

Supplementary table1 The result of different algorithm in initial training

Model	Mean AUC	AUC 95% CI Lower	AUC 95% CI Upper	Mean Accuracy	Sensitivity	Specificity	PPV	NPV
Logistic Regression	0.931	0.889	0.972	0.874	0.975	0.99	0.975	0.99
Random Forest	0.943	0.905	0.981	0.867	1	1	1	1
Gradient Boosting	0.955	0.921	0.989	0.840	1	1	1	1
Support Vector Classifier	0.959	0.927	0.992	0.882	0.975	1	1	0.99
Decision Tree	0.776	0.708	0.844	0.770	1	1	1	1
K-Nearest Neighbors	0.873	0.819	0.928	0.803	0.725	0.99	0.966	0.902
Naive Bayes	0.860	0.804	0.917	0.811	0.775	0.912	0.775	0.912
XGBoost	0.938	0.899	0.978	0.861	1	1	1	1
LightGBM	0.919	0.875	0.964	0.819	1	1	1	1
Ridge Classifier	0.865	0.801	0.924	0.895	0.9	0.98	0.947	0.961
Extra Trees	0.957	0.925	0.990	0.902	1	1	1	1
AdaBoost	0.919	0.874	0.963	0.861	1	1	1	1
Voting Classifier	0.945	0.908	0.983	0.846	1	1	1	1

AUC: Area Under Curve. CI: Confidence Interval, PPV: Positive Predict Value, NPV: Negative Predict Value