

# Exploring Influencing Factors Including CYP2C19 Genotypes, and Developing a Machine Learning-Based Predictive Model for Clopidogrel Resistance in Chinese Patients with Ischemic Stroke

Beiming Xu<sup>1,\*</sup>, Lanting Gu<sup>1,\*</sup>, Guanwen Lin<sup>2</sup>, Hongyang Ni<sup>3</sup>, Haoqiang Shi<sup>1</sup>, Bing Chen<sup>1</sup>

<sup>1</sup>Department of Pharmacy, Ruijin Hospital affiliated to Shanghai Jiaotong University School of Medicine, Shanghai, People's Republic of China;

<sup>2</sup>Department of Anesthesiology, Hainan General Hospital, Hainan Affiliated Hospital of Hainan Medical University, Haikou, Hainan, People's Republic of China; <sup>3</sup>Department of Neurosurgery, Ruijin Hospital Affiliated to Shanghai Jiaotong University School of Medicine, Shanghai, People's Republic of China

\*These authors contributed equally to this work

Correspondence: Haoqiang Shi; Bing Chen, Department of Pharmacy, Ruijin Hospital, Shanghai Jiaotong University, Shanghai, People's Republic of China, Email shihaoqiang@126.com; chchenbing@163.com

**Background:** Clopidogrel resistance (CR) may diminish its antiplatelet effect, thereby increasing the risk of cardiovascular and cerebrovascular events. The cause of CR remains unclear, and it may be related to pharmacogenomics and coagulation markers. Machine learning is a novel approach to investigate the correlations among various factors. This study aimed to investigate the factors influencing CR in Chinese patients with ischemic stroke and to develop a precise and reliable predictive model for CR using machine learning.

**Methods:** Thromboelastography (TEG), a standard technique for assessing platelet inhibition, was used to measure the adenosine diphosphate (ADP)-induced platelet inhibition rate. CR was defined as an ADP-induced platelet inhibition rate of less than 30%. Genotypes of *CYP2C19* and *PON1* were identified using fluorescence in situ hybridization. The relationships between genotypes, laboratory indicators, and ADP-induced platelet inhibition rates or CR were examined. An extreme gradient boosting (XGBoost) machine learning method was applied to predict the occurrence of CR. Adaptive Synthetic technique was used for reliable data augmentation and the predictive model was internally validated via nested cross-validation.

**Results:** A total of 208 patients were enrolled in the study. Participants were categorized into the CR group (n=14) and the non-CR group (n=194). The CR group exhibited significantly lower activated partial thromboplastin time (APTT) levels compared with the non-CR group (P<0.05). Carriers of at least one loss-of-function (LOF) allele of *CYP2C19* had a significantly higher risk of CR than individuals without LOF alleles. Other risk factors for ischemic stroke, such as age, sex, and body weight, did not significantly affect platelet inhibition rates or CR. Based on the XGBoost model, *CYP2C19* genotype, D-dimer levels, platelet count, and total bilirubin were major contributors to the prediction of CR in Chinese patients with ischemic stroke. The area under the receiver operating characteristic curve was 0.9925±0.0067. The model's accuracy and sensitivity was 97.44% and 91.82%, respectively.

**Conclusion:** Genetic polymorphisms in *CYP2C19* are the primary factors influencing CR. A machine learning model may be useful for early prediction of CR and for guiding the rational use of clopidogrel.

**Keywords:** *CYP2C19*, genetic polymorphism, clopidogrel resistance, ischemic stroke, machine learning

## Introduction

Ischemic stroke is frequently caused by rupture of atherosclerotic plaques or thromboembolism, with platelet activation playing a central role in thrombosis. The global age-standardized incidence rate of ischemic stroke is estimated to reach 89.32 per 100,000 individuals by 2030.<sup>1</sup> As ischemic stroke is often managed with interventional therapy, patients

are likely to experience a period of hypercoagulability, during which effective antiplatelet therapy is crucial. Clopidogrel, a second-generation thienopyridine drug, is widely used in the treatment of myocardial infarction, acute coronary syndrome (ACS), and ischemic stroke.<sup>2</sup> Clopidogrel exerts its effect by irreversibly blocking the P2Y<sub>12</sub> receptor, thereby inhibiting ADP-mediated platelet activation and aggregation, reducing platelet–fibrinogen interactions, and preventing thrombosis. The pharmacological action of clopidogrel involves two distinct metabolic steps in the human body. Initially, clopidogrel is metabolized to 2-oxo-clopidogrel primarily by the enzyme CYP2C19. Subsequently, this intermediate metabolite is further catalyzed by CYP2C19 and CYP1A2 to generate active thiol metabolites.<sup>3</sup>

CYP2C19 exhibits marked genetic polymorphism. CYP2C19\*2, \*3, and \*17 alleles are among the most commonly occurring alleles. The CYP2C19\*2 and \*3 alleles can cause decreased CYP2C19 activity, whereas CYP2C19\*17 is associated with increased CYP2C19 activity. The CYP2C19\*2 allele is the most prevalent variant in Caucasian, African American, and Asian populations. In contrast, the CYP2C19\*3 allele is more frequently observed in Asian populations, with an occurrence rate of approximately 10%.<sup>4</sup> According to the Clinical Pharmacogenetics Implementation Consortium (CPIC) guidelines (2013), CYP2C19 phenotypes are categorized as ultrarapid metabolizers (UMs: \*1/\*17, \*17/\*17), extensive metabolizers (EMs: \*1/\*1), intermediate metabolizers (IMs: \*1/\*2, \*1/\*3, \*2/\*17), and poor metabolizer (PMs: \*2/\*2, \*2/\*3, \*3/\*3).<sup>5</sup> The presence of intermediate and poor metabolizer phenotypes has been considered a contributing factor to reduced platelet inhibition and an increased risk of major adverse cardiovascular and cerebrovascular events.

