

# Machine Learning Models of Hemorrhage/Ischemia in Moyamoya Disease and Analysis of Its Risk Factors

Zhongjun Chen

Department of Neurology, the Second Affiliated Hospital of Nanchang University

Haowen Luo

Medical Big-Data Center, the Second Affiliated Hospital of Nanchang University

Lijun Xu (✉ [56460984@qq.com](mailto:56460984@qq.com))

Department of Neurology, the Second Affiliated Hospital of Nanchang University

---

## Research Article

**Keywords:** LR, XGboost, MLP, Prediction model, Moyamoya disease

**Posted Date:** February 22nd, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-215146/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

**Object:** Identify the risk factors for hemorrhage/ischemia in patients with moyamoya disease and establish models using Logistic regression (LR), XGboost and Multilayer Perceptron (MLP), evaluating and comparison the effects of those models; providing theoretical basis for moyamoya disease patients to prevent stroke recurrence.

**Methods:** This retrospective study used data from the database of Jiang Xi Province Medical Big Data Engineering & Technology Research Center; the data of patients with moyamoya disease admitted to the second affiliated hospital of Nanchang university from January 1, 2012 to December 31, 2019 were collected. A total of 994 patients with moyamoya disease were screened, including 496 patients with cerebral infarction and 498 patients with cerebral hemorrhage. LR, XGboost and MLP were used to establish models for hemorrhage /ischemia in moyamoya disease, the effects of different models were verified and compared.

**Result:** LR, XGboost and MLP models all had good discrimination ( $AUC > 0.75$ ), and their AUC value are  $0.9227$  [95%CI:  $0.9215-0.9239$ ]  $0.9677$  (95%CI:  $0.9657-0.9696$ )  $0.9672$  [95%CI:  $0.9643-0.9701$ ]. Compared with LR model, the prediction ability of XGboost and MLP model in training and test set is improved, which is increased by 18.11% and 14.34% respectively in training set, and there is a significant difference.

**Conclusion:** Compared with the traditional LR model, the machine learning models are more effective in predicting hemorrhage/ischemia in moyamoya disease.

## Background

Moyamoya disease (MMD) is a rare cerebrovascular disease characterized by chronic progressive stenosis or occlusion at the onset of bilateral internal carotid arteries, anterior cerebral arteries and middle cerebral arteries, and abnormal vascular proliferation. The incidence rate is high in Asian countries such as Japan, Korea and China. At present, the etiology is not clear, and no clear pathogen has been detected; it is considered that multiple factors lead to the occurrence of moyamoya disease. People with susceptibility genes, under the influence of various external environmental factors, such as infection and inflammation, have abnormal immunologic mechanism, and the disease is caused by abnormal immunologic mechanism [1-3]. Family related studies report that the prevalence of familial MMD in China is 1.5% [4]. Vascular injury has been proved in autopsy reports of moyamoya disease patients [5]. Studies have shown that hyperhomocysteinemia, hypertension and smoking are the risk factors of vascular injury in patients with moyamoya disease [6].

The main clinical manifestations of moyamoya disease are cerebral ischemia or cerebral hemorrhage [7]. Most children patients take ischemic symptoms as clinical manifestations, and adult patients may be cerebral ischemia, hemorrhage or both. Intracranial hemorrhage is the initial symptom in half of adult patients with moyamoya disease [8] and it is basically consistent with this study. Study about the initial

symptoms of moyamoya disease using univariate analysis showed that family history of moyamoya disease and cerebral hemorrhage are risk factors of hemorrhagic stroke, and family history of cerebral infarction is risk factor of ischemic stroke, but multivariate analysis showed no statistical significance [4]. The study of Won-sang Cho indicated that cerebral hemorrhage is more common in familial moyamoya disease [9]. The risk or protective factors of moyamoya disease with cerebral hemorrhage as the initial symptom have not been fully identified, and hemorrhagic stroke is the main cause of death in patients with moyamoya disease. There is no ideal medical treatment for moyamoya disease, and the key treatment for moyamoya disease is reducing the incidence of hemorrhagic stroke and preventing of stroke recurrence. There are many controversies about the prevention of stroke recurrence in moyamoya disease patients, because of no exact theoretical basis for it, and lack of means to predict whether stroke recurrence is ischemic or hemorrhagic. There are few studies on the risk and protective factors for prediction the initial symptoms of moyamoya disease. In this study, we used three different methods to establish models of hemorrhage/ischemia in moyamoya disease to predict the type of stroke in patients with moyamoya disease. According to the models constructed in this study, the risk or protective factors of patients are input into the models for testing, aspirin can be recommended to prevent the recurrence of ischemic stroke if the probability of ischemic stroke is high; Conversely, aspirin treatment is not recommended to reduce the risk of cerebral hemorrhage if the risk of cerebral hemorrhage is high. The study provides a theoretical basis for the prevention of stroke recurrence in patients with moyamoya disease.

## Methods

### *Data Source*

In this retrospective study, the data of patients with moyamoya disease admitted to the Second Affiliated Hospital of Nanchang University from January 1, 2012 to December 31, 2019 were collected from the database of Jiang Xi Province Medical Big Data Engineering & Technology Research Center. All patients were required to meet the following criteria: 1. Moyamoya disease was diagnosed by CTA and / or MRA and / or DSA [10]. 2. The first onset of the disease 3. clinical neurological deficit symptoms; and the following cases were excluded: 1. Initial symptom is not cerebral infarction or cerebral hemorrhage. 2. Cases with incomplete information. 3. Cases with connective tissue disease or autoimmune disease or hyperthyroidism. 4. Cases with multiple neurofibroma. Patients with ischemic stroke as the initial symptom were selected as the cerebral infarction group and hemorrhagic stroke are selected as the cerebral hemorrhage group. The experimental protocol was approved by the ethics committee of the Second Affiliated Hospital of Nanchang University, and all the methods were carried out according to the relevant guidelines and regulations. Consent was obtained from all subjects or their parents (under 18 years old).

### *Variables*

According to the inclusion and exclusion criteria, the patients' information including gender, age of onset, ethnicity, marital status, long-term residence (urban or rural), type of medical insurance, fasting blood glucose, platelet count, high-density lipoprotein, low-density lipoprotein, triglyceride, total cholesterol, apolipoprotein, hypertension and diabetes history, smoking and drinking history, involved vessels, stenosis or occlusion, combined with aneurysm, other disease history and other data, and Suzuki rating performed according to the vascular condition of patients [4] were collected;

### *Statistical Analysis*

This study uses Python version 3.6 as the development tool to establish the models. The data set is randomly divided into training set and test set according to the ratio of 4:1. Different random seeds are used to repeat grouping for 20 times, based on the 3-fold cross validation method, the optimal parameters of XGboost and MLP in the training set are selected through the grid optimization method, and the parameters corresponding to the model with the largest area under the curve (AUC) are selected. Parameters range of XGboost model: max depth range is (2,3,4,5); n estimators range is (40,60,80,100); subsample range is (0.8,0.9,1); learning rate range is (0.01,0.05,0.1). MLP set Max ITER = 1000, random state = 0; Parameters range of MLP model: activation range is (identity, logistic, tanh, relu); the range of alpha is (0.01, 0.001, 0.0001); the range of hidden layer sizes is (15, 20, 25). MLP and LR call the corresponding functions in sklearn package; XGboost calls XGboost package; the feature importance is filtered by calling the PermutationImportance function of eli5. Sklearn package; SPSS statistics 18 software is used for univariate analysis of general data, t test or Mann Whitney U test for quantitative data, and c2 test for qualitative data. Z test is used for AUC value comparison between two models; in this paper, with  $P < 0.05$  as the standard, the difference is statistically significant.

