Supplementary

Table S1 CT protocols of the four hospitals

Parameters	Hospital I		Hospital II	Hospital III	Hospital IV
CT scanners	Smoatom Definition AS, Siemens	Smoatom Definition Flash, Siemens	Smoatom Definition AS, Siemens	GE Revolution, GE	Uct 710, United Shadow Medical
Tube voltage (kV)	100–120	100–120	100	100–120	120
Tube current	Automatic tube-current	Automatic tube-current	Automatic tube-current	Automatic tube-current	Automatic tube-current
FOV (mm)	350–400	350–400	350–400	350–400	350–400
Gantry rotation time (s)	0.6	0.5	0.28	0.28	0.33
Detector collimation (mm)	128×0.625	128×0.625	128×0.625	128×0.6	64×0.6
Section thickness (mm)	5	5	5	5	5
Section interval (mm)	5	5	5	5	5
Image matrix	512×512	512×512	512×512	512×512	512×512

Hospital I, Shunde Hospital, Southern Medical University (The First People's Hospital of Shunde); hospital II, Xingtan Hospital Affiliated to Shunde Hospital of Southern Medical University; hospital III, The Third Xiangya Hospital of Central South University; hospital IV, Chenzhou First People's Hospital; FOV, field of view; CT, computed tomography.

Appendix 1 A brief description of six machine learning (ML) classifiers

Logistic regression (LR) analysis is a statistical method used to analyze a dataset in which one or more independent variables determine the outcome. The outcome was measured using a dichotomous variable in which there were only two possible outcomes.

Extreme gradient boosting (XGBoost) is an ensemble learning model based on a gradient boosting algorithm. It builds a strong classifier by iteratively training multiple weak classifiers and optimizing the loss function using a gradient descent. XGBoost performs well in processing structured and large-scale datasets with high efficiency and accuracy.

Support vector machine (SVM) is a ML approach based on the structural risk minimization principle of statistical learning. It projects data onto a multidimensional space and classifies them into hyperplanes.

Random forest (RF) is an ensemble ML method for classification and regression that operates by constructing a large number of decision trees (DTs) and outputting classes as a single tree (classification) or average prediction (regression) model.

A DT is a basic classification and regression method that can generate multibranch DTs, and the target variables are categorical.

K-nearest neighbor (KNN) is an instance-based learning algorithm. It classifies new samples into the most common category among the KNNs by measuring the distance between samples. The KNN algorithm is simple, easy to understand, and suitable for multiclass classification and nonlinear decision boundary problems.

Artificial neural network (ANN) is a computational model based on the structure and functions of a biological neural network. The main advantage of an ANN is its ability to approximate nonlinear mathematical functions.

https://dx.doi.org/10.21037/qims-2025-887