An ADP-induced platelet inhibition rate of less than 30% is generally considered indicative of clopidogrel resistance (CR).<sup>6,7</sup> The prognosis of patients with ischemic stroke including encountering recurrent stroke or stent thrombosis are significantly impacted by CR.<sup>7</sup> Currently, the mechanisms underlying CR remain inadequately elucidated. Various factors have been investigated for their association with CR. Previous studies have demonstrated that CR is significantly associated with extrinsic factors such as under-dosing and non-compliance,<sup>8</sup> as well as intrinsic factors like CYP2C19 and PONI genetic polymorphisms.<sup>6,8,9</sup> For instance, a study by Li reported that the CYP2C19\*2 and PONI Q192R alleles were key contributors to platelet hyperreactivity in patients with ACS receiving clopidogrel therapy, with these mutations increasing the risk of CR.<sup>3</sup> However, the findings of a study by Chang contradict this viewpoint.<sup>10</sup> Moreover, the majority of research on CR has focused on cardiovascular diseases, such as ACS and percutaneous coronary intervention (PCI), with relatively few studies conducted in patients with ischemic stroke.<sup>6,10,11</sup>

Thromboelastography (TEG) is widely used to assess the inhibition of adenosine diphosphate (ADP)-induced platelet aggregation, which shows the antiplatelet effect of clopidogrel after administration. However, TEG is an invasive detection method and is frequently carried out after administration of clopidogrel. Prediction antiplatelet effect based on factors influence CR before taking medicines is valuable.

Therefore, this study aimed to investigate the relationship between CYP2C19 genetic polymorphisms and CR in Chinese patients with ischemic stroke, to assess the various factors that may influence CR, and to develop a machine learning model to predict the occurrence of CR in this population.

## Materials and Methods

### Study Population

Patients with ischemic stroke admitted to the Neurosurgery Department of Ruijin Hospital (Shanghai, China) between July 10, 2017 and September 30, 2019 were retrospectively enrolled in this study. The inclusion criteria were as follows: (1) patients aged over 18 years; (2) patients requiring angiography to determine the necessity of stent implantation; (3) CYP2C19 genotypes detected; (4) TEG testing conducted after the administration of clopidogrel (75 mg, daily) for at least three days prior to angiography; and (5) National Institutes of Health Stroke Scale (NIHSS) score of the patient was evaluated. The exclusion criteria were as follows: (1) patients who failed to adhere to prescribed medication regimens; (2) NIHSS score exceeding four; and (3) lack of important clinical data such as coagulation function. Ultimately, 208 out of 326 patients met the inclusion and exclusion criteria and were enrolled in the study.

## Clinical Data Collection

Demographic characteristics and baseline data of the 208 patients were systematically collected and evaluated through a review of hospital medical records. These data included age, sex, body weight, and vascular risk factors such as smoking, alcohol consumption, hypertension, diabetes mellitus (DM), history of ischemic stroke or transient ischemic attack (TIA), peripheral arterial disease, previous myocardial infarction, arterial aneurysm, and other cardiovascular or cerebrovascular diseases. In addition, information on angiography and stenting, as well as laboratory indicators of liver and kidney function and coagulation status, were collected and evaluated.

## CYP2C19 and PONI Genotyping by Fluorescence in situ Hybridization

Leucocyte DNA was extracted from peripheral blood samples using the TIANamp Blood DNA Kit (Tiangen Biotech Co., Beijing, China) according to the manufacturer's standard protocol. The extracted DNA samples were stored at 4 °C and analyzed within one month. Genotyping of single-nucleotide polymorphisms (SNPs) for *CYP2C19\*2* (rs4244285), *CYP2C19\*3* (rs4986893), *CYP2C19\*17* (rs12248560), and *PONI* (rs662) was performed. SNP genotyping was performed using fluorescence in situ hybridization with a TL988A fluorescence detector (Xi'an TianLong, Shaanxi Province, China), strictly following the manufacturer's instructions.

## Platelet Function Testing

Assessment of ADP-induced platelet inhibition was performed using the TEG<sup>®</sup> 5000 Thrombelastograph<sup>®</sup> Hemostasis Analyzer system (Haemonetics Corporation, Boston, USA). Venous blood samples were collected from patients who received clopidogrel (75 mg, once daily) for a minimum of three days. Specifically, 2.7 mL of venous blood was anticoagulated with 3.2% sodium citrate, and 4.0 mL was anticoagulated with 14.7 U/mL lithium heparin. The analyzer was equipped with three channels: the first channel was loaded with 20 µL of 0.2 mol/L CaCl<sub>2</sub> and 340 µL of citrate-anticoagulated blood; the second channel contained 10 µL of activator F and 360 µL of heparin-anticoagulated blood; and the third channel was filled with 10 µL of activator F, 10 µL of ADP, and 360 µL of heparin-anticoagulated blood. The instrument software calculated the rate of platelet inhibition and presented the results as percentages.

## Statistical Methods

Data analysis was performed using SPSS Statistics 22 (IBM, Armonk, New York, USA), and the Kolmogorov–Smirnov test was applied to assess the normality of the continuous data distribution. Continuous variables were presented as mean (standard deviation) for normally distributed data and as median (interquartile range) for non-normal data distribution. Categorical variables were presented as counts (percentages) when the data did not follow a normal distribution. For comparative analysis, independent *t*-tests were used for continuous variables with normal distributions, whereas the Mann–Whitney *U*-test was applied for variables with non-normal distributions. Categorical data were compared using chi-square tests or Fisher's exact tests. P-values less than 0.05 were considered statistically significant. Chi-square tests were also applied to assess deviations of SNPs from Hardy–Weinberg equilibrium and to evaluate genotype differences between groups. The correlation between laboratory indicators and ADP-induced platelet inhibition rates or CR were analyzed using the Spearman correlation coefficient test or binary regression analysis.

