Clinical Research Protocol

Project Proposal

Scheme name: Research on the Ischemic Stroke Recurrence Risk Prediction Model Using XGBoost Combined with Convolutional Neural Network Algorithm

Scheme number: 2021efyB03

Research types: Observational study

Source of funding: The Second Affiliated Hospital of Nanchang University

Version number: V 1.0

Applicant: Yi Yingping Phone: 13807089678

Unit: The Second Affiliated Hospital of Nanchang University

Contact address: No.1 Minde Road,Nanchang,Jiangxi Province,China

Email address: yyp66@126.com

Declaration date: April 29, 2021

The Second Affiliated Hospital of Nanchang University Clinical Research Center

一.Abstract of the Research Plan (should include research title, research objective, design type, research subjects, sample size, inclusion criteria, observation indicators, statistical analysis methods, etc.)

1. Research Title: Study on the Recurrence Risk Prediction Model of Ischemic Stroke Using XGBoost Combined with Convolutional Neural Network Algorithms.
2. Research Abstract:
Stroke is the leading cause of disability and death among adults in China, with an annual death toll of approximately 1.5 million and a disability rate of about 75%. About 70% of stroke patients suffer from ischemic stroke, and the recurrence rates at 3 months, 6 months, and 12 months after the first ischemic stroke are approximately 10.9%, 13.4%, and 14.7%, respectively. The risk of death and disability after the recurrence of ischemic stroke is 9.4 times that of the initial stroke. This project increases the sample size based on the existing ischemic stroke follow-up cohort and conducts a 1-year follow-up to observe the recurrence situation, mRs scores, and adverse events after discharge. Combining complex multi-dimensional data such as demographic data, physical examination, medical history, imaging images, medication information, scale scores, CYP2C19 genotype test results, laboratory examinations, etc.; using convolutional neural networks for lesion segmentation of ischemic stroke patient imaging and feature extraction; using Cox regression models to obtain factors influencing recurrence and adverse events; constructing an ischemic stroke recurrence risk prediction model within 1, 3, 6, and 12 months using the XGBoost combined with convolutional neural network algorithm (Convolutional Neural Network, CNN). Exploring the predictive effect of XGBoost on ischemic stroke recurrence risk at different times and providing new ideas and methods for the prevention and control of major chronic diseases.
3. Design Type: Cohort Study.
4. Research Subjects: Inpatients with ischemic stroke from the Department of Neurology at the Second Affiliated Hospital of Nanchang University and more than ten tertiary hospitals in the province.
5. Sample Size: 2400 cases.
6. Inclusion Criteria:
(1) Inclusion criteria:
1) Aged 18-85; 2) Diagnosed with ischemic stroke or transient ischemic attack (diagnosis in line with the standards set by the Cerebrovascular Disease Group of the Neurology Society of the Chinese Medical Association in 2014); 3) Within the acute phase of onset (2 weeks); 4) Voluntarily participate