contour:

$$HD95 = max_{k95\%}[d(G,M), d(M,G)] = max_{k95\%}[\max_{g\in G}\min_{m\in M} d\{g,m\}, \max_{m\in M}\min_{g\in G} d\{m,g\}], \quad (5)$$

$$ASD = \frac{1}{S(G)+S(M)}\left(\sum_{g\in S(G)} d\left(g,\ S(M)\right) + \sum_{m\in S(M)} d\left(m, S(G)\right)\right), \quad (6)$$

where g represents points in G and m represents points in M. $d(g,m)$ represents the distance between point g and point m. $S(M)$ and $S(G)$ represent the surfaces of M and G, respectively. $d\left(g,\ S(M)\right)$ and $d\left(m, S(G)\right)$ represent the shortest distance from any point g to $S(M)$ and from any point m to $S(G)$, respectively.

Implementation

The Adam algorithm with a batch size of 16, β1 = 0.9, β2 = 0.999, and decay = 1e-6 was adopted to optimize the segmentation and classification models[4]. The initial learning rate for both the segmentation and classification models was set at 0.001. The weights of the networks were initialized using the default initialization mechanism of the Keras framework. All experiments were performed in the Tensorflow and Keras framework. The training strategies were optimized in the same computer system with 32 GB RAM and a GeForce GTX 2080 graphics processing unit (GPU).

Compute performance measures on an hourly basis

First step: Convert dicom data to nii data for CT images, use python code to extract window width and window level, and convert thin layer (thickness of <5mm) to thick layer (thickness of 5mm) (if necessary).

Second step: Input all the thick layer CT image data into the trained 3D spatially weighted U-Net to generate the coarse segmentation results. This step takes about 50 minutes to perform coarse segmentation for 104 patients.

Third step: Input the nii data and the coarse segmentation results for CT images into the trained 2D classical U-Net to generate the fine segmentation results. This step takes

about 10 minutes to perform fine segmentation for 104 patients.

Fourth step: Input the CT volume and its 3D pleural effusion fine masks into the classification model to generate the classification scores. This step takes about 44 seconds to perform classification for 104 patients.

Excluding the intermediate data processing steps, it takes about an hour to process 104 patients through the segmentation and classification model.

Reference

1. Wang S, Tian S, Li Y, et al. Development and validation of a novel scoring system developed from a nomogram to identify malignant pleural effusion. EBioMedicine 2020;58:102924.

2. Luo P, Mao K, Xu J, et al. Metabolic characteristics of large and small extracellular vesicles from pleural effusion reveal biomarker candidates for the diagnosis of tuberculosis and malignancy. J Extracell Vesicles 2020;9(1):1790158.

3. Bhalla AS, Das A, Naranje P, et al. Imaging protocols for CT chest: A recommendation. Indian J Radiol Imaging 2019;29(3):236-246.

4. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. arXiv preprint 2014;arXiv:14126980.

Supplementary Figure

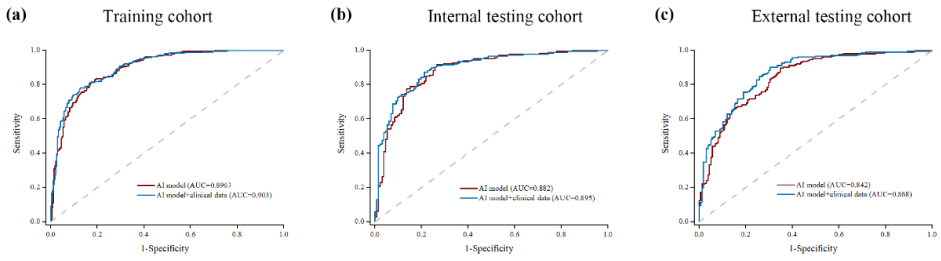

Figure S1. ROC curves of AI model and AI model combined with clinical data in the training (a), internal testing (b) and external testing (c) cohorts.