Table 1 Demographic and clinical characteristics of the patients

Parameter	Number
Gender	
Male	250(47.17)
Female	280(52.83)
Age (years)	
≤ 44	121(22.83)
45 ≤ Age ≤ 59	193(36.42)
60 ≤ Age ≤ 74	132(24.91)
≥ 75	84 (15.85)
Body mass index (BMI, kg/m^2)	
< 18.5	48(9.06)
18.5 ≤ BMI ≤ 23.9	275(51.89)
≥ 24	175(33.02)
Charlson comorbidity index (Score)	
0	104(19.62)
1 or 2	190(35.85)
3 or 4	123(23.21)
≥ 5	113(21.32)

Data presented as number (%)

Table 2 Variable assignment

Number	Variable	Assignment
	Adverse drug reaction	1, Yes; 0, No
X1	Gender	1, Male; 0, Female
X2	Age (years)	1, ≤ 44; 2, 45 ≤ Age ≤ 59; 3, 60 ≤ Age ≤ 74; 4, ≥ 75
X3	Body mass index (BMI, kg/m^2)	1, < 18.5; 2, 18.5 ≤ BMI ≤ 23.9; 3, ≥ 24
X4	Asians	1, Yes; 0, No
X5	Genetic family history	1, Yes; 0, No
X6	History of allergy	1, Yes; 0, No
X7	Smoking	1, Yes; 0, No
X8	Alcohol	1, Yes; 0, No
X9	Temperature (℃)	1, < 36.1; 2, 36.1 ≤ Temperature ≤ 37.2; 3, > 37.3
X10	Pulse (beats/min)	1, < 60; 2, 60 ≤ Pulse ≤ 100, 3, > 100
X11	Breathe (times/min)	1, < 12; 2, 12 ≤ Breathe ≤ 20; 3, > 20
X12	Blood pressure	0, Normal (systolic pressure ≤ 139 mmHg or diastolic pressure ≤ 89 mmHg); 1, Grade I (140 mmHg ≤ systolic pressure ≤ 159 mmHg or 90 mmHg ≤ diastolic pressure ≤ 99 mmHg); 2, Grade II (160 mmHg ≤ systolic pressure ≤ 179 mmHg or 100 mmHg ≤ diastolic pressure ≤ 109 mmHg); 3, Grade III (systolic pressure ≥180 mmHg or diastolic pressure ≥110 mmHg)
X13	Charlson comorbidity index (Score)	1, 0; 2, 1 or 2; 3, 3 or 4; 4, ≥ 5
X14	Cardiovascular disease	1, Yes; 0, No

X15	Endocrine diseases	1, Yes; 0, No
X16	Respiratory diseases	1, Yes; 0, No
X17	Nervous diseases	1, Yes; 0, No
X18	Digestive diseases	1, Yes; 0, No
X19	Neoplastic diseases	1, Yes; 0, No
X20	Orthopedic diseases	1, Yes; 0, No
X21	Genito-urinary diseases	1, Yes; 0, No
X22	Hematopathy	1, Yes; 0, No
X23	Oculopathy	1, Yes; 0, No
X24	Ear-nose-throat diseases	1, Yes; 0, No
X25	Dermatoses	1, Yes; 0, No
X26	Immune rheumatism	1, Yes; 0, No
X27	Other diseases	1, Yes; 0, No
X28	Solvent	1, 0.9% sodium chloride injection; 2, 5% glucose injection; 3, Other solvents
X29	Dose (mg)	1, < 1.6; 2, =1.6; 3, > 1.6
X30	Anti-infective agents	1, Yes; 0, No
X31	Cardiovascular medicines	1, Yes; 0, No
X32	Medicines for digestive system	1, Yes; 0, No
X33	Respiratory medicines	1, Yes; 0, No
X34	Nervous system medicines	1, Yes; 0, No
X35	Medication in mental disorders	1, Yes; 0, No

X36	Non-steroidal anti-inflammatory drugs	1, Yes; 0, No
X37	Antiallergic agent	1, Yes; 0, No
X38	Genito-urinary system medicines	1, Yes; 0, No
X39	Medicines for hematopathy	1, Yes; 0, No
X40	Endocrine agents or hormone drugs	1, Yes; 0, No
X41	Antineoplastic drugs	1, Yes; 0, No
X42	Amino acids, vitamins, minerals or other nutrition preparations	1, Yes; 0, No
X43	Regulating water, electrolyte or acid-base balance drugs	1, Yes; 0, No
X44	Adjuvant agents to anesthesia or anesthetics	1, Yes; 0, No
X45	Diagnostic agents	1, Yes; 0, No
X46	Biological agents	1, Yes; 0, No
X47	Obstetrical-gynecological drugs	1, Yes; 0, No
X48	Stomatological preparations	1, Yes; 0, No
X49	Ophthalmic medication	1, Yes; 0, No
X50	Ear-nose-throat medication	1, Yes; 0, No
X51	Dermatology medication	1, Yes; 0, No
X52	Other traditional Chinese medicines	1, Yes; 0, No

	or Chinese patent medicines	
X53	Urea	1, Below the normal range; 2, Within the normal range; 3, Above the normal range
X54	Serum creatinine	1, Below the normal range; 2, Within the normal range; 3, Above the normal range
X55	Renal function	1, Glomerular filtration rate ≥ 90 ml/(min·1.73m^2); 2, 60ml/(min·1.73m^2) ≤ Glomerular filtration rate ≤ 89ml/(min·1.73m^2); 3, 30ml/(min·1.73m^2) ≤ Glomerular filtration rate ≤59 ml/(min·1.73m^2); 4, 15ml/(min·1.73m^2) ≤ Glomerular filtration rate ≤29 ml/(min·1.73m^2); 5, Glomerular filtration rate < 15 ml/(min·1.73m^2)
X56	Blood glucose	1, Below the normal range; 2, Within the normal range; 3, Above the normal range
X57	Serum potassium	1, Below the normal range; 2, Within the normal range; 3, Above the normal range
X58	Serum sodium	1, Below the normal range; 2, Within the normal range; 3, Above the normal range
X59	Total cholesterol	1, Below the normal range; 2, Within the normal range; 3, Above the normal range
X60	Triglyceride	1, Below the normal range; 2, Within the normal range; 3, Above the normal range
X61	High-density lipoprotein	1, Below the normal range; 2, Within the normal range; 3, Above the normal range
X62	Low-density lipoprotein	1, Below the normal range; 2, Within the normal range; 3, Above the normal range
X63	Albumin	1, Below the normal range; 2, Within the normal range; 3, Above the normal range
X64	Hypoproteinemia	1, Yes; 0, No
X65	Globulin	1, Below the normal range; 2, Within the normal range; 3, Above the normal range
X66	Albumin/globulin (A/G)	1, Below the normal range; 2, Within the normal range; 3, Above the normal range
X67	Aspartate aminotransferase	1, Below the normal range; 2, Within the normal range; 3, Above the normal range
X68	Alanine aminotransferase	1, Below the normal range; 2, Within the normal range; 3, Above the normal range

X69	Liver function	1, Less than 3 times upper limit of normal range of liver function tests (ULN of LFTs); 2, 3~5 times ULN of LFTs; 3, More than 5 times ULN of LFTs
X70	Total bilirubin	1, Below the normal range; 2, Within the normal range; 3, Above the normal range
X71	Lactic dehydrogenase	1, Below the normal range; 2, Within the normal range; 3, Above the normal range
X72	Creatine kinase	1, Below the normal range; 2, Within the normal range; 3, Above the normal range
X73	White blood cell	1, Below the normal range; 2, Within the normal range; 3, Above the normal range
X74	Neutrophil granulocyte	1, Below the normal range; 2, Within the normal range; 3, Above the normal range
X75	Lymphocyte percentage	1, Below the normal range; 2, Within the normal range; 3, Above the normal range
X76	Monocyte percentage	1, Below the normal range; 2, Within the normal range; 3, Above the normal range
X77	Eosinophils	1, Below the normal range; 2, Within the normal range; 3, Above the normal range
X78	Red blood cell	1, Below the normal range; 2, Within the normal range; 3, Above the normal range
X79	Hemoglobin	1, Below the normal range; 2, Within the normal range; 3, Above the normal range
X80	Platelet count	1, Below the normal range; 2, Within the normal range; 3, Above the normal range
X81	Hypersensitive C-reactive protein	0, Within the normal range; 1, Above the normal range
X82	Pre-treatment indicators of carcinoma	0, Within the normal range; 1, Above the normal range
X83	Pre-treatment serum levels	0, Within the normal range; 1, Above the normal range

Table 3 Results of different variable preprocessing methods

Method	Included variables
Column deletion	X1, X2, X3, X5, X7, X8, X12, X13, X14, X15, X16, X17, X18, X19, X20, X21, X22, X28, X29, X30, X31, X32, X33, X34, X35, X36, X39, X40, X41, X42, X43, X44, X45, X46, X51, X52, X54, X55, X56, X57, X58, X59, X60, X61, X62, X63, X65, X66, X67, X68, X71, X72, X73, X74, X75, X76, X77, X78, X79, X80, X81, X82, X83
Lasso	X1, X2, X18, X29, X30, X31, X33, X51, X52, X54, X55, X65, X66, X68, X78
Boruta	X1, X2, X5, X12, X13, X16, X17, X18, X20, X29, X30, X31, X33, X39, X40, X51, X52, X54, X55, X63, X66, X67, X68, X74, X75, X77, X78, X79

Variable names were shown in Supplementary Table 2.

Table 4 The effect of different data processing methods and machine learning algorithms on model prediction performance (Ten-fold cross-validation)

		AUC		Accuracy		Precision		Recall rate		F1 value	
		Mean±SD	95%CI	Mean±SD	95%CI	Mean±SD	95%CI	Mean±SD	95%CI	Mean±SD	95%CI
Data filling											
	No filling	0.868±0.099	0.864-0.872	0.820±0.093	0.816-0.823	0.772±0.190	0.765-0.779	0.720±0.254	0.710-0.730	0.729±0.217	0.721-0.737
	Simple filling	0.881±0.097	0.877-0.885	0.828±0.100	0.824-0.832	0.793±0.165	0.787-0.799	0.746±0.243	0.737-0.756	0.751±0.197	0.744-0.759
	RF filling	0.885±0.095	0.881-0.888	0.831±0.095	0.827-0.835	**0.802±0.157**	0.796-0.808	0.749±0.237	0.740-0.759	**0.757±0.189**	0.750-0.764
	RF improve filling	**0.887±0.094**	0.883-0.890	**0.832±0.096**	0.828-0.835	0.799±0.158	0.793-0.806	**0.751±0.240**	0.742-0.760	0.757±0.191	0.749-0.764
	p value	***p*<0.0001**		***p*<0.0001**		***p*<0.0001**		***p*<0.0001**		***p*<0.0001**	
Data sampling											
	No sampling	0.824±0.088	0.820-0.828	0.832±0.050	0.830-0.835	0.641±0.271	0.629-0.653	0.399±0.197	0.391-0.408	0.464±0.193	0.455-0.472
	Random over sampler	**0.923±0.063**	0.920-0.925	0.858±0.085	0.854-0.861	**0.849±0.079**	0.845-0.852	0.872±0.118	0.867-0.877	0.857±0.089	0.854-0.861
	Random under sampler	0.815±0.107	0.810-0.819	0.732±0.104	0.728-0.737	0.783±0.145	0.776-0.789	0.678±0.188	0.670-0.686	0.707±0.132	0.701-0.713
	SMOTE over sampler	0.920±0.072	0.917-0.923	0.857±0.081	0.853-0.860	0.844±0.071	0.841-0.848	0.875±0.125	0.869-0.880	0.856±0.089	0.852-0.860
	Borderline SMOTE	0.919±0.077	0.916-0.923	**0.859±0.085**	0.855-0.862	0.841±0.074	0.837-0.844	**0.885±0.130**	0.879-0.890	**0.859±0.093**	0.855-0.863
	p value	***p*<0.0001**		***p*<0.0001**		***p*<0.0001**		***p*<0.0001**		***p*<0.0001**	
Variable selection											

	No selection	0.870±0.105	0.867-0.874	0.820±0.104	0.817-0.824	0.780±0.178	0.774-0.786	0.733±0.254	0.725-0.742	0.737±0.208	0.730-0.744
	Lasso selection	**0.889±0.089**	0.886-0.892	**0.835±0.090**	0.832-0.838	**0.801±0.165**	0.796-0.807	**0.751±0.240**	0.743-0.759	**0.758±0.196**	0.752-0.765
	Boruta selection	0.881±0.094	0.878-0.884	0.827±0.093	0.824-0.830	0.794±0.162	0.788-0.799	0.741±0.236	0.733-0.749	0.750±0.191	0.744-0.757
	p value	***p*<0.0001**		***p*<0.0001**		***p*<0.0001**		***p*<0.0001**		***p*<0.0001**	
machine learning algorithms											
	AdaBoost	0.871±0.092	0.864-0.879	0.813±0.093	0.806-0.820	0.784±0.136	0.773-0.795	0.731±0.202	0.715-0.747	0.745±0.160	0.733-0.758
	Bagging	0.907±0.102	0.898-0.915	0.854±0.101	0.846-0.863	0.805±0.158	0.793-0.818	0.791±0.245	0.771-0.810	0.785±0.196	0.769-0.801
	Bernoulli NB	0.866±0.082	0.860-0.873	0.802±0.085	0.795-0.809	0.771±0.144	0.759-0.783	0.719±0.178	0.705-0.733	0.736±0.148	0.724-0.748
	DT	0.815±0.110	0.806-0.824	0.805±0.089	0.797-0.812	0.773±0.158	0.760-0.786	0.715±0.237	0.696-0.734	0.724±0.184	0.709-0.739
	ET	0.829±0.110	0.821-0.838	0.809±0.092	0.801-0.816	0.767±0.164	0.754-0.780	0.714±0.255	0.694-0.735	0.720±0.207	0.704-0.737
	Gaussian NB	0.845±0.089	0.838-0.852	0.786±0.085	0.779-0.793	0.734±0.155	0.722-0.747	0.743±0.164	0.730-0.756	0.730±0.143	0.719-0.742
	Gradient Boosting	0.891±0.102	0.883-0.899	0.841±0.099	0.833-0.849	0.822±0.149	0.810-0.834	0.746±0.252	0.725-0.766	0.762±0.194	0.747-0.778
	KNN	0.896±0.084	0.890-0.903	0.830±0.098	0.822-0.838	0.747±0.296	0.724-0.771	0.687±0.381	0.656-0.717	0.674±0.326	0.648-0.700
	LDA	0.897±0.073	0.891-0.903	0.835±0.081	0.829-0.842	0.805±0.117	0.796-0.815	0.768±0.191	0.753-0.783	0.777±0.144	0.765-0.788
	LR	0.893±0.076	0.886-0.899	0.834±0.082	0.827-0.840	0.815±0.119	0.805-0.824	0.754±0.216	0.737-0.772	0.767±0.157	0.755-0.780
	Multinomial NB	0.839±0.071	0.834-0.845	0.773±0.078	0.766-0.779	0.753±0.161	0.740-0.766	0.653±0.235	0.634-0.672	0.676±0.190	0.660-0.691
	Passive Aggressive	0.836±0.098	0.828-0.844	0.780±0.091	0.772-0.787	0.723±0.161	0.711-0.736	0.720±0.205	0.703-0.736	0.712±0.172	0.698-0.725

QDA	0.915±0.081	0.909-0.922	0.860±0.089	0.853-0.868	0.827±0.152	0.814-0.839	0.798±0.184	0.783-0.812	0.805±0.156	0.792-0.817
RF	0.919±0.097	0.911-0.926	0.871±0.100	0.863-0.879	0.843±0.154	0.831-0.856	0.775±0.268	0.753-0.796	0.788±0.214	0.771-0.805
SGD	0.895±0.075	0.889-0.901	0.832±0.082	0.825-0.839	0.803±0.197	0.787-0.819	0.710±0.287	0.687-0.733	0.726±0.238	0.707-0.745
SVM	**0.926±0.086**	0.919-0.933	**0.875±0.096**	0.867-0.883	**0.858±0.144**	0.847-0.870	0.776±0.271	0.754-0.797	0.791±0.217	0.773-0.808
XGBoost	0.922±0.092	0.914-0.929	0.869±0.100	0.861-0.877	0.825±0.153	0.812-0.837	**0.810±0.229**	0.792-0.828	**0.808±0.185**	0.793-0.822
p value	***p*<0.0001**		***p*<0.0001**		***p*<0.0001**		***p*<0.0001**		***p*<0.0001**	

AUC, Area under curve; RF, Random Forest; SMOTE, Synthetic minority oversampling technique; Bernoulli NB, Bernoulli Naïve Bayes; DT, Decision Tree; ET, Extra Tree; Gaussian NB, Gaussian Naïve Bayes; KNN, K-Nearest Neighbor; LDA, Latent Dirichlet Allocation; LR, Logistic Regression; Multinomial NB, Multinomial Naïve Bayes; QDA, Quadratic Discriminant Analysis; SGD, Stochastic Gradient Descent; SVM, support vector machine. XGBoost, eXtreme Gradient Boosting