Appendix 1 methods

Hospital name

The hospitals included Central South University Xiangya School of Medicine Affiliated Haikou Hospital, Hainan Affiliated Hospital of Hainan Medical University, The First Affiliated Hospital of Hainan Medical University, The Second Affiliated Hospital of Hainan Medical University, Hainan Hospital of Traditional Chinese Medicine, Hainan Women and Children's Medical Center, Sanya Central Hospital, Sanya People's Hospital, Hainan Hospital of Chinese PLA General Hospital, Wenchang People's Hospital, and LeDong Second People's Hospital.

Preprocessing

First, the raw resolution varied from 512×512 to 726×726 pixels and from 58 to 113 slices. For resolution regularization, we rescaled the images to 512×512 in bilinear fashion. In addition, for intensity normalization, a standard lung window [width(w) =1,200 HU and level(l) =–600 HU] was used to normalize each slice to float matrix between 0 and 1 in float 32.

Postprocessing

In the postprocessing stage, we employed the morphology to refine the segmentation results. Based on the lung segmentation results, the two largest connected components that were predicted to be positive were segmented as the lung in the 3D space. Considering that the left and right lungs may be attached in segmentation, the components with fewer than 2,000 3D pixel numbers were also eliminated. The smaller regions were typically segmented due to random noise. In addition, the 3D negative areas closed by positive areas were completed according to the morphology. Typically, the 26-connect operator was used in the computing of the connected components, and the Skimage tool box was used in the computing of the 3D connected components.

Feature extraction

Area

From the raw DICOM file of chest CT, we obtained the number, the real physical sizes of pixels in the HU arrays, and slice thickness. By multiplying the pixel numbers and the real physical volumes of the pixels, the involvement of the lesions could be easily obtained. The features of involvement included the real physical volume and the rate of lung volume.

Amount

Based on the lesions annotated and the 3D expression, we counted the numbers of each type of lesion to express the distribution of the lesions in the scans.

Intensity

The intensity of each annotated lesion was described by mathematical evaluation values of the HU values. The evaluated values contained the max, mean, and median. As the cavity contained air and noise, the min value of HU was not considered. Typically, each of the annotated lesions was evaluated in 3D space. The max HU features were obtained from the smoothed HU arrays using a 3×3 mean filter to control random imaging noise.

The rate of volume change

To evaluate disease progression, the rates of change in the volume characteristics were calculated because this feature was the most variable among all the features. The rate was defined as the difference divided by the time between two scans of the same patient. If there was only one CT scan, the time was defined as the duration of onset, and we assumed that there was no visible lesion on the onset CT. Therefore, the CT lesion value at the initial onset time was set to zero. The distribution of

some features within the timeline is shown in *Figure S2*.

Supplementary result

Comparative experimental

Using clinical knowledge, we selected 12 image features and obtained the importance of the 12 features by recalculating. We also compared the classification accuracy of the 12 most important features that were automatically selected with our method's 12 selected features. We obtained a 73.73% accuracy on all 24 features, a 76.37% accuracy on the automatically picked features, and a 78.73% accuracy on our selected features. Therefore, the experimental results were consistent with the limitations of the random forest algorithm reported in the literature (32).

Additionally, we compared the CT features between different groups (severe death *vs* severe survival, severe death *vs*. mild survival). We found significant differences in the lesions ratio and consolidation ratio between severe death and mild survival in the CT scans, while there were no apparent differences between severe death and severe survival (*Table S1*).