## **Results**

### *General Patient Characteristics and establishment of the model*

The data of patients with moyamoya disease who were admitted to the Second Affiliated Hospital of Nanchang University from January 1, 2012 to December 31, 2019 and met the inclusion criteria but did not meet the exclusion criteria were collected, 994 patients were enrolled (496 in cerebral infarction group and 498 in cerebral hemorrhage group); According to the parameters described in the chapter of statistical methods, LR, XGboost and MLP methods were used to establish the prediction model of hemorrhage/ischemia in moyamoya disease. The patient screening process is shown in Fig.1, the general information of the two groups is shown in Table 1.

### **Table 1** General information of patients

	CI n=496	CH n=498	t / z / $\chi^2$ P
<b>General information</b>			
*Gender [n (%)]			
Male	223(44.96%)	281(56.43%)	$\chi^2=13.071, P<0.01$
Female	273(55.04%)	217(43.57%)	
*Age [n (%)]			
Juvenile $\leq 17$ years	7 (1.41%)	0 (0.00%)	$\chi^2=26.501, P<0.01$
Young adult 18-44 years	114(22.99%)	180(36.14%)	
The elderly $\geq 45$	375(75.60%)	318(63.86%)	
*Residence [n (%)]			
Urban	127 25.60%	31 6.22%	$\chi^2=71.64, P<0.01$
Rural	369 74.40%	467 93.78%	
* Hospitalization times $\bar{X} \pm S$	1.853 $\pm$ 1.4662	1.402 $\pm$ 0.978	t= 5.709, P<0.05
* Health insurance [n (%)]			
Employees	145 29.23%	97 19.48%	$\chi^2=81.665, P<0.01$
Urban residents	107 21.57%	89 17.87%	
Rural cooperative	120 24.20%	228 45.78%	
Retirement	2 0.40%	12 2.41%	
Own expense	92 18.55%	71 14.26%	
Other types	30 6.05%	1 0.20%	
PLT			
Higher than normal	44 8.87%	40 8.03%	$\chi^2=0.863, P=0.649$
Normal range	436 87.90%	437 87.75%	
Below normal	16 3.23%	21 4.22%	
<b>Risk factors of vessels</b>			
* Hypertension history [n (%)]	185 37.30%	86 17.27%	$\chi^2=50.271, P<0.01$

* DM history [n%]	45 [9.07%]	15 [3.01%]	$\chi^2=16.092, P<0.01$
Smoking [n%]	86 [17.34%]	96 [19.28%]	$\chi^2=0.624, P=0.429$
Alcohol using: [n%]	82 [16.53%]	92 [18.47%]	$\chi^2=0.649, P=0.421$
<b>* FBG</b>			
Higher than normal	135 [27.22%]	87 [17.47%]	$\chi^2=23.473, P<0.01$
Normal range	346 [69.76%]	408 [81.93%]	
Below normal	15 [3.02%]	3 [0.60%]	
<b>Blood fat [median (IQR)]</b>			
*HDL	1.080 (0.213)	1.106 (0.064)	$z= -4.678, P<0.01$
LDL	2.684 (0.720)	2.654 (0.158)	$z= -1.892, P=0.058$
*TG	1.400 (0.585)	1.354 (0.140)	$z= -4.678, P<0.01$
TC	4.364 (0.752)	4.357 (0.082)	$z= -0.712, P=0.476$
Apo A	1.050 (0.207)	1.055 (0.067)	$z= -1.164, P=0.143$
Apo B	0.836 (0.200)	0.837 (0.027)	$z= -0.241, P=0.810$
<b>Vessels condition [n%]</b>			
<b>* Involved vessels</b>			
MCA	264 [53.23%]	187 [37.55%]	$\chi^2=24.64, P<0.01$
CA	232 [46.77%]	311 [62.45%]	
<b>* Stenosis or occlusion</b>			
Stenosis	431 [86.90%]	353 [70.88%]	$\chi^2=38.23, P<0.01$
Occlusion	65 [13.10%]	145 [29.12%]	
* Aneurysm [n%]	12 [2.42%]	118 [23.69%]	$\chi^2=100.17, P<0.01$
*Suzuki stage X±S	2.748±0.7245	3.952±0.808	$t=-24.72, P<0.01$

The marked \* in the table indicates the risk factors with significant difference ( $P < 0.05$ )

### *Verification and comparison of models*

## Verification of models

According to the collected data, LR, XGboost and MLP are used in the test set of the two groups to verify the models, and some evaluation index of the models are compared between the test and the training set. The AUC values of the test set in the three models are all above 0.9 (> 0.75), which indicates that the predictive ability of the three models is very good. The details of the evaluation index of the models obtained from the training and test set are shown in Table 2.

## Comparison of evaluation index between training and test set

The AUC value, sensitivity, accuracy, specificity, positive predictive value, negative predictive value, positive likelihood ratio, negative likelihood ratio, Youden index and threshold of the three models are compared between the training set and the corresponding test set. The details of the evaluation index of the models are shown in Table 2.

**Table 2** evaluation index of training and test set of the models

	LR		XGboost		MLP	
	training set	test set	training set	test set	training set	test set
AUC	0.9227	0.9112	0.9677	0.9260	0.9672	0.9149
sensitivity	0.8291	0.8284	0.9085	0.8449	0.9185	0.8317
accuracy	0.8453	0.8402	0.9096	0.8621	0.91428	0.8482
specificity	0.8642	0.8627	0.9131	0.8907	0.9124	0.8780
PPV	0.8602	0.8598	0.9138	0.8861	0.9143	0.8713
NPV	0.8331	0.8291	0.9079	0.8439	0.9153	0.8311
PLR	6.2492	7.3518	11.3990	12.3768	12.0077	8.0112
NLR	0.1971	0.1964	0.0997	0.1730	0.0896	0.1903
Youden index	0.6933	0.6911	0.8216	0.7356	0.8309	0.7097
threshold	0.5154	0.5065	0.4769	0.5189	0.4744	0.5262
95% lower limit	0.9215	0.9065	0.9657	0.9225	0.9643	0.9115
95% upper limit	0.9239	0.9160	0.9696	0.9295	0.9701	0.9182
AUCstandard error	0.0006	0.0024	0.0010	0.0018	0.0015	0.0017

PPV= positive predictive value; NPV= negative predictive value; PLR= positive likelihood ratio

NLR= negative likelihood ratio.

### *ROC curve comparison between training and test set*

Taking random state=0 to divide test set and training set, the best parameters of XGboost model are: max depth =4, n estimators =100, subsample =0.9, learning rate =0.05, the best parameters of MLP model are: activation ='logistic ', alpha = 0.001, hidden layer sizes =15. The ROC curve of training and test set of the models is shown in Fig.2, 3.

It can be seen from table 2 that there is no significant difference between the AUC values of the training set and the corresponding test set of models,  $P > 0.05$ , and all the AUC values are above 0.9(>0.75), which indicates that the three models have good discrimination ability in predicting the hemorrhage/ischemia of moyamoya disease, but the AUC value of XGboost model is the closest to 1, which indicates that XGboost model may be better than other models in discrimination ability. The details of discrimination ability comparison between models are shown the section "comparison of discrimination ability".

### *Comparison between XGboost model and other models*

#### *Comparison of discrimination ability*

Comparison between training sets: In the training set, the AUC of XGboost model is 0.9677, which is larger than that of LR model (AUC = 0.9227) and MLP model (AUC = 0.9672), which indicated that XGboost model had the best discrimination ability in predicting the hemorrhage/ischemia of moyamoya disease; AUC between the two models is tested by Z test, Comparing the AUC value of XGboost and LR model, the statistic  $z = 0.3430$ ,  $P > 0.05$ , there is no significant difference between the two models in discrimination ability; comparing the AUC value of XGboost and MLP model,  $z = -0.2774$ ,  $P > 0.05$ , there is no significant difference also between the two models in discrimination ability.

Comparison between test sets: In the test set, AUC of XGboost model is 0.9260, which is larger than that of LR model (AUC = 0.9112) and MLP model (AUC = 0.9149). Comparing AUC of XGboost model and LR model, the statistic  $z = -0.2000$ ,  $P > 0.05$ , there is no significant difference between the two models in discrimination ability; comparing the AUC value of XGboost model and MLP model,  $z = 0.0404$ ,  $P > 0.05$ , there is no significant difference also between the two models in discrimination ability.

XGboost model better than LR and MLP model in discrimination ability, but there is no significant difference in training and test set.

#### *Comparison of accuracy*

ROC curve and AUC of the model are often used to compare the discrimination ability between two models, while Net Reclassification Index (NRI) is often used to compare the accuracy of the prediction ability between two models <sup>[11]</sup>.

Comparison between training sets: In the training set, Compared with LR model, the NRI of XGboost model is 0.1811,  $z = 7.9471$ ,  $P < 0.01$ , which indicates that the prediction ability of XGboost model has

been improved by 18.11%, and there is a significant difference. Compared with MLP model, the NRI of XGboost model is 0.0377,  $z = 2.1850$ ,  $P = 0.0290 < 0.05$ , which indicates that the prediction ability of XGboost model has also been improved by 3.77%, and there is a significant difference with  $P < 0.05$  as the standard. Compared with LR model, the NRI of MLP model is 0.1434,  $z = 6.5760$ ,  $P < 0.01$ , which indicates that the prediction ability of MLP model has also been improved by 18.11%, and there is a significant difference.

Comparison between test sets: In the test set, Compared with LR model, the NRI of XGboost model is 0.1107,  $z = 2.5247$ ,  $P = 0.0120 < 0.05$ , which indicates that the prediction ability of XGboost model has been improved by 11.07%, and there is a significant difference with  $P < 0.05$  as the standard. Compared with MLP model, the NRI of XGboost model is 0.0306,  $z = 0.7838$ ,  $P = 0.4330$ , which indicates that the prediction ability of XGboost model has been improved by 3.06%, but there is no significant difference. Compared with LR model, the NRI of MLP model is 0.0801,  $z = 2.1390$ ,  $P = 0.0320 < 0.05$ , which indicates that the prediction ability of MLP model has been improved by 8.01% in the test set, and there is a significant difference with  $P < 0.05$  as the standard.

In terms of the accuracy of the models, XGboost model is better than LR and MLP models in the training set, and there is a significant difference; in the test set, XGboost model is also better than LR and MLP models, and there is a significant difference with LR model. MLP model is better than LR model in training and test set, and there is a significant difference.

### *Feature importance of models*

The data of 994 patients with moyamoya disease (496 patients with cerebral infarction and 498 patients with cerebral hemorrhage) who met the inclusion criteria and did not meet the exclusion criteria were collected from the database of Jiang Xi Province Medical Big Data Engineering; LR, XGboost and MLP are used to establish the prediction model for hemorrhage/ischemia in moyamoya disease, according to the feature importance of each model, the top ten variables are shown in Fig. 4-6.

S00=stenosis or occlusion; FBG= fasting blood glucose; HT=hospitalization times; TC=total cholesterol

TG= Triglyceride; TOHI2=own expense for hospitalization; TOHI3=other type of health insurance;

TOHI6= rural cooperative health insurance;

The Suzuki grades of the two groups are compared by boxplot (Fig.7). It can be seen from Fig.7 that the Suzuki grades of the cerebral hemorrhage group are mainly concentrated in grades 4 and 5, while the Suzuki grades of the cerebral infarction group are mainly concentrated in grades 2 and 3.

Compared with the two groups of patients with or without aneurysm by histogram (Fig. 8), the proportion of patients with aneurysm in cerebral infarction group is only 2.42%, while that in cerebral hemorrhage group is 23.69%.

# Discussion

## *Effectiveness of the models*

At present, there are some related researches on prediction model of hemorrhage/ischemia in moyamoya disease, which mainly focus on the risk factors of hemorrhagic moyamoya disease, and most of them use LR to establish the model. Although XGboost and MLP have been widely used in artificial intelligence and other fields, they are rarely used in medical fields. In this study, LR, XGboost and MLP methods were used to establish the prediction model of hemorrhage/ischemia in moyamoya disease, and the important model evaluation indexes such as discrimination ability, accuracy and specificity of the three methods are compared. All the three methods have good discrimination ( $AUC > 0.75$ ); In the training set of this study, the AUC of XGboost model is 0.9677, which is larger than that of LR ( $AUC = 0.9227$ ) and MLP model ( $AUC = 0.9672$ ), indicating that XGboost model may have the best discriminant ability in predicting hemorrhage/ischemia of moyamoya disease, but there is no significant difference with LR and MLP model. In terms of NRI, an important index to evaluate the prediction accuracy of the model, XGboost model is superior to LR and MLP model in training set with significant difference, and its prediction ability is improved by 18.11% and 3.77% respectively; XGboost model is also superior to LR and MLP models in the test set, and with significant difference compared with LR model, but no significant difference with MLP model, and its prediction ability is improved by 11.07% and 3.06% respectively; The prediction ability of MLP model in training set and test set is better than LR model, and there is a significant difference; Compared with LR model, the prediction ability of MLP model in training set and test set is improved by 18.11% and 8.01% respectively. In summary, Discrimination ability of XGboost model may be better than that of LR and MLP model, but there is no significant difference with LR and MLP model; The prediction accuracy of XGboost model is better than that of LR and MLP models, and there is a significant difference; The prediction accuracy of MLP model is better than that of LR model, and there is a significant difference. The prediction accuracy of the model established by machine learning method is better than that of LR<sup>[12-13]</sup>, which is consistent with the research results of Gingal, churpek<sup>[14-16]</sup>. The accuracy, specificity, positive predictive value and negative predictive value of XGboost model are better than those of LR and MLP models, the reasons may be as follows: Firstly, XGboost model is a nonlinear ensemble learning algorithm model, and the tree model can be infinitely split, which can better fit the data than LR model<sup>[12]</sup>; Secondly, each factor in LR is independent and assumes a linear relationship, and XGboost has strong plasticity and flexibility<sup>[13]</sup>, which can automatically find and use the interaction effect and nonlinear relationship between related factors, making the prediction effect more accurate<sup>[17]</sup>. Finally, as a kind of forward structure artificial neural network, MLP can deal with the problem of nonlinear separability, and has good fault tolerance, strong adaptive and self-learning function<sup>[18]</sup>. However, its learning speed is slow, and it is very difficult to select the number of hidden nodes in the network, which may lead to insufficient learning. MLP model in this paper is better than LR model, and worse than XGboost model, which may be due to insufficient learning, but it needs further verification.

## *Feature importance of models*

In the univariate statistical analysis of general data, all risk factors are analyzed by univariate analysis, which is obviously not rigorous and accurate enough to predict the initial symptoms of moyamoya disease. The traditional LR method assumes that the contribution of all factors to the model is linear, but it is not so in clinical practice. However, machine learning can automatically discover and utilize the interaction effect and nonlinear relationship between related factors when establish models, which is also the difference between traditional regression analysis and machine learning method in modeling<sup>[12]</sup>. According to the top ten factors of feature importance of the three models established in this study, LR model indicated that hemorrhagic stroke as the first symptom may be closely related to Suzuki rating, with aneurysm or not, residence, involved vessels, vascular stenosis or occlusion and hospitalization times; MLP model indicated that hemorrhagic stroke as the first symptom may be closely related to Suzuki rating, with aneurysm or not, other types of medical insurance, apolipoprotein A and hospitalization times; XGboost model suggests that hemorrhagic stroke as the initial symptom may be closely related to Suzuki rating, with aneurysm or not, fasting blood glucose, hospitalization times and residence. In this paper, the top ten factors of feature importance of the three models all indicated that hemorrhagic stroke as the initial symptom of moyamoya disease patients may be closely related to Suzuki rating, with aneurysm or not, hospitalization times, residence and age of onset. In conclusion, multiple factors determine the initial symptoms of patients with moyamoya disease, Suzuki rating, with aneurysm or not, hospitalization times, residence, age of onset and other risk factors may be important risk factors for hemorrhagic stroke, which is similar to some reports<sup>[19]</sup>.

## Limitations

In this study, the data of patients is collected only from the big database of Jiang Xi Province Medical Big Data Engineering & Technology Research Center; the earlier the cases in the big database, the higher the proportion of incomplete data, which leads to the higher proportion of cases not included in this study because of incomplete information; Due to the limitation of medical level, the earlier the case is, the lower the proportion is diagnosed as moyamoya disease. In this study, these factors increase the sampling error of data collection. Although the sample size of this study is 994 cases, but the sample size of 994 cases is still not very large for MLP method, which has a certain impact on the performance of MLP model. Due to the geographical limitations of the patients, some risk factors in the collected data are eventually eliminated, such as the ethnicity of the patients, because 100% of them are Han majority.

In conclusion, XGboost model can better predict the initial symptoms of patients with moyamoya disease, and provide an important theoretical basis for preventing stroke recurrence of patients with moyamoya disease. We can reduce the impact of regional limitations on the research results by multi center research and increasing the sample size, and reduce the sampling error by collecting the case data in recent years, so as to improve the performance of the model.

## Conclusion

In this study, the data of 994 patients with moyamoya disease collected from database, LR, XGboost and MLP are used to establish prediction model for hemorrhage/ischemia in moyamoya disease, and the three models are tested and compared. The conclusion is as follows: 1) All the three models have good predictive ability in predicting hemorrhage / ischemia in patients with moyamoya disease, XGboost model is better than the other two models in accuracy, specificity, positive predictive value, negative predictive value and other important evaluation indexes. 2) XGboost model may be better than LR and MLP models in terms of prediction discrimination ability, but there is no significant difference in training and test set. 3) In terms of the accuracy of the model, XGboost model is better than LR and MLP models in the training set, and there is a significant difference; in the test set, XGboost model is also better than LR and MLP models, and there is a significant difference with LR model. MLP model is better than LR model in training set and test set, and there is a significant difference. 4) According to the order of the feature importance of the three models, Suzuki rating, whether with aneurysm, involved vessels, vascular stenosis or occlusion, residence in villages and towns, hospitalization times all contributed a lot to the three models, which indicates that they may be important risk factors of hemorrhagic stroke; These risk factors also have significant differences in univariate analysis, which also indicates that they may be important risk factors for hemorrhagic stroke.

## Declarations

### Author information

### Affiliations

**1 Department of Neurology, the Second Affiliated Hospital of Nanchang University,**

**Nanchang, Jiangxi, 330006, P.R. China;**

Zhongjun chen, Lijun Xu

**2 Medical Big-Data Center, the Second Affiliated Hospital of Nanchang University**

**Nanchang, Jiangxi, 330006, P.R. China;**

Haowen Luo

### Contribution

All authors designed the study, collected and processed the data, and reviewed the final version of the manuscript; Zhongjun Chen contributed to the analysis and writing of manuscripts; Haowen Luo contributed to statistical analysis, model establishing and manuscript preparation.

### Corresponding author

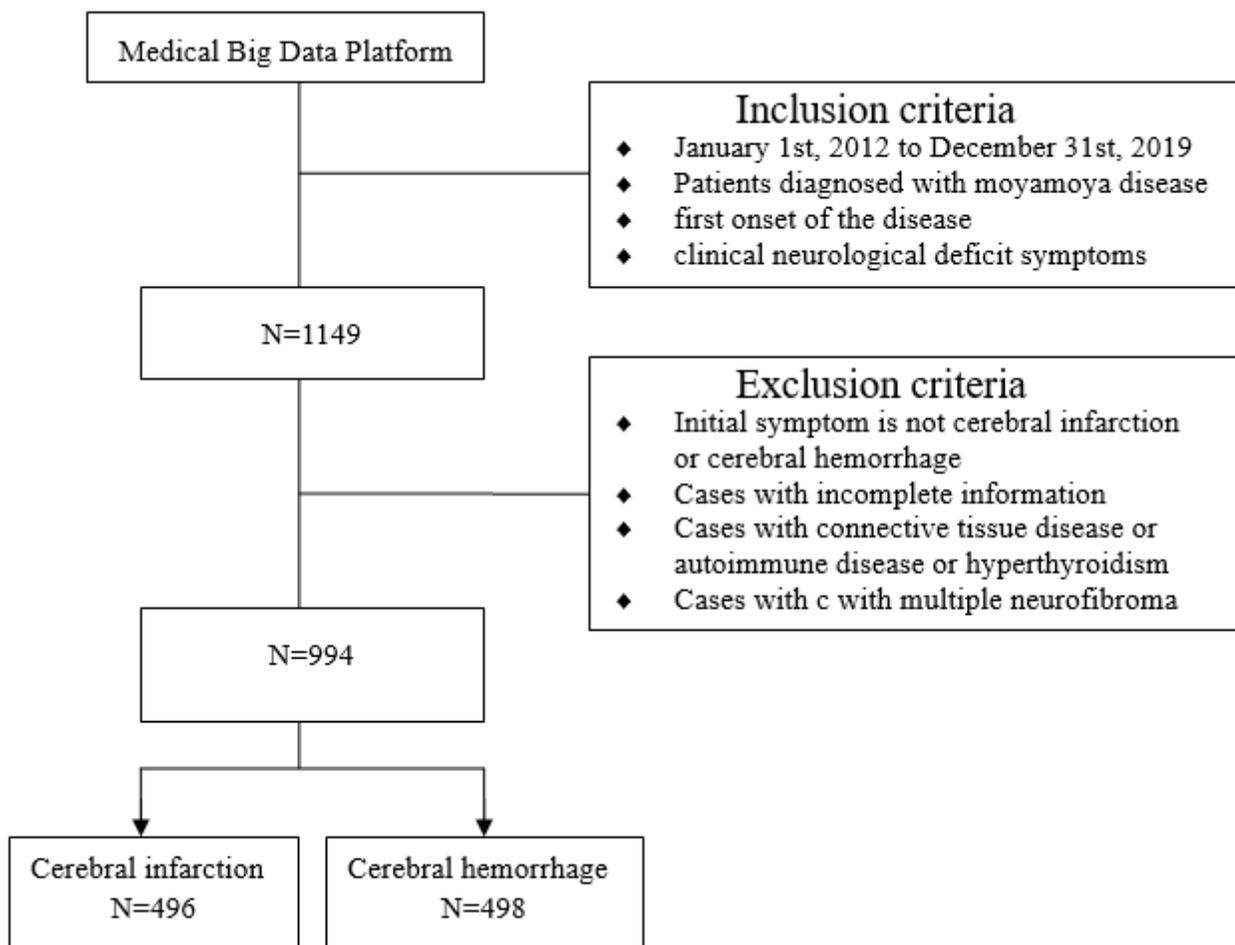
Haowen Luo (Email: [370904129@qq.com](mailto:370904129@qq.com))

## References

1. Miki Fujimura, Taku Fujimura, Aya Kakizaki, et al. Increased serum production of soluble CD163 and CXCL5 in patients with moyamoya disease: Involvement of intrinsic immune reaction in its pathogenesis. *Brain Research*, 2018:39-44.
2. Juan C. Mejia-Munne, Jason A. Ellis, Neil A. Feldstein, et al. Moyamoya and Inflammation. *World Neurosurg*, 2017, 100:575-578.
3. Tong Zhang, Congrong Guo, Xin Liao, et al. Genetic analysis of RNF213 p.R4810K variant in non-moyamoya intracranial artery stenosis/occlusion disease in a Chinese population. *Environmental Health and Preventive Medicine* (2017) 22:41
4. Miao W, Zhao P L, Zhang Y S, et al. Epidemiological and clinical features of Moyamoya disease in Nanjing, China[J]. *Clin Neurol Neurosurg*, 2010, 112(3):199-203.
5. Weinberg D G, Arnaout O M, Rahme R J, et al. Moyamoya disease: a review of histopathology, biochemistry, and genetics.[J]. *Neurosurgical Focus*, 2011, 30(6):E20.
6. Spencer C G, Martin S C, Felmeden D C, et al. Relationship of homocysteine to markers of platelet and endothelial activation in "high risk" hypertensives: a substudy of the Anglo-Scandinavian Cardiac Outcomes Trial.[J]. *International Journal of Cardiology*, 2004, 94(2):293-300.
7. Suzuki J, Takaku A. Cerebrovascular "moyamoya" disease. Disease showing abnormal net-like vessels in base of brain. *Arch Neurol*, 1969, 20(3): 288-299.
8. Baba T, Houkin K, Kuroda S. Nove1 epidemio1ogica1 features of moyamoya disease. *J Neurol Neurosurg Psychiatry*, 2008, 79(8) 900-904.
9. Cho W S, Batchuluun B, Lee S J, et al. Recurrent subdural hematoma from a pseudoaneurysm at the cortical branch of the middle cerebral artery after mild head injury: case report. [J]. *Neurol Med Chir*, 2011, 51(3):217-221.
10. Research Committee on the Pathology and Treatment of Spontaneous Occlusion of the Circle of Willis. Guidelines for diagnosis and treatment of moyamoya disease (spontaneous occlusion of the circle of Willis) [J]. *Neurol Med Chir (Tokyo)*, 2012, 52(5):245-266.
11. Farr S A, Banks W A, Morley J E. Effects of leptin on memory processing[J]. *Peptides* 2006(27):1420-1425
12. Shimoda A, Ichikawa D, Oyama H. Using machine-learning approaches to predict non-participation in a nationwide general health check-up scheme. *Comput Method Progr Biomed* 2018(163):39-46.
13. Hong WS, Haimovich AD, Taylor RA. Predicting hospital admission at emergency department triage using machine learning. *PLoS One* 2018;3:e201016.
14. Morley John E. Current opinion in clinical nutrition and metabolic care, Cognition and nutrition [J]. 2014(17):1-4
15. Beam AL, Kohane IS. Big data and machine learning in health care. *JAMA* 2018;319(13):1317-1318.

16. Qing Li, Xueqin Yao, Damien Échevin. How Good Is Machine Learning in Predicting All-Cause 30-Day Hospital Readmission? Evidence From Administrative Data. ScienceDirect. 2020; 23(10):1307–1315.
17. Cherrier M M, Matsumoto A M, Amory K. Testosterone improves spatial memory in men with Alzheimer disease and mild cognitive impairment [J]. Neurology, 2005(64):2063-2068.
18. Reddy Vivek, Doshi Shephal K, Sievert Horst, et al. Percutaneous left atrial appendage closure for stroke prophylaxis in patients with atrial fibrillation: 2.3-Year Follow-up of the PROTECT AF (Watchman Left Atrial Appendage System for Embolic Protection in Patients with Atrial Fibrillation) Trial [J]. Circulation. 2013(127):720-729.
19. Arash Ghaffari-Rafi, Shadeh Ghaffari-Rafi, Jose Leon-Rojas. Socioeconomic and demographic disparities of moyamoya disease in the United States. Clinical Neurology and Neurosurgery. 192(2020)105719.

## Figures



**Figure 1**

Flow chart of patient screening

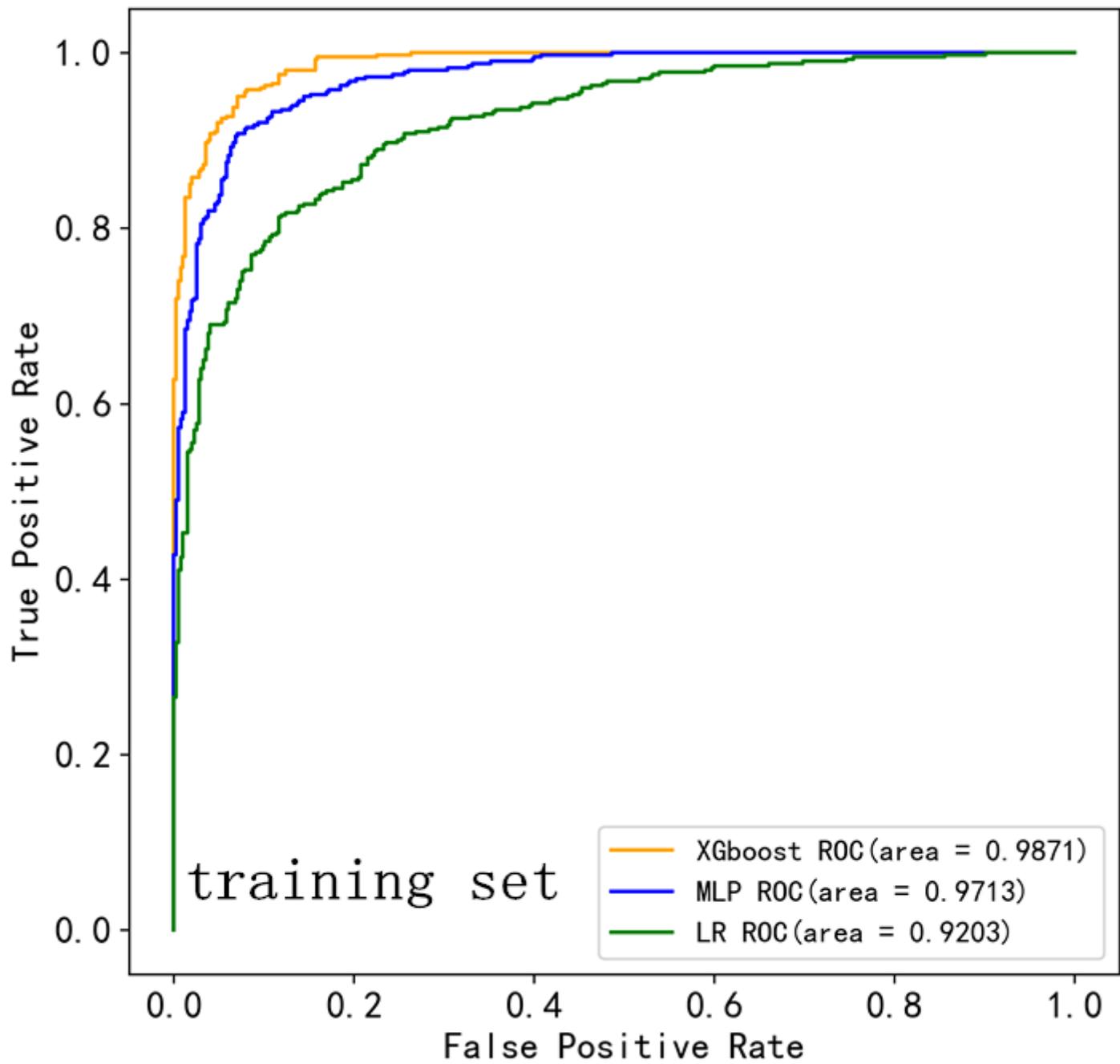


Figure 2

ROC curve of training set

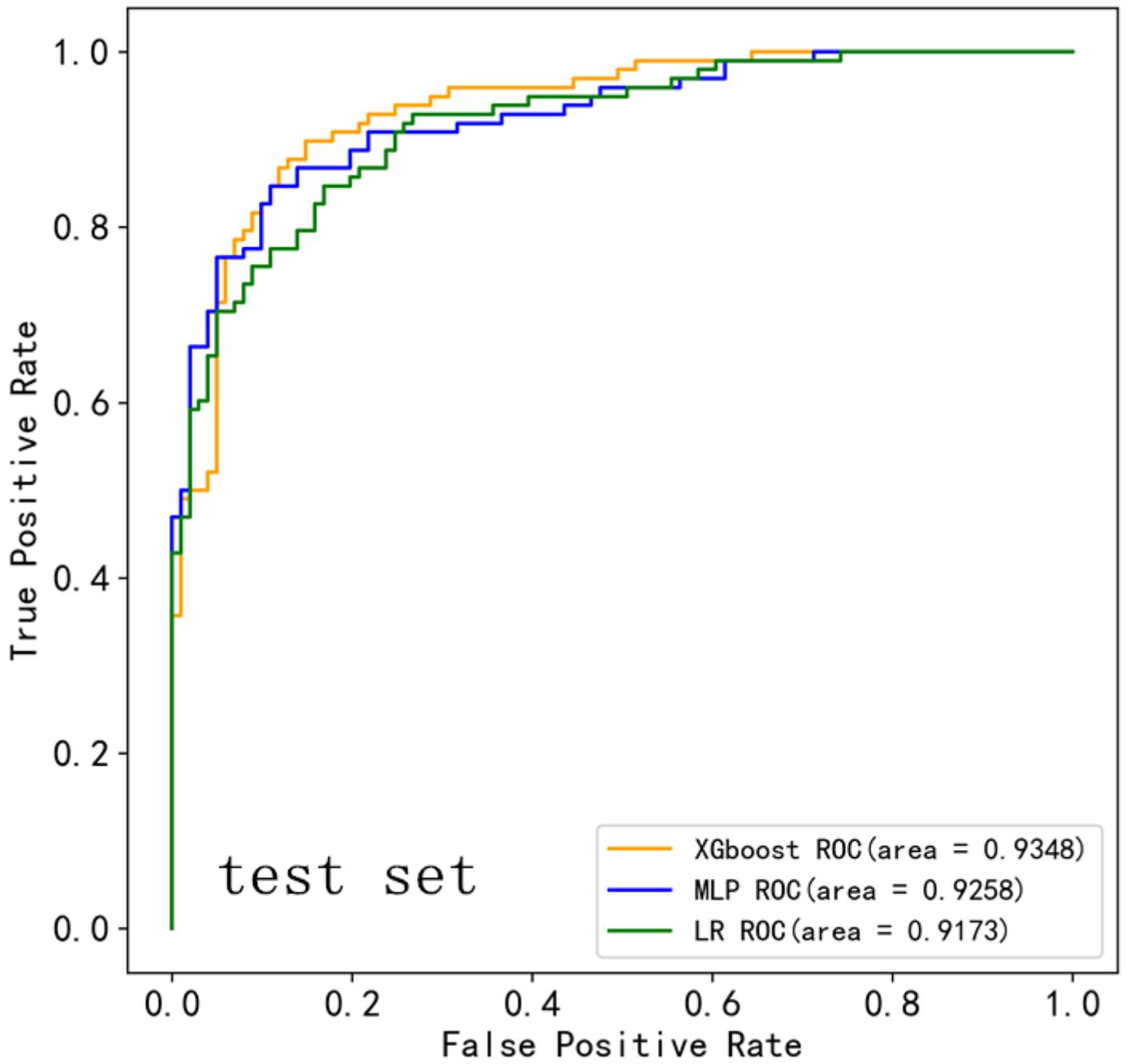


Figure 3

ROC curve of test set

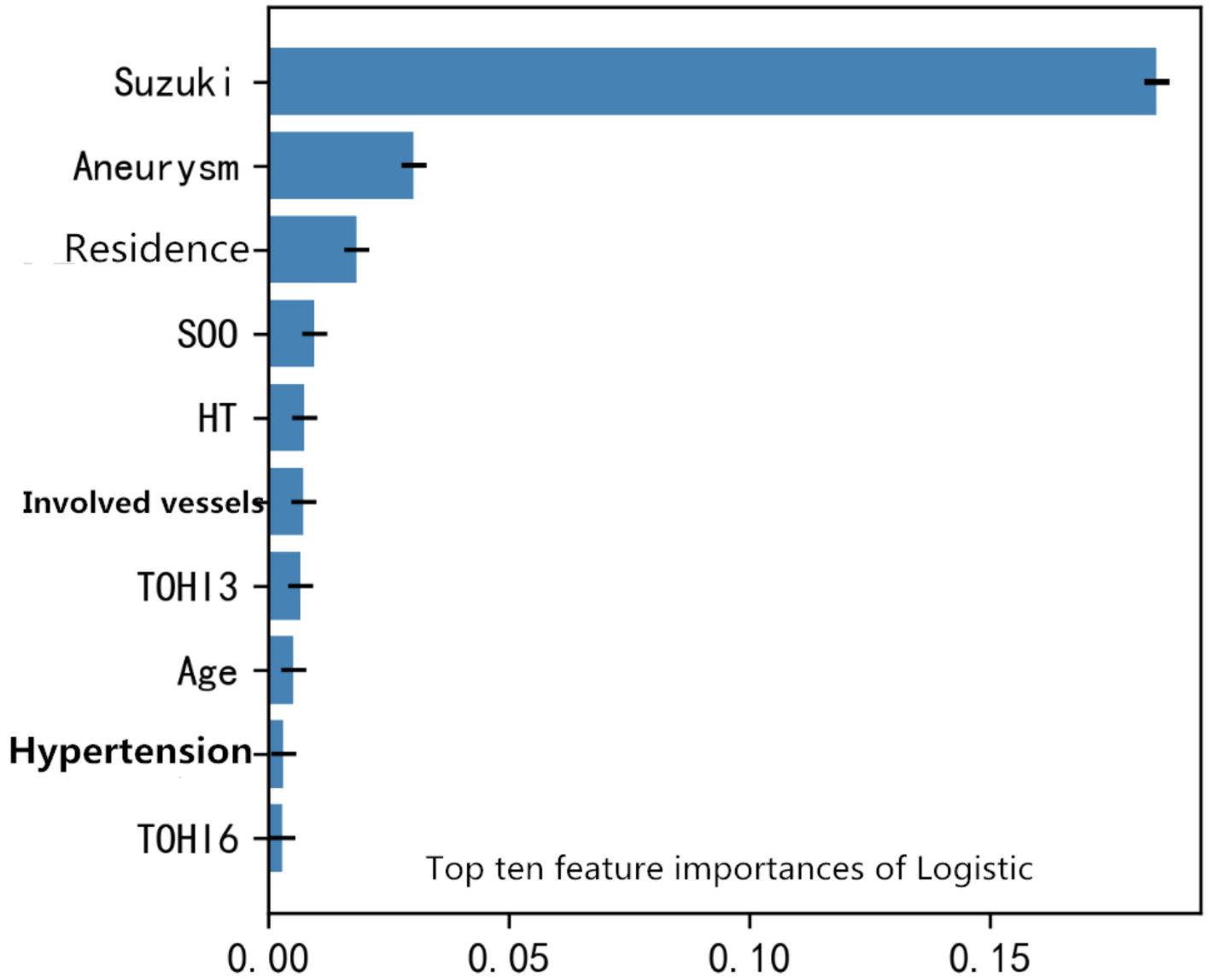


Figure 4

top ten variables of LR

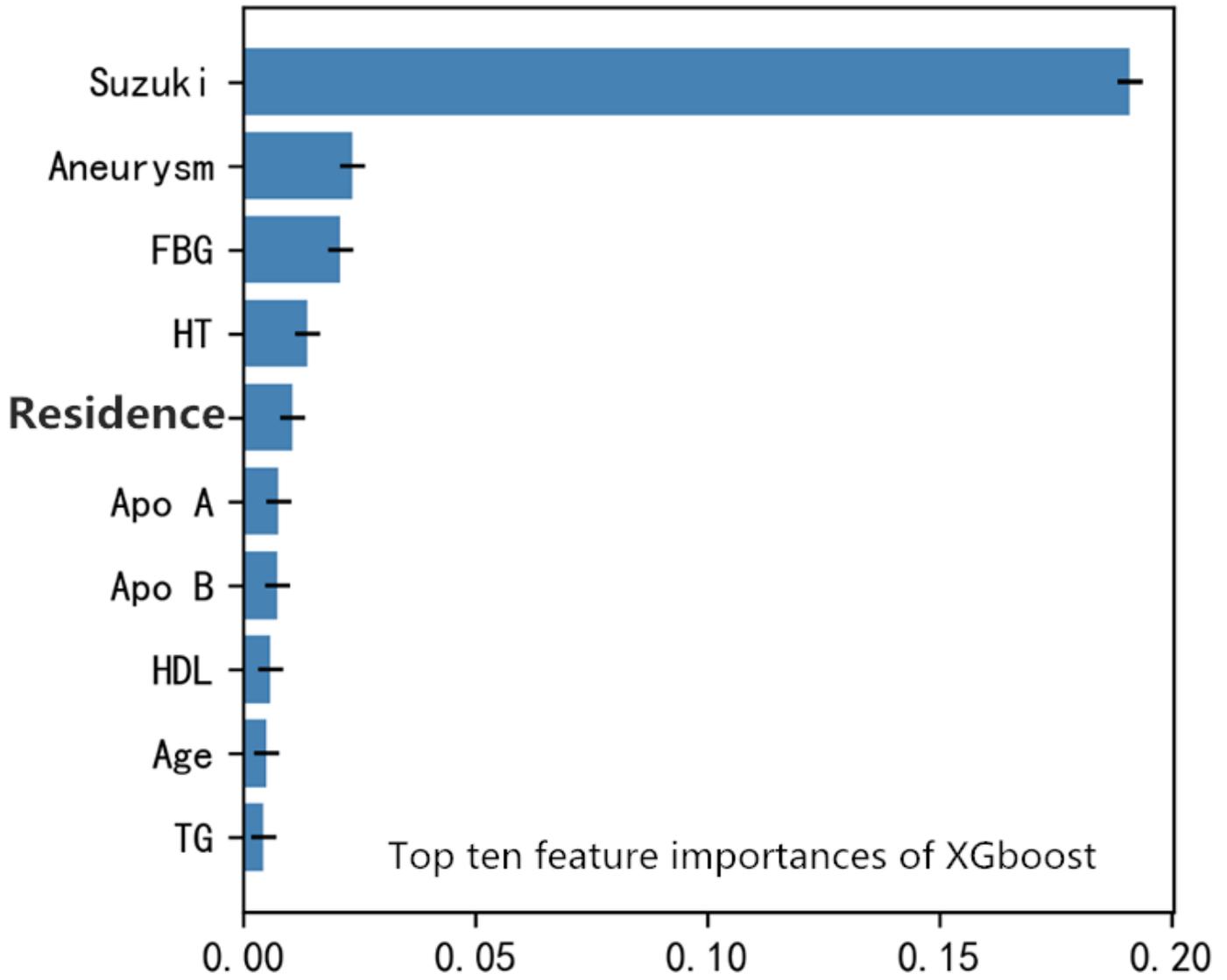


Figure 5

top ten variables of XGboost

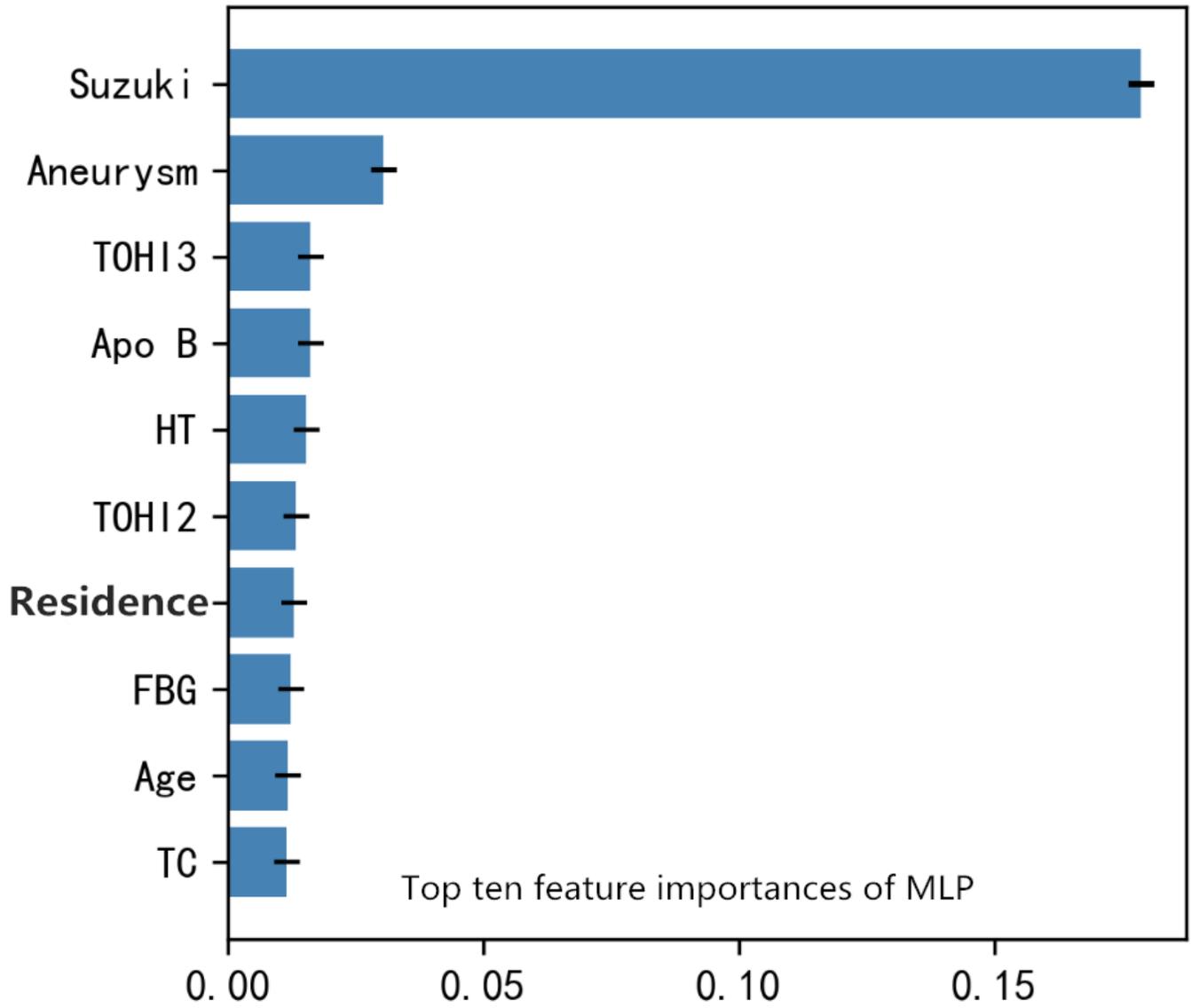
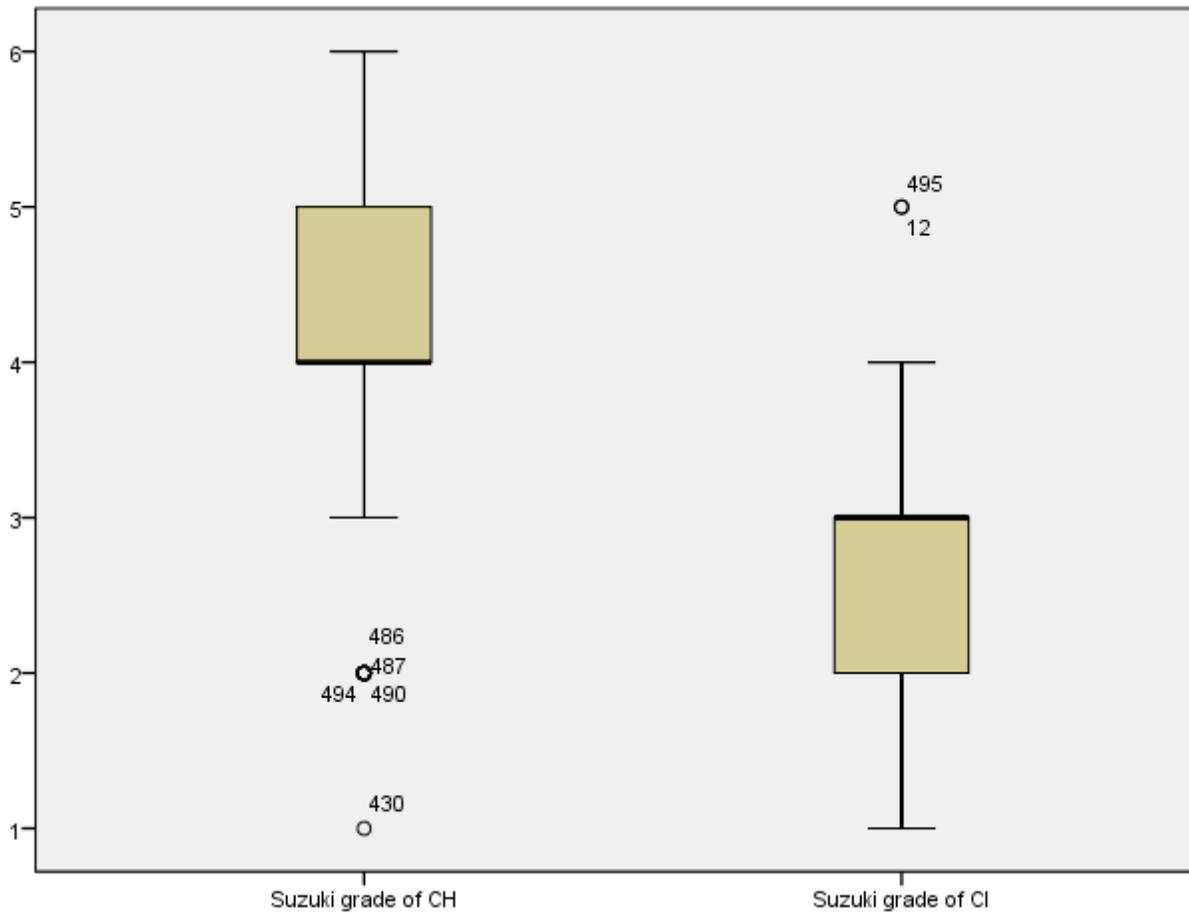


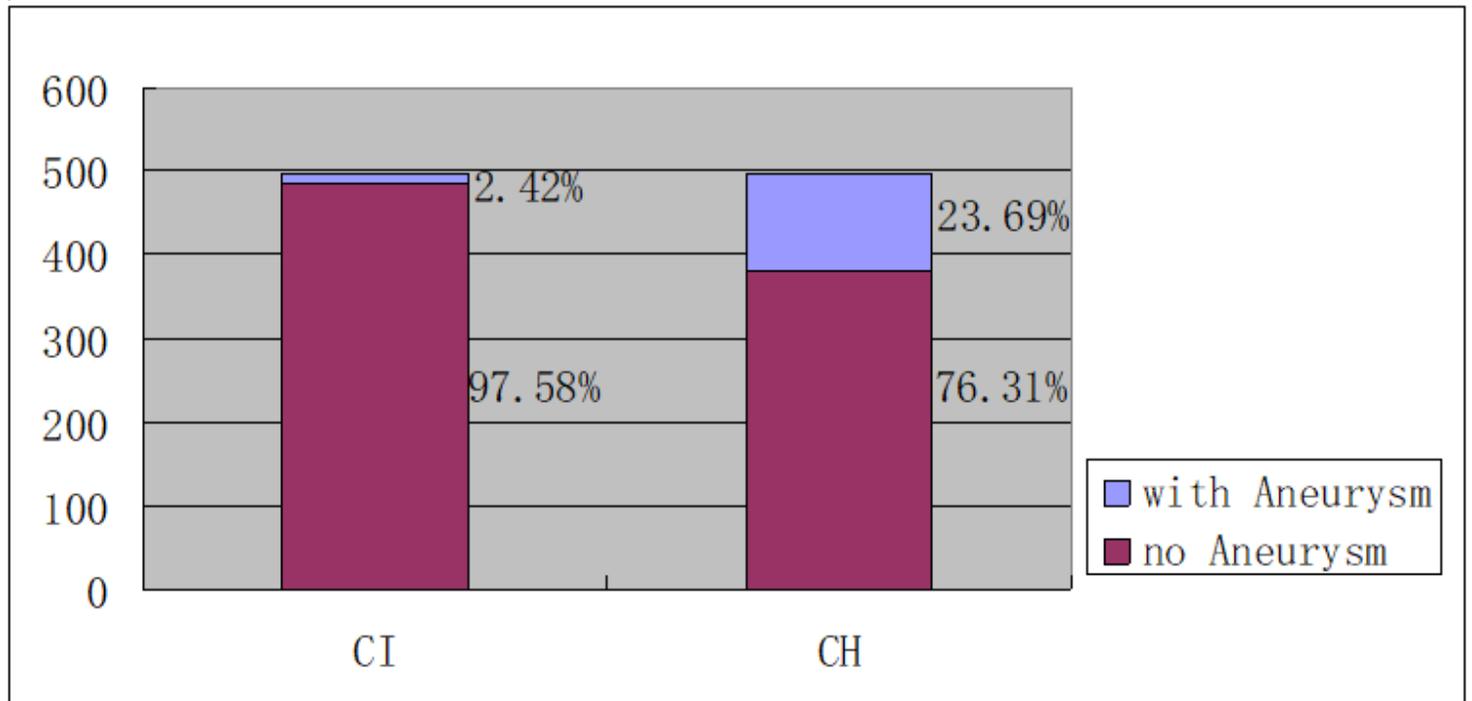
Figure 6

top ten variables of MLP



**Figure 7**

Suzuki grade Boxplot of two groups



## Figure 8

Aneurysm for two groups