## Machine Learning (ML)

Extreme gradient boosting (XGBoost) was used to develop the ML models. The indicators included in the analysis were as follows: demographic data, clinical data, and *CYP2C19* and *PONI* genotypes. Among the continuous variables, serum creatinine (SCR), white blood cell (WBC) count, red blood cell (RBC) count, fibrinogen (FG), and D-dimer levels exhibited positively skewed distributions, indicating the need for normalization during raw data preprocessing. Logarithmic transformation was applied to these features to achieve a normal distribution; consequently, the original features—SCR, WBC, RBC, FG, and D-DIMER—were deprecated. For categorical features, one-hot encoding was applied to spread them, as each feature had a limited number of categories.

To address the extreme class imbalance of the CR group (14 of 208 patients, 6.73%), which served as the target variable of model training, the original dataset was augmented using the Adaptive Synthetic (ADASYN) technique.<sup>12</sup> The dataset was then divided into five folds of training, validation, and test subsets with a ratio of 64:16:20 to conduct a 5×5 nested cross-validation pipeline. Samples within these subsets were shuffled, and their label distributions were maintained using a stratified splitting strategy. The model target variable, termed “Response to Clopidogrel,” was defined by categorizing the ADP-induced platelet inhibition rate: values below 30% were classified as the resistant group, and the remaining values as the non-resistant group. The model pipeline incorporated an embedded Random Forest Classifier, which served as both a feature selector and a multi-class label predictor.

In this study, the XGBoost Classifier—a Gradient Boosting Decision Tree algorithm widely recommended for tabular data and classification tasks<sup>13</sup>—was employed as the multi-class label predictor. The Tree Parzen Estimator<sup>14</sup> was used as the objective function for hyperparameter optimization and implemented using the Python package Optuna.<sup>15</sup> To reduce metric bias, five distinct dataset shuffle patterns were employed. A total of 1,000 hyperparameter optimization trials were conducted for each model type. The optimal feature set, as determined by the feature selector and the fine-tuned model, was subsequently obtained.

## Results

### Patient Characteristics

Based on the ADP-induced platelet inhibition rate, the cohort of 208 patients was divided into two groups: those with an inhibition rate of less than 30% (classified as CR) and those with normal platelet responsiveness (classified as non-CR). Fourteen patients were categorized into the CR group, while 194 patients were categorized into the non-CR group. The baseline characteristics of the two groups are presented in Table 1. No significant differences were observed between the two groups with respect to age, sex, body weight, BMI, incidence of hypertension or DM, history of ischemic stroke or TIA, previous myocardial infarction, arterial aneurysm, current smoking and drinking habits, or most laboratory indicators. In addition, subgroups categorized by smoking status (never, ex-smoker, and current smoker), drinking status (never, social drinker, and regular drinker), and stenting via angiography (intracranial stent and extracranial stent) were analyzed, and no significant differences were observed between the groups. In contrast, the CR group demonstrated significantly lower levels of activated partial thromboplast time (APTT) compared with the non-CR group ( $P < 0.05$ ).

### Genotype Distribution

The genotype distributions and allele frequencies of the genetic variants are presented in Table 2. All genotype distributions conformed to Hardy–Weinberg equilibrium. Among the 208 patients, 56.25% were carriers of the *CYP2C19*\*2 allele (47.12% heterozygous and 9.13% homozygous), 12.5% were carriers of the *CYP2C19*\*3 allele (12.02% heterozygous and 0.48% homozygous), and 0.48% were carriers of the *CYP2C19*\*17 allele (with no homozygotes). The *PON1* genotype was determined in 114 of the 208 patients, of whom 84.21% were carriers of the *PON1* Q192R allele, including 53 (46.49%) heterozygotes and 43 (37.72%) homozygotes.

### Relationship Between Genotypes, Phenotypes, and CR

No significant difference were observed consequenced CR among ultrarapid, extensive, intermediate, and poor metabolizer groups as shown in Table 3. However, individuals carrying at least one loss-of-function (LOF) of *CYP2C19* alleles group exhibited a higher likelihood of CR compared with those without LOF *CYP2C19* alleles ( $P = 0.038$ ,  $X^2 = 4.317$ ). Conversely, no significant differences were observed among the different *PON1* genotypes in the 114 individuals ( $P = 0.874$ ,  $X^2 = 0.358$ ).

### Machine Learning Models

A dataset comprising 65 features and 324 samples was constructed following data augmentation. The dataset was divided into five folds of training, validation, and test subsets according to the aforementioned strategy. Sixteen features were selected as predictors. The validation and test result, evaluated using the area under the receiver operating characteristic curve (AUROC), was  $0.9925 \pm 0.0067$ . The ROC–AUC curves of the five test subsets are shown in Figure 1.