References

32. Probst P, Wright MN, Boulesteix AL. Hyperparameters and tuning strategies for random forest. WIREs Data Mining and Knowledge Discovery 2019;9:e1301.
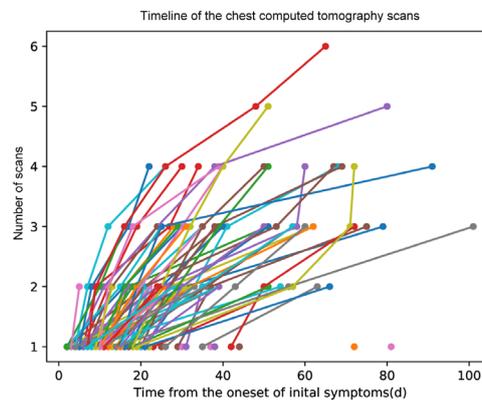
Figure S1 Timeline of the chest computed tomography scans. Colors represent different patients, while dots on the line indicate multiple CT scans from the same patient. d, days; CT, computed tomography.
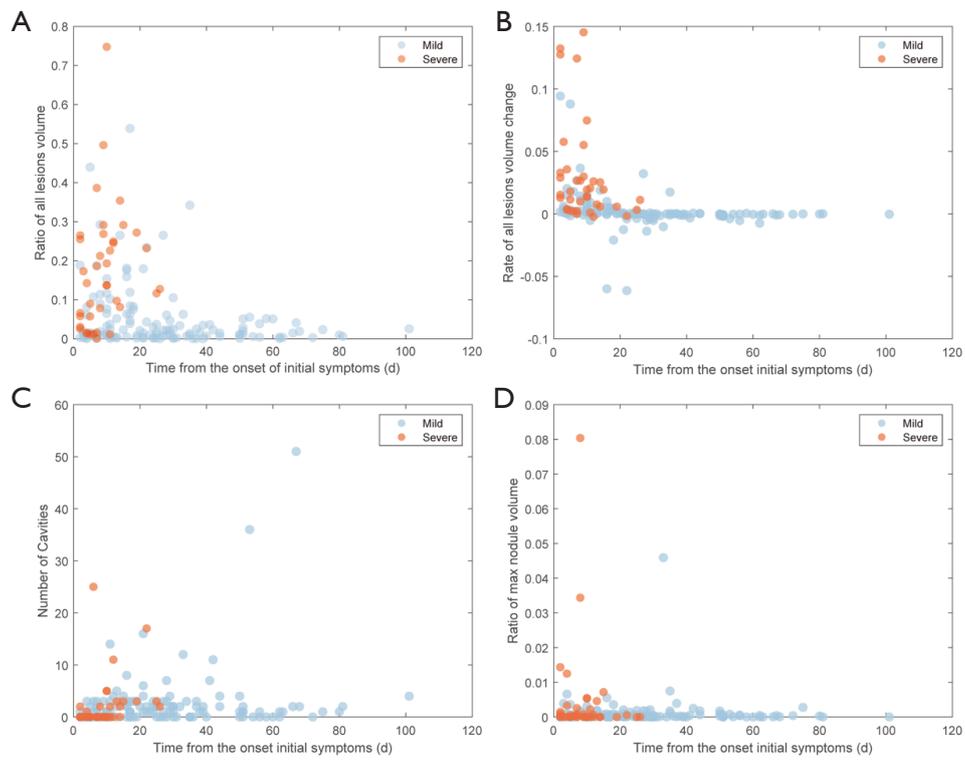
Figure S2 The distribution of a few features within the timeline.
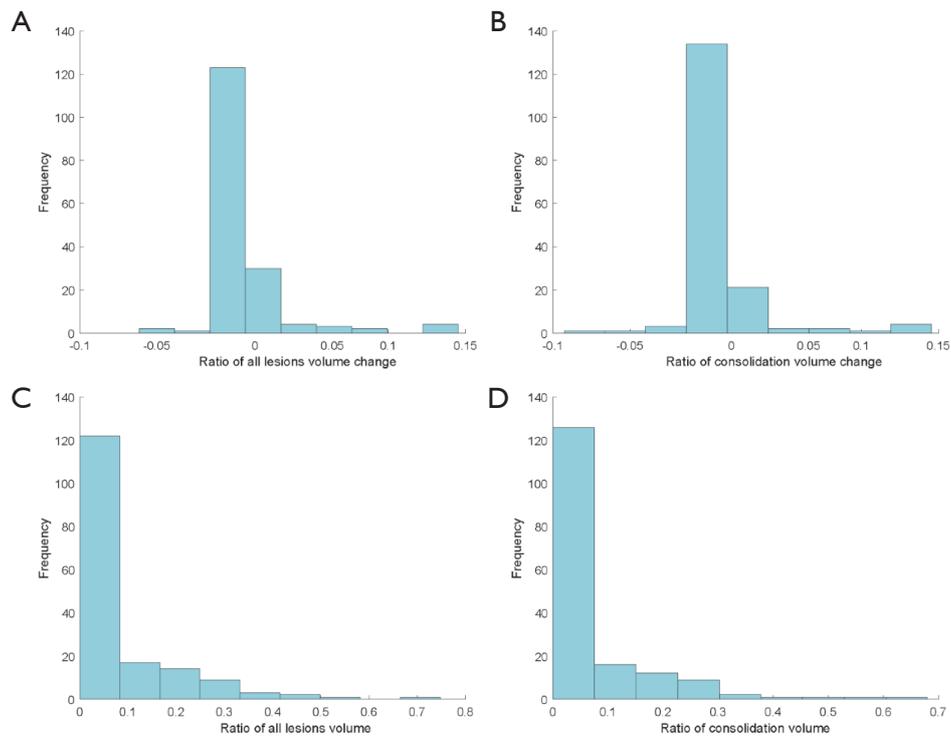
Figure S3 Bar graph showing the histograms of computed tomography features.
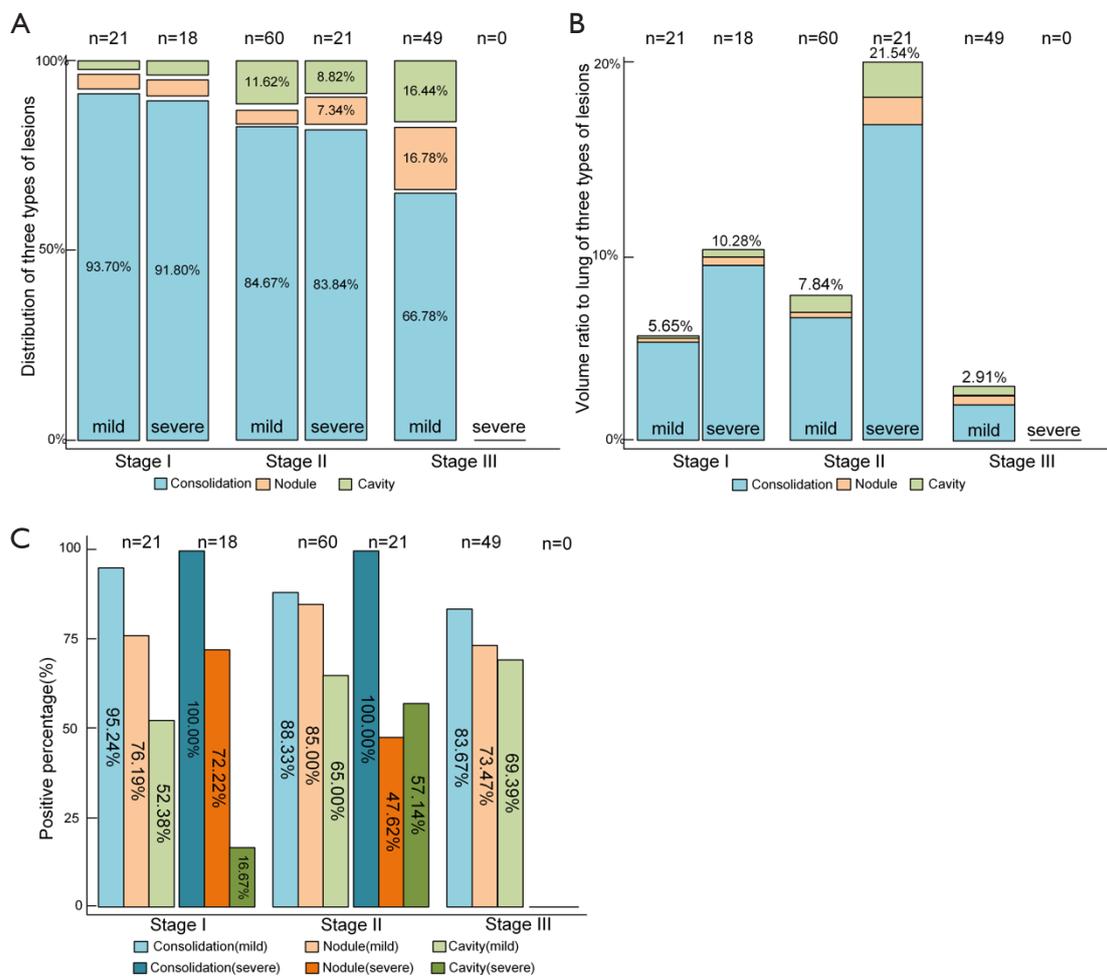
Figure S4 Bar graph showing the changes in the distribution of the three lesions at different stages. (A) The composition ratio of the three lesions in different stages. (B) The absolute ratio of the three lesions in the lung and the ratio of volume averaged by the number of scans. (C) Frequency of CTs with nodules and cavities at the different stages. Note that there are significant differences among the stages and the severe and mild scans. n indicates the number of scans. CT, computed tomography.

Table S1 Comparison of computed tomography features in patients with different outcomes

Features	Severe death (N=6)	Severe survived (N=24)	P	Severe death (N=6)	Mild survived (N=67)	P
Lesions ratio	0.160±0.29	0.163±0.15	0.967	0.160±0.29	0.056±0.10	0.025
Consolidation ratio	0.147±0.27	0.141±0.14	0.930	0.147±0.09	0.047±0.09	0.018
Nodules ratio	0.013±0.03	0.011±0.03	0.860	0.013±0.14	0.004±0.14	0.142
Num of cavities	1.000±1.38	1.385±4.08	0.822	1.000±1.55	2.331±5.71	0.571