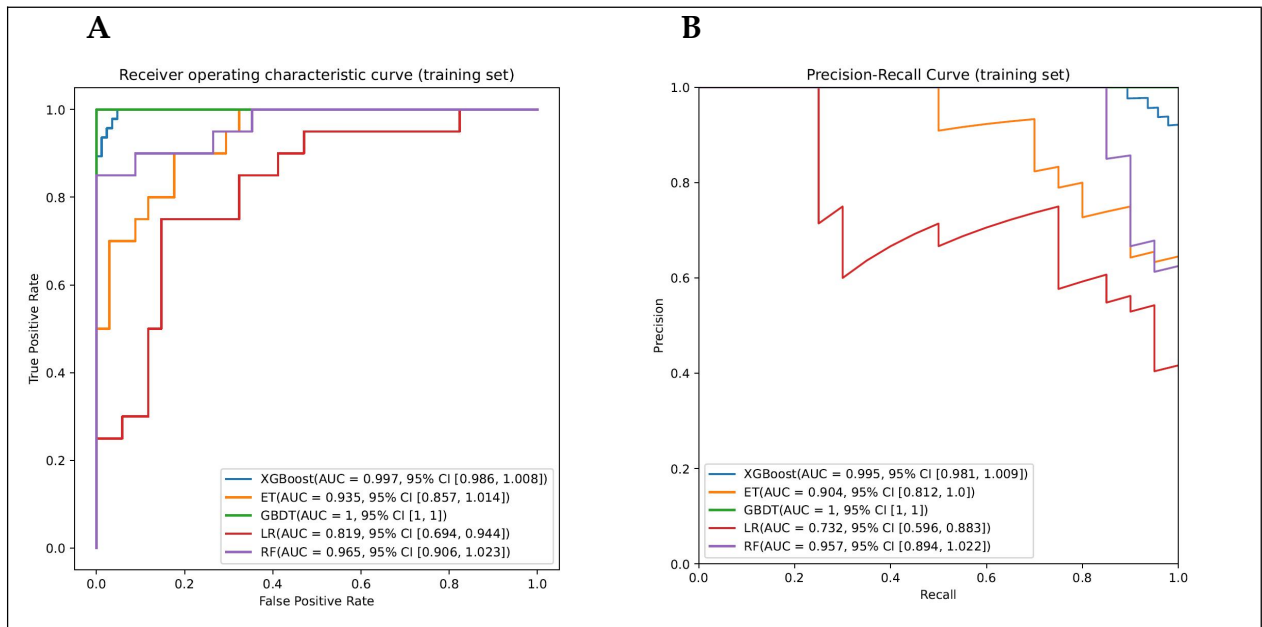
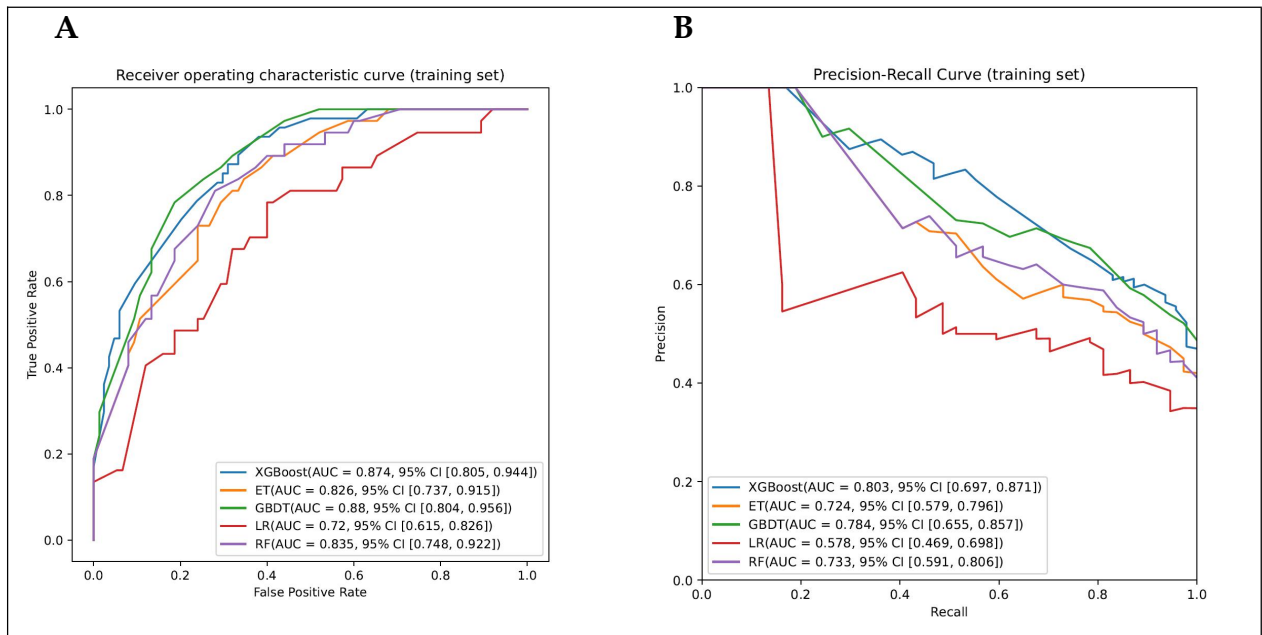


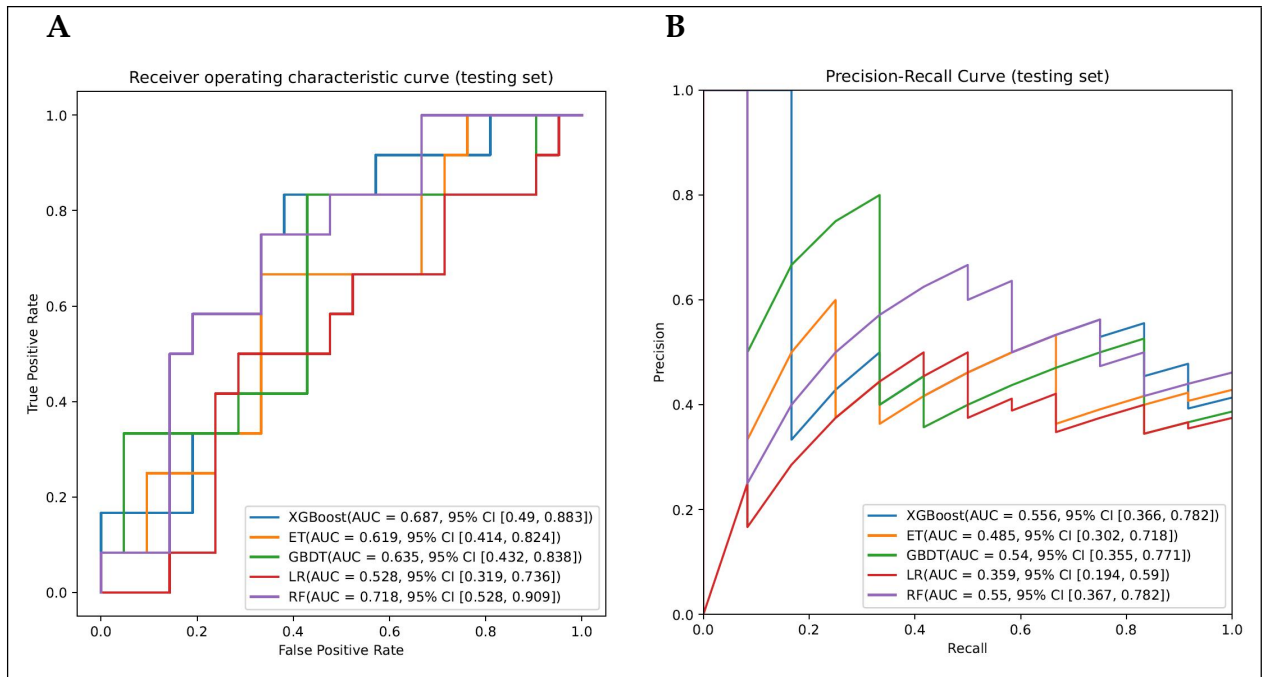
Supplementary Figure 1 Evaluation of the predictive models. A, B The picture shows average area under the receiver operating characteristic curve and precision recall curve of the five models inputting 18 clinical features in the training set. Average area and 95% confidence intervals (CIs) of different predictive models are displayed in the box. XGBoost: Extreme gradient boosting gradient; ET: Extremely random trees; GBDT: Gradient boosting decision tree; LR: Logistic regression; RF: Random forest.



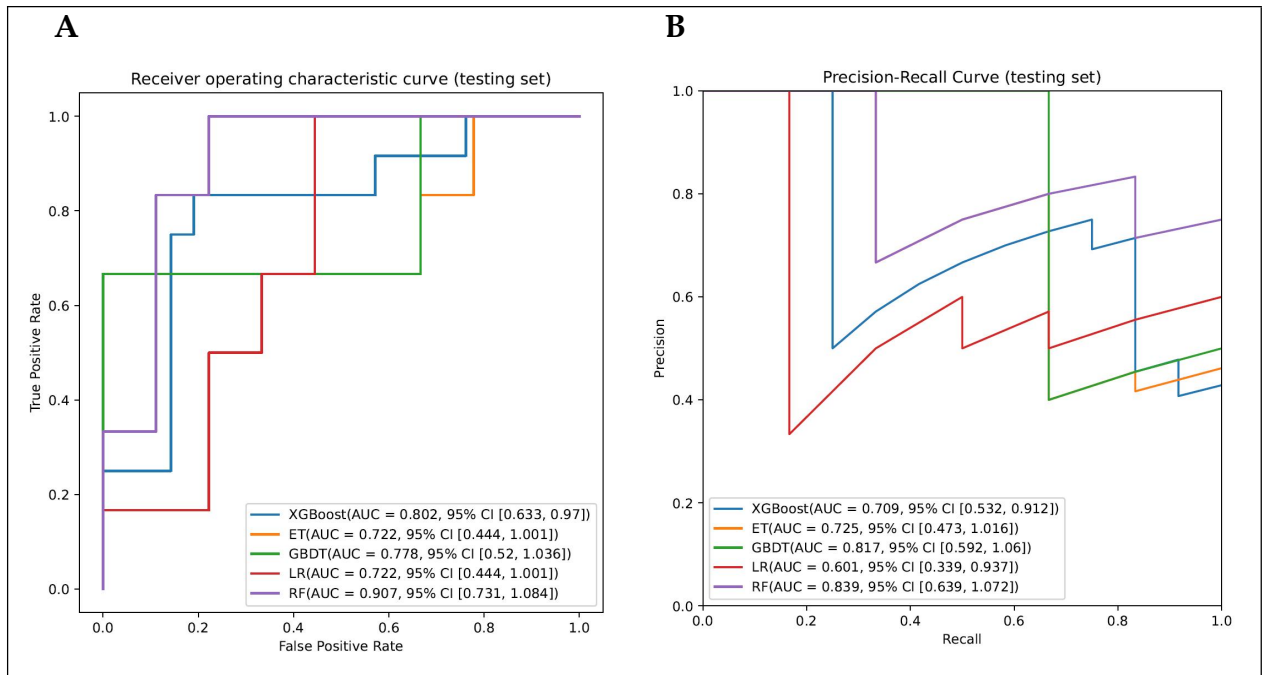
Supplementary Figure 2 Evaluation of the predictive models. A, B The picture shows average area under the receiver operating characteristic curve and precision recall curve of the five models inputting 14 SNPs in the training set. Average area and 95% confidence intervals (CIs) of different predictive models are displayed in the box. XGBoost: Extreme gradient boosting gradient; ET: Extremely random trees; GBDT: Gradient boosting decision tree; LR: Logistic regression; RF: Random forest.



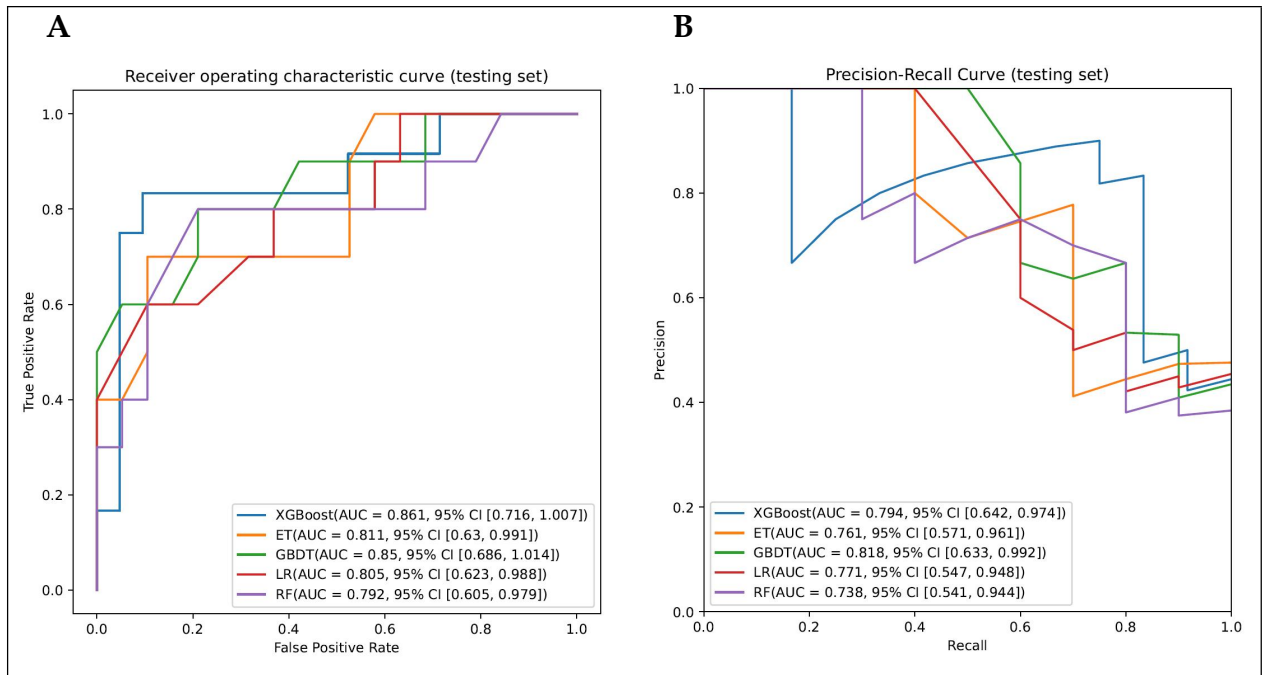
Supplementary Figure 3 Evaluation of the predictive models. A, B The picture shows average area under the receiver operating characteristic curve and precision recall curve of the five models inputting top five ranked features in the training set. Average area and 95% confidence intervals (CIs) of different predictive models are displayed in the box. XGBoost: Extreme gradient boosting gradient; ET: Extremely random trees; GBDT: Gradient boosting decision tree; LR: Logistic regression; RF: Random forest.



Supplementary Figure 4 Validation of the training set. A, B The picture shows average area under the receiver operating characteristic curve and precision recall curve of the five models inputting 18 clinical features in the test set. AUC: Area under the curve; CI: Confidence interval; XGBoost: Extreme gradient boosting gradient; ET: Extremely random trees; GBDT: Gradient boosting decision tree; LR: Logistic regression; RF: Random forest.



Supplementary Figure 5 Validation of the training set. A, B The picture shows average area under the receiver operating characteristic curve and precision recall curve of the five models inputting 14 SNPs in the test set. AUC: Area under the curve; CI: Confidence interval; XGBoost: Extreme gradient boosting gradient; ET: Extremely random trees; GBDT: Gradient boosting decision tree; LR: Logistic regression; RF: Random forest.



Supplementary Figure 6 Validation of the training set. A, B The picture shows average area under the receiver operating characteristic curve and precision recall curve of the five models inputting the top five features in the test set. AUC: Area under the curve; CI: Confidence interval; XGBoost: Extreme gradient boosting gradient; ET: Extremely random trees; GBDT: Gradient boosting decision tree; LR: Logistic regression; RF: Random forest.

Supplementary Table 1 Patients characteristics of clinical variable

Demographic	Values (n = 164)
Females	46, (26.4)
Age, mean \pm SD	34.3 \pm 12.7
Gender (male/female)	119/45
Daily dose (mg)	77.7 (25-200)
ALT (U/L)	22.5 \pm 17.9
AST(U/L)	20.3 \pm 9.3
Hb (%)	123.0 \pm 22.9
BUN (mg/dl)	4.2 \pm 1.4
CRE (mg/dl)	63.1 \pm 16.5
NEUTP (%)	0.5 \pm 0.1
PLT (10e9/L)	278.8 \pm 103.6
HCT (%)	2.5 \pm 26.9
CRP (mg/L)	9.1 \pm 15.4
WBC (10e9/L)	1.3 \pm .29
LP(a)(mg/L)	294.1 \pm 318.8
HDL-c(mmol/L)	1.2 \pm .41
LDL-c(mmol/L)	2.34 \pm .78
TG (mmol/L)	1.4 \pm 2.0

ALT: Alanine aminotransferase levels before toxicity; AST: Aspartate aminotransferase levels before toxicity; Hb: Hemoglobin; BUN: blood urea nitrogen; CRE: Creatinine; NEUTP: neutrophil; PLT: Platelet count before toxicity; HCT: Red blood cell specific volume; TG: Triglyceride levels before toxicity; CRP: C-reactive protein; WBC: white blood cell; LP: Lipoprotein; HDL-c: Hardware Description Language Cholesterol; LDL-c: Low-Density Lipoprotein Cholesterol; TG: Triglyceride.

Supplementary Table 2 All single nucleotide polymorphisms included in this study

Genes of pharmacokinetic pathway-related proteins		Genes of transcriptional regulation proteins		Genes of disease related proteins		CD related Genes of TiPN related proteins	
Gene	SNPs	Gene	SNPs	Gene	SNPs	Gene	SNPs
CYP2C1 9	rs1152808 7	ABCA1	rs2777795	VEGF A	rs3025020	IL-12	rs1353248
CYP2C1 9	rs4244285	ABCA1	rs2575876	VEGF A	rs3025011	IL-12	rs571099
CYP2C1 9	rs8014735 9	ABCA1	rs6256603 2	VEGF A	rs3025009	IL-12	rs609907
CYP3A4	rs3735451	ABCA1	rs4149275	VEGF A	rs62401162	IL-12	rs2279742
CYP3A4	rs1267085 0	ABCA1	rs2297406	VEGF A	rs3025018	IL-12	rs4680558
CYP3A4	rs3556427 7	ABCA1	rs1099141 9	VEGF A	rs3025040	IL-12	rs4680564
CYP3A4	rs2246709	ABCA1	rs4149261	CRBN	rs77784443	IL-12	rs1428285
CYP3A4	rs4646440	ABCA1	rs4149282	CRBN	rs4685611	IL-12	rs7826869 7
		ABCA1	rs2254708	CRBN	rs1705814	IL-12	rs1721710 2
		ABCA1	rs4149287	CRBN	rs14309127 8	IL-12	rs7890293 1
		ABCA1	rs7857983	CRBN	rs2306767	IL-12	rs2243138
		ABCA1	rs3758294	CRBN	rs3736223	IL-12	rs1249273 0

ABCA1	rs2740486	IKZF1	rs6583442	IL-12	rs9868698
ABCA1	rs2249891	IKZF1	rs12718731	IL-12	rs6441282
ABCA1	rs3416541	IKZF1	rs11766800	IL-12	rs1263148
	9			8	
ABCA1	rs1099141	IKZF1	rs12668621	IL-12	rs1192298
	7			8	
ABCA1	rs2274871	IKZF1	rs7791054	IL-12	rs1498736
ABCA1	rs2859975	IKZF1	rs12668621	BDNF	rs7964255
	9			7	
ABCA1	rs1082074	IKZF1	rs7791054	BDNF	rs2030324
	3				
ABCA1	rs4149341			BDNF	rs1103010
					1
ABCA1	rs4149339			BDNF	rs7127507
ABCA1	rs2740486			BDNF	rs1103010
					4
ABCA1	rs2777789			BDNF	rs1103010
					0
ABCA1	rs2066716			BDNF	rs6265
ABCA1	rs1929842			MMP	rs3787268
				9	
ABCA1	rs2020926			MMP	rs3918249
				9	
ABCA1	rs363717			MMP	rs2274755
				9	
ABCB1	rs4728709			MMP	rs3918254
				9	
ABCB1	rs7795818			TRPV	rs161394
				1	

ABCB1	rs7795817	TRPV	rs9902581
		1	
ABCB1	rs7795840	TRPV	rs1695329
		1	5
ABCB1	rs7795820	TRPV	rs161374
		1	
ABCB1	rs7795848	TRPV	rs3826503
		1	
ABCB1	rs7795823	TRPV	rs161386
		1	
ABCB1	rs7795846	TRPV	rs1695316
		1	3
ABCB1	rs7795827	TRPV	rs1566790
		1	
ABCB1	rs4148751	TRPV	rs224546
		1	
ABCB1	rs7795834	TRPV	rs7719946
		1	3
ABCB1	rs7795839	TRPV	rs150908
		1	
ABCB1	rs7795847	TRPV	rs4790151
		1	
SLC12A	rs7640259	TRPV	rs1085078
6	2	4	1
SLC12A	rs4780235	TRPV	rs6194087
6		4	0
SLC12A	rs7164902	TRPV	rs6194377
6		4	4
		TRPV	rs7508775

4	8
TRPV	rs1106829
4	0
TRPV	rs4635144
4	
TRPV	rs3742033
4	
TRPV	rs1077490
4	0
TRPV	rs7608808
4	5
TRPV	rs7971845
4	

Supplementary Table 3 Performance of the models for training set (Clinical features)

Model	Precision	Sensitivity	Specificity	Accuracy	AUROC	F1
XGBoost	0.691	0.809	0.798	0.802	0.837	0.745
ET	0.431	1	0.262	0.527	0.966	0.603
GBDT	0.887	1	0.929	0.954	0.999	0.94
LR	0.439	0.915	0.345	0.55	0.759	0.593
RF	0.474	0.957	0.405	0.603	0.865	0.634

ROC: Receiver operating characteristic; XGBoost: Extreme gradient boosting gradient; ET: Extremely random trees; GBDT: Gradient boosting decision tree; LR: Logistic regression; RF: Random forest.

Supplementary Table 4 Performance of the models for training set (genetic features)

Model	Precision	Sensitivity	Specificity	Accuracy	AUROC	F1
XGBoost	0.904	1	0.94	0.962	0.997	0.949
ET	0.556	1	0.529	0.704	0.935	0.714
GBDT	0.952	1	0.971	0.981	1	0.976
LR	0.586	0.85	0.647	0.722	0.819	0.694
RF	0.645	1	0.676	0.796	0.965	0.784

ROC: Receiver operating characteristic; XGBoost: Extreme gradient boosting gradient; ET: Extremely random trees; GBDT: Gradient boosting decision tree; LR: Logistic regression; RF: Random forest.

Supplementary Table 5 Performance of the models for training set (top five features)

Model	Precision	Sensitivity	Specificity	Accuracy	AUROC	F1
XGBoost	0.568	0.894	0.619	0.718	0.874	0.694
ET	0.472	0.919	0.493	0.634	0.826	0.624
GBDT	0.487	1	0.48	0.652	0.88	0.655
LR	0.425	0.838	0.44	0.571	0.72	0.564
RF	0.507	0.919	0.56	0.679	0.835	0.654

ROC: Receiver operating characteristic; XGBoost: Extreme gradient boosting gradient; ET: Extremely random trees; GBDT: Gradient boosting decision tree; LR: Logistic regression; RF: Random forest.

Supplementary Table 6 Performance of the models for testing set (clinical features)

Model	Precision	Sensitivity	Specificity	Accuracy	ROC	F1
XGBoost	0.526	0.833	0.571	0.667	0.687	0.645
ET	0.429	1	0.238	0.515	0.619	0.6
GBDT	0.4	0.5	0.571	0.545	0.635	0.444
LR	0.4	0.833	0.286	0.485	0.528	0.541
RF	0.429	1	0.238	0.515	0.718	0.6

ROC: Receiver operating characteristic; XGBoost: Extreme gradient boosting gradient; ET: Extremely random trees; GBDT: Gradient boosting decision tree; LR: Logistic regression; RF: Random forest.

Supplementary Table 7 Performance of the models for testing set (gene features)

Model	Precision	Sensitivity	Specificity	Accuracy	ROC	F1
XGBoost	0.692	0.75	0.81	0.788	0.802	0.72
ET	0.667	0.667	0.778	0.733	0.722	0.667
GBDT	1	0.667	1	0.867	0.778	0.8
LR	0.714	0.833	0.778	0.8	0.722	0.769
RF	0.714	0.833	0.778	0.8	0.907	0.769

ROC: Receiver operating characteristic; XGBoost: Extreme gradient boosting gradient; ET: Extremely random trees; GBDT: Gradient boosting decision tree; LR: Logistic regression; RF: Random forest.

Supplementary Table 8 Performance of the models for testing set (top five features)

Model	Precision	Sensitivity	Specificity	Accuracy	ROC	F1
XGBoost	0.769	0.833	0.857	0.848	0.861	0.8
ET	0.583	0.7	0.737	0.724	0.811	0.636
GBDT	0.529	0.9	0.579	0.69	0.85	0.667
LR	0.435	1	0.316	0.552	0.805	0.606
RF	0.471	0.8	0.526	0.621	0.792	0.593

ROC: Receiver operating characteristic; XGBoost: Extreme gradient boosting gradient; ET: Extremely random trees; GBDT: Gradient boosting decision tree; LR: Logistic regression; RF: Random forest.