**Table 1** Demographic and Clinical Information in CR Group and Non-CR Group of Chinese Ischemic Stroke Patients

Variables	Total (n=208)	Non-CR (n=194)	CR (n=14)	P
<b>Age (years)</b>	65(56,70)	65(56,70)	65.5(60.25,68.0)	0.798
<b>Gender (male) n (%)</b>	128(61.54%)	121(62.37%)	7(50.00%)	0.358
<b>WEIGHT (kg)*</b>	66.74±11.73	66.62±11.37	68.4±12.48	0.575
<b>BMI (kg/m<sup>2</sup>)*</b>	24.37±3.41	24.34±3.41	24.81±3.49	0.619
<b>Smoker, n (%)</b>				
YES	23(11.06%)	22(11.34%)	1(7.14%)	0.966
NO	185(88.94%)	172(88.66%)	13(92.86%)	
<b>Drinker, n (%)</b>				
YES	14(6.73%)	14(7.22)	0(0%)	0.605
NO	194(93.27%)	180(92.78%)	14(100.00%)	
<b>Risk factors, n (%)</b>				
Hypertension	145(69.71%)	135(69.59%)	10(71.43%)	1.000
Diabetes mellitus	52(25.00%)	47(24.23%)	5(35.72%)	0.338
History of ischemic stroke or TIA	59(28.37%)	55(28.35%)	4(28.57%)	1.000
Previous myocardial infarction	208(100.00%)	194(100.00%)	14(100.00%)	1.000
Arterial aneurysm	78(37.50%)	73(37.63%)	5(35.71%)	0.886
Other Cardiovascular or cerebrovascular diseases	90(43.27%)	84(43.30%)	6(42.86%)	0.974
<b>Angiography and stenting</b>				
Only angiography, n (%)	82(39.42%)	74(38.14%)	8(57.14%)	0.160
Stenting via angiography, n (%)	126(60.58%)	120(61.86%)	6(42.86%)	
<b>Laboratory measurement</b>				
ALT (IU/L)	20.00(14.00,29.00)	20.00(14.00,29.00)	18.00(16.00,20.00)	0.395
AST (IU/L)	20.00(17.00,25.00)	20.00(17.00,25.00)	22.00(18.25,23.00)	0.709
ALP (U/L)	72.00(60.50,91.00)	73.00(61.00,91.00)	66.50(54.75,82.25)	0.286
TBIL (μmol/L)	11.30(9.50,14.28)	11.20(9.50,14.23)	12.80(10.28,15.28)	0.241
DBIL (μmol/L)	2.10(1.60,2.60)	2.10(1.60,2.60)	2.30(2.00,2.68)	0.373
BUN (μmol/L)	5.80(4.70,6.70)	5.70(4.50,6.70)	6.15(5.58,6.70)	0.077
SCR (μmol/L)	73.00(62.00,87.00)	73.00(69.00,85.00)	71.50(59.00,96.00)	0.985
UA (μmol/L)	306.00(256.50,379.50)	305.00(255.00,379.00)	324.00(291.25,375.25)	0.602
WBC (×10 <sup>9</sup> /L)	6.86(5.30,8.39)	6.86(5.30,8.62)	7.00(5.65,7.73)	0.817
PLT (×10 <sup>9</sup> /L)	199.00(166.75,230.25)	198.00(163.25,288.75)	216.50(178.00,234.50)	0.253
APTT (s)	29.40(27.30,31.60)	29.50(27.30,31.68)	27.65(26.90,28.98)	0.034
PT (s)	11.60(10.90,12.10)	11.60(10.90,12.10)	11.35(10.37,11.73)	0.128
TT (s)	18.20(17.60,19.30)	18.20(17.58,19.20)	18.40(18.20,19.68)	0.395
FG (g/L)	2.70(2.40,3.20)	2.70(2.40,3.20)	2.55(2.43,3.20)	0.307
D-dimer (mg/L)	0.31(0.19,0.59)	0.32(0.20,0.60)	0.30(0.19,0.36)	0.818

**Notes:** \*In accord with normal distribution.

**Abbreviations:** BMI, body mass index; ALT, alanine aminotransferase; AST, aspartate aminotransferase; ALP, alkaline phosphatase; TBIL, total bilirubin; DBIL, direct Bilirubin; BUN, blood urea nitrogen; SCR, serum creatinine; UA, urine acid; PLT, platelet; PT, prothrombin time; TT, thrombin time; FG, fibrinogen.

At the test stage, the confusion matrix for model predictions was generated using the test subset (Figure 2) and SHapley Additive exPlanations (SHAP)<sup>16</sup> values were generated using the complete dataset. The model’s accuracy, reflecting specificity in the clinical diagnostic domain, was 97.44%, with only 5 out of 190 CR samples in the test subset not correctly identified. The sensitivity was 91.82%. Other ML Metrics (using CR as positive), such as F1 Score and Recall are 0.9503 ± 0.0215 and 0.9744 ± 0.0162, respectively. Relevant hyperparameter search space and the optimized hyperparameters for the five XGBoost classifiers were provided in Table S1 and S2.

To address the extreme class imbalance (14 positive cases out of 208 total samples), we employed a strategy of oversampling prior to data splitting. The ADASYN technique was applied to the entire dataset to generate a class-balanced, synthetic dataset. A 5-fold cross-validation (CV) was then performed on this augmented dataset.

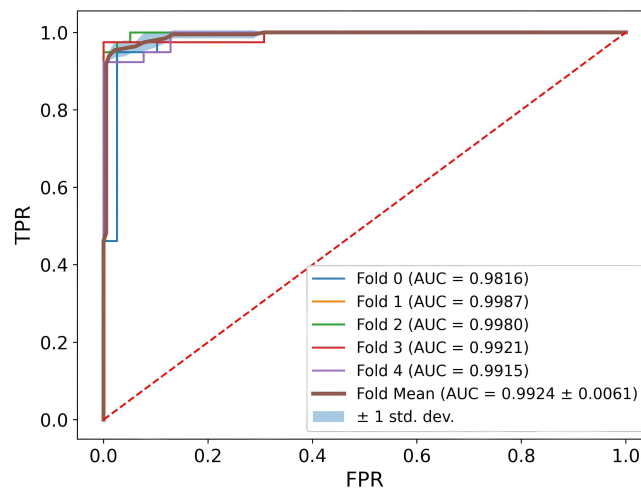
**Table 2** Gene Frequency of *CYP2C19* and *PONI* Distribution in Chinese Ischemic Stroke Patients

		Total	Minor Allele Frequency	Hardy-Weinberg Equilibrium (P value)
<i>CYP2C19</i> *2 (n=208)	GG(*1/*1)	91(43.75%)	32.69%	0.596
	GA(*1/*2)	98(47.12%)		
	AA(*2/*2)	19(9.13%)		
<i>CYP2C19</i> *3 (n=208)	GG(*1/*1)	182(87.50%)	6.49%	0.990
	GA(*1/*3)	25(12.02%)		
	AA(*3/*3)	1(0.48%)		
<i>CYP2C19</i> *17 (n=208)	CC(*1/*1)	207(99.52%)	0.24%	0.999
	CT(*1/*17)	1(0.48%)		
	TT(*17/*17)	0		
<i>PONI</i> (n=114)	QQ(AA)	18(15.79%)	39.04%	0.970
	QR(AG)	53(46.49%)		
	RR(GG)	43(37.72%)		

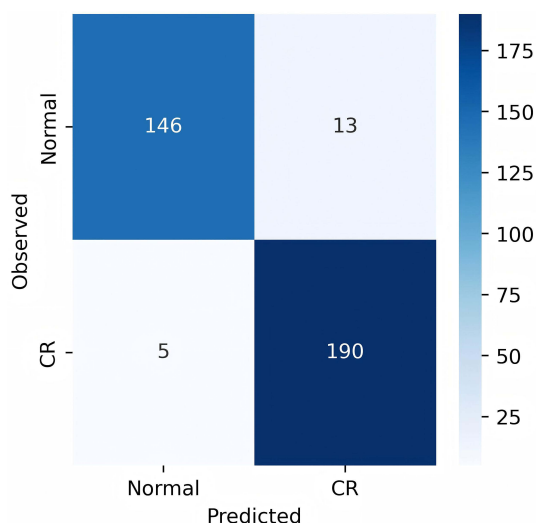
**Table 3** Relationship Between *CYP2C19* Phenotypes and Occurring of CR

<i>CYP2C19</i> Phenotypes	Total (n=208)	Non-CR (n=194)	CR (n=14)	P value	$\chi^2$ value
UM (*1/*17,*17/*17)	2(0.96%)	2(1.03%)	0	0.064	7.045
EM (*1/*1)	74(35.58%)	73(37.63%)	1(7.14%)		
IM (*1/*2, *1/*3,*2/*17)	100(48.08%)	89(45.88%)	11(78.57%)		
PM (*2/*2, *2/*3, *3/*3)	32(15.38%)	30(15.46%)	2(14.29%)		
UM and EM (no LOF of <i>CYP2C19</i> alleles)	76(36.54%)	75(38.66%)	1(7.14%)	0.038	4.317
IM and PM (at least one LOF of <i>CYP2C19</i> alleles)	132(63.46%)	119(61.34%)	13(92.86%)		

SHAP values, illustrated in Figure 3, indicate that *CYP2C19*\*2 genotype, D-dimer, PLT, TBIL, SEX and APTT contribute significantly to the resistance model, consistent with established clinical knowledge. Given that XGBoost is an ensemble decision tree estimator, 68 trees were constructed. We can identify the split point and the decision pathways for each response to clopidogrel. Similar patterns are observed across different decision trees, indicating that the IM and PM group exhibits a higher likelihood of CR.



**Figure 1** Average AUROC of the five test subsets using the XGBoost method.  
**Abbreviations:** TPR, True Positive Rate; FPR, False Positive Rate.



**Figure 2** Confusion Matrix Heat Map of test subsets (summarized).

In conclusion, our ML model demonstrates high sensitivity and specificity after hyperparameter optimization in cross-validation context, which rigorously evaluates its reliability and robustness.

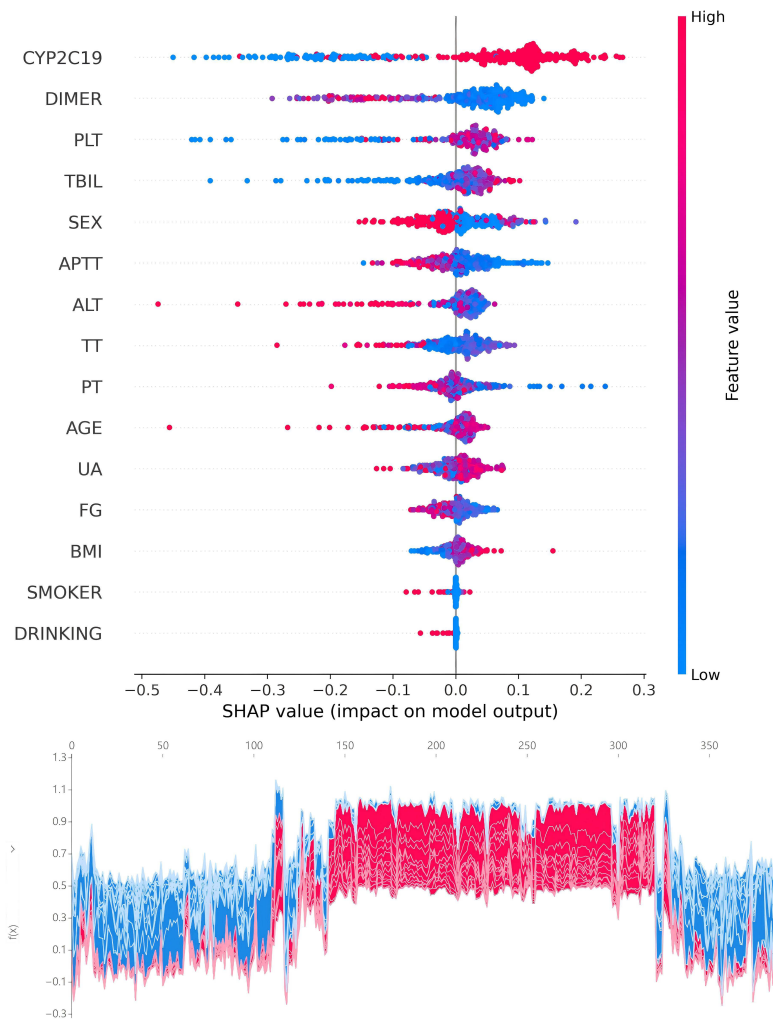
Figure 4 is a screenshot from a global explain program based on SHAP, which can display every individual case's predicted result and its corresponding SHAP value. Figure 4A, showing significant force difference between types of *CYP2C19*\*2, and the absolute value added to SHAP expectation value was the largest. Figure 4B and C showed the D-dimer and PLT has a similar force distribution among samples.

## Discussion

In the present study, *CYP2C19* and *PON1* genotypes in Chinese patients with ischemic stroke were determined, demographic data and laboratory indicators of the patients were recorded, and their relationship with CR was analyzed. Patients carrying at least one LOF allele of *CYP2C19* exhibited a significantly higher risk of developing CR compared with those without LOF alleles. APTT levels were significantly lower in the CR group than in the non-CR group. In addition, an XGBoost-based ML model was developed to predict CR in Chinese patients, in which *CYP2C19* genotypes, D-dimer, PLT, and TBIL were identified as valuable prediction factors.

Currently, the mechanisms underlying CR are not fully understood; however, several contributing factors have been identified, including both external and internal factors. External factors, such as patient compliance, obesity, smoking, and drug–drug interactions,<sup>8</sup> were not observed in this study. Internal factors include genetic polymorphisms, increased platelet receptor expression, and ADP release. Genetic polymorphism, in particular, plays a crucial role in inter-individual variability in clopidogrel efficacy.<sup>17–19</sup> The LOF alleles including *CYP2C19* \*2 and \*3 alleles are shown to decrease enzymatic activity, which can lower circulating levels of active clopidogrel metabolite, may result in a diminished antiplatelet response and a higher thrombosis risk.<sup>8</sup> Previous studies indicate that the *CYP2C19* gene contains at least one missing allele in approximately 60% of the Asian population. In China, the *CYP2C19*\*2 allele is the predominant LOF variant, accounting for 75–85% of intermediate and poor metabolizers.<sup>9</sup> The frequencies of *CYP2C19*\*2 and *CYP2C19*\*3 alleles observed in our study are consistent with those reported by Zhuo et al.<sup>19</sup> Although no statistically significant differences were observed between *CYP2C19*\*2, \*3, \*17 alleles and CR, patients carrying at least one LOF *CYP2C19* allele (IM and PM) exhibited a significantly higher risk of CR compared with non-carriers; notably, 13 of 14 patients with CR carried at least one LOF allele ( $P=0.038$ ). These findings are consistent with those reported by Peng<sup>6</sup> and Zhuo.<sup>20</sup>

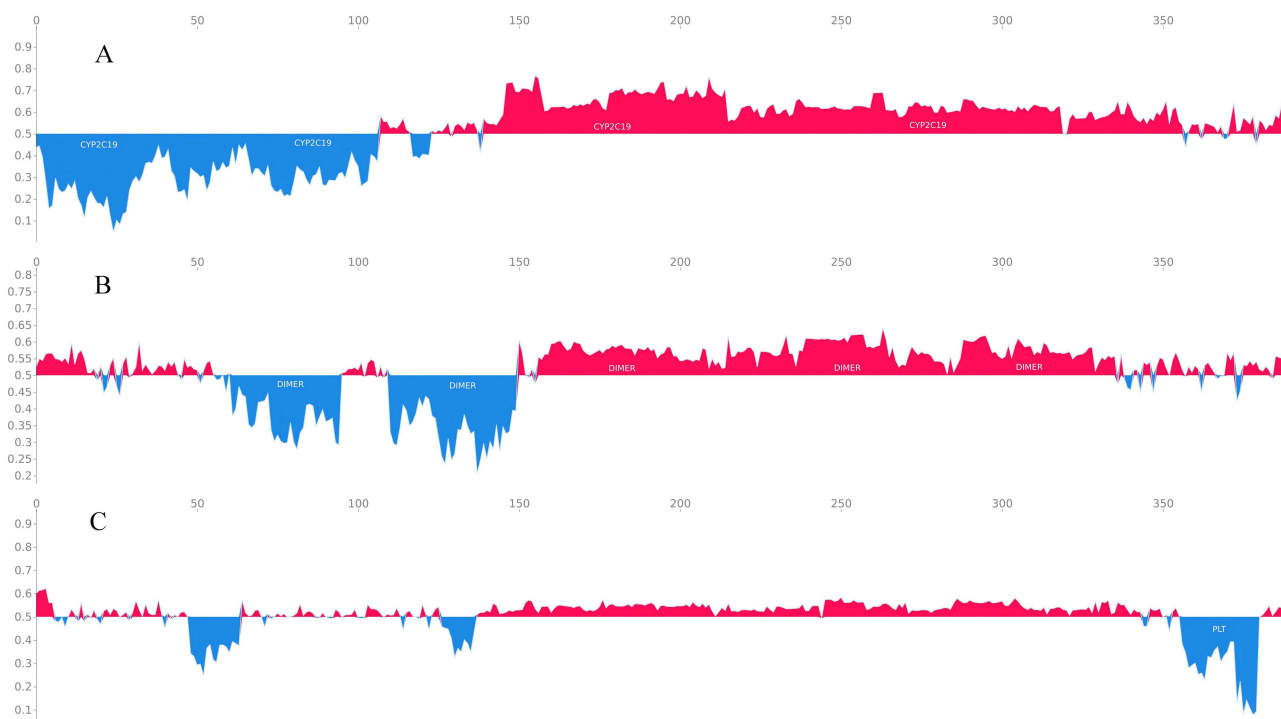
*PON1* is a crucial enzyme involved in the metabolism of clopidogrel. It has been proposed that *PON1* gene polymorphisms may reduce enzyme activity, thereby diminishing the antiplatelet efficacy of clopidogrel and potentially leading to CR.<sup>6,21</sup> However, Sibbing et al reported no significant association between *PON1* genotypes and clinical



**Figure 3** SHAP value summary in the final model for the estimation of CR.

outcomes or three-month prognosis in patients with cerebral infarction treated with clopidogrel.<sup>22</sup> Similarly, our study did not demonstrate any significant association between *PON1* genotype and CR. Moreover, the number of patients with *PON1* genotypes was limited; *PON1* is involved in the secondary metabolic step of clopidogrel and its metabolites may not be in the active forms. Consequently, the *PON1* 192Q mutation did not show a significant influence on platelet function or CR in this study.

In contrast, APTT levels were significantly lower in the CR group than in the non-CR group, indicating a tendency towards easier blood coagulation. This finding is consistent with the results from a study by Chang, which identified APTT as an independent risk factor for CR.<sup>10</sup> However, due to its modest contribution to CR prediction, we speculate that the reduction in APTT may be a consequence rather than a cause of CR. Additionally, D-dimer levels, PLT, and TBIL were identified by the model as factors influencing CR. These associations have been rarely reported in previous studies, except in a study by Zheng, which demonstrated the influence of PLT on CR risk.<sup>23</sup> In Wang’s predictive model for CR,<sup>24</sup> D-dimer and PLT were also included as predictive variables, along with WBC, hemoglobin, FG, PT, UA, glycated hemoglobin, and apolipoprotein B. However, the AUC and accuracy of Wang’s model (0.8730 and 0.8033, respectively) were lower than those of our model (0.9925 and 0.9744), which may be attributable to differences in sample size and selected predictive variables. D-dimer levels and PLT, which reflect a platelet-related hypercoagulable state, showed similar force distribution among samples, suggesting that common coagulation function indicators may be used to assist the identification of CR. TBIL also showed a strong influence on CR; apart from being a biochemical indicator



**Figure 4** SHAP force in the final model for the estimation of CR (X-axis represents the sample index; Y-axis represents the SHAP value; Red regions indicate positive contributions; Blue regions indicate negative contributions). (A) *CYP2C19* genotype; (B) D-dimer; (C) PLT).

of liver function, it may affect DNA methylation of the ADP receptor P2Y12, thereby contributing to CR risk, as reported by Su et al<sup>25</sup> Larger-scale studies are warranted to further assess the associations between CR and other potential factors.

ML tools have emerged as powerful computational approaches in data-intensive research fields. ML algorithms are computationally efficient, possess robust predictive capabilities, and facilitate learning in environments characterized by large datasets.<sup>26</sup> Individualized modeling of disease and drug dynamics is crucial in precision medicine, and ML-based computational techniques are gaining increasingly prominence. In future, ML tools are expected to be applied across diverse areas, including pharmacokinetic modeling, exposure/effect modeling, pharmacometric simulations, optimal dose selection, decision support through therapeutic drug monitoring, and systems pharmacology.<sup>27</sup> In our study, the developed resistance model achieved a sensitivity of 91.82% and a specificity of 97.44%, enabling the estimation CR based on several key features. Notably, the influence of *CYP2C19* genotype was primarily associated with CR prediction, consistent with previous research. Given the complexity of CR assessment and its dependence on numerous coagulation-related indices, the integration of ML techniques alongside genetic testing may facilitate the early identification of CR risks, thereby assisting clinical decision makers in modifying treatment plans to enhance the efficacy–cost ratio of medications in a data-driven manner.

## Limitations

Firstly, the sample size of CR is relatively small, which may impact the robustness of the model. Secondly, some other potential confounding factors such as *ABCB1* genotypes,<sup>28</sup> low-density lipoprotein levels<sup>18</sup> or the use of concomitant medications, such as aspirin<sup>10</sup> were not assessed, which may influence the clinical effectiveness of clopidogrel. Furthermore, the present study is a single-center study, an internal validation method was used to test the model. Therefore, future studies should investigate the roles of additional genotypes in CR, broaden the scope of the study population and complement external validation of the model.

## Conclusion

In this study, we investigated the relationship between *CYP2C19* polymorphisms, patient demographic and laboratory indicators, and clopidogrel resistance (CR) in Chinese patients with ischemic stroke. Our findings indicate that the presence of a LOF *CYP2C19* allele is significantly associated with an increased risk of CR. D-dimer levels, PLT, and TBIL were also found to contribute to CR. Based on the validation of data from larger cohorts, the developed ML model may be valuable for the early prediction of CR and for guiding the rational use of clopidogrel.

## Data Sharing Statement

Due to ethical restrictions given by Ethics Committee of Ruijin Hospital, our data sets are available from the corresponding author Bing Chen (chchenbing@163.com) upon reasonable request.

## Ethics Approval and Consent to Participate

The study was conducted in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of Ruijin Hospital (No. KY-2022-161). Due to the retrospective nature of the study, the requirement for written informed consent was waived by the ethics committee, and all the data of patients were anonymously and confidentially stored.

## Acknowledgments

Thank all staff members for data collection, data analysis, and monitoring as part of this study.

## Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

## Funding

This work was supported by the National Natural Science Foundation of China (Grant Numbers: 81973387).

## Disclosure

The authors state that the research was carried out without any commercial or financial ties that might be seen as a potential conflict of interest.

## References

1. Pu L, Wang L, Zhang R, et al. Projected global trends in ischemic stroke incidence, deaths and disability-adjusted life years from 2020 to 2030. *Stroke*. 2023;54(5):1330–1339. doi:10.1161/STROKEAHA.122.040073
2. Pereira NL, Rihal CS, So DY, et al. Clopidogrel Pharmacogenetics. *Circ Cardiovasc Interv*. 2019;12(4):e007811. doi:10.1161/CIRCINTERVENTIONS.119.007811
3. Li X, Wang Z, Wang Q, et al. Clopidogrel-associated genetic variants on inhibition of platelet activity and clinical outcome for acute coronary syndrome patients. *Basic Clin Pharmacol Toxicol*. 2019;124(1):84–93. doi:10.1111/bcpt.13110
4. Shetkar SS, Ramakrishnan S, Seth S, et al. CYP 450 2C19 polymorphisms in Indian patients with coronary artery disease. *Indian Heart J*. 2014;66(1):16–24. doi:10.1016/j.ihj.2013.10.001
5. Scott SA, Sangkuhl K, Stein CM, et al. Clinical pharmacogenetics implementation consortium guidelines for *CYP2C19* genotype and clopidogrel therapy: 2013 update. *Clin Pharmacol Ther*. 2013;94(3):317–323. doi:10.1038/clpt.2013.105
6. Peng W, Shi X, Xu X, et al. Both *CYP2C19* and *PON1* Q192R genotypes influence platelet response to clopidogrel by thrombelastography in patients with acute coronary syndrome. *Cardiovasc Ther*. 2019;2019:3470145. doi:10.1155/2019/3470145
7. Lavanya S, Babu D, Dheepthi D, et al. Clopidogrel resistance in ischemic stroke patients. *Ann Indian Acad Neurol*. 2024;27(5):493–497. doi:10.4103/aian.aian\_79\_24
8. Pradhan A, Bhandari M, Vishwakarma P, et al. Clopidogrel resistance and its relevance: current concepts. *J Family Med Prim Care*. 2024;13(6):2187–2199. doi:10.4103/jfmpc.jfmpc\_1473\_23
9. Lee CR, Luzum JA, Sangkuhl K, et al. Clinical pharmacogenetics implementation consortium guideline for *CYP2C19* genotype and clopidogrel therapy: 2022 update. *Clin Pharmacol Ther*. 2022;112(5):959–967. doi:10.1002/cpt.2526

10. Chang R, Zhou W, Ye Y, et al. Relationship between CYP2C19 polymorphism and clopidogrel resistance in patients with coronary heart disease and ischemic stroke in China. *Genet Res.* 2022;2022:1901256. doi:10.1155/2022/1901256
11. Zhang L, Lv Y, Dong J, et al. Assessment of risk factors for drug resistance of dual anti platelet therapy after PCI. *Clin Appl Thromb Hemost.* 2022;28:10760296221083674. doi:10.1177/10760296221083674
12. He H, Bai Y, Garcia EA, et al. ADASYN: adaptive synthetic sampling approach for imbalanced learning. *IEEE.* 2008. doi:10.1109/IJCNN.2008.4633969
13. Shwartz-Ziv R, Armon A. Tabular data: deep learning is not all you need. 2021. doi:10.1016/J.INFFUS.2021.11.011
14. Watanabe S. Tree-Structured parzen estimator: understanding its algorithm components and their roles for better empirical performance. *arXiv preprint.* 2023 arXiv:2304.11127v3
15. Akiba T, Sano S, Yanase T, et al. Optuna: a next-generation hyperparameter optimization framework. *ACM.* 2019. doi:10.1145/3292500.3330701
16. Lundberg S, Lee SI. A unified approach to interpreting model predictions. *Advanc Neural Informat Proc Syst.* 2017;30. doi:10.48550/arXiv.1705.07874
17. Sun Y, Lu Q, Tao X, et al. CYP2C19\*2 polymorphism related to clopidogrel resistance in patients with coronary heart disease, especially in the asian population: a systematic review and meta-analysis. *Front Genet.* 2020;11:576046. doi:10.3389/fgene.2020.576046
18. Shi GX, Zhao ZH, Yang XY, et al. Correlation study of CYP2C19 gene polymorphism and clopidogrel resistance in Han Chinese patients with cerebral infarction in Guizhou region. *Medicine.* 2021;100(6):e24481. doi:10.1097/MD.00000000000024481
19. Kim Y, Weissler EH, Pack N, et al. A systematic review of clopidogrel resistance in vascular surgery: current perspectives and future directions. *Ann Vasc Surg.* 2023;91:257–265. doi:10.1016/j.avsg.2022.12.071
20. Zhuo ZL, Xian HP, Long Y, et al. Association between CYP2C19 and ABCB1 polymorphisms and clopidogrel resistance in clopidogrel-treated Chinese patients. *Anatol J Cardiol.* 2018;19(2):123–129. doi:10.14744/AnatolJCardiol.2017.8097
21. Tresukosol D, Suktitipat B, Hunnangkul S, et al. Effects of cytochrome P450 2C19 and paraoxonase 1 polymorphisms on antiplatelet response to clopidogrel therapy in patients with coronary artery disease. *PLoS One.* 2014;9(10):e110188. doi:10.1371/journal.pone.0110188
22. Sibbing D, Koch W, Massberg S, et al. No association of paraoxonase-1 Q192R genotypes with platelet response to clopidogrel and risk of stent thrombosis after coronary stenting. *Eur Heart J.* 2011;32(13):1605–1613. doi:10.1093/eurheartj/ehr155
23. Zheng N, Yin F, Yu Q, et al. Associations of PER3 polymorphisms with clopidogrel resistance among Chinese Han people treated with clopidogrel. *J Clin Lab Anal.* 2021;35(4):e23713. doi:10.1002/jcla.23713
24. Wang RY, Yan SD, Zeng JQ, et al. Construction of a machine learning-based clopidogrel resistance risk prediction model. *Cardiovasc Toxicol.* 2025;25(10):1548–1560. doi:10.1007/s12012-025-10026-2
25. Su J, Li X, Yu Q, et al. Association of P2Y12 gene promoter DNA methylation with the risk of clopidogrel resistance in coronary artery disease patients. *Biomed Res Int.* 2014;2014:450814. doi:10.1155/2014/450814
26. McComb M, Bies R, Ramanathan M. Machine learning in pharmacometrics: opportunities and challenges. *Br J Clin Pharmacol.* 2022;88(4):1482–1499. doi:10.1111/bcp.14801
27. Stankevičiūtė K, Woillard JB, Peck RW, et al. Bridging the worlds of pharmacometrics and machine learning. *Clin Pharmacokinet.* 2023;62(11):1551–1565. doi:10.1007/s40262-023-01310-x
28. Zhang J, Dong ZF, Bian CX, et al. The correlation between MDR1 gene polymorphism and clopidogrel resistance in people of the hui and han nationalities. *Clin Appl Thromb Hemost.* 2022;28:10760296211073272. doi:10.1177/10760296211073272

International Journal of General Medicine

Publish your work in this journal

The International Journal of General Medicine is an international, peer-reviewed open-access journal that focuses on general and internal medicine, pathogenesis, epidemiology, diagnosis, monitoring and treatment protocols. The journal is characterized by the rapid reporting of reviews, original research and clinical studies across all disease areas. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/international-journal-of-general-medicine-journal>

**Dovepress**  
Taylor & Francis Group