Supplementary

Table S1 The selected ranges of the evaluated hyperparameters in the machine learning methods

Method	Hyperparameters
Naïve Bayes	Using GaussianNB(), not hyperparameters settin
Decision tree	max_depth=[50, 100, 150, 200), min_samples_leaf=[5,10,15), criterion=['entropy')
k-nearest neighbors	n_neighbors=[6,7,8,9), P=[2,3)
Logistic regression	penalty=['l1', 'l2'), C=[0.001,0.01, 0.1), tol=[1e-3,1e-4), max_iter=[300,400,500),solver=['liblinear')
Multilayer perceptron	hidden_layer_sizes=[(3,),(10,)) solver=['sgd') learning_rate_init=[0.01, 0.1, 0.5) learning_rate=['adaptive') max_iter=[100,200,300) tol=[1e-2,1e-3)
Random forest	n_estimators=[100,200,300) max_depth=[50, 100)
Support vector machine	kernel=['linear','sigmoid') probability=[True) C=[0.01,0.1,10)
Gradient boosting decision tree	objective=['binary: logistic') n_estimators=[50,500) max_depth=[3,5,7,9) learning_rate=[0.05,0.1,0.5) eval_metirc=['auc')

The parameter random_state in the above methods was set to 7052.

Table S2 Hyperparameter setting of the selected gradient boosting decision trees

Hyperparameter	Setting
Objective	binary: logistic
n_estimators	500
max_depth	5
learning_rate	0.5
eval_metric	"auc"
random_state	7052

Note: Other hyperparameters were the default values in scikit-learn package.

https://dx.doi.org/10.21037/qims-22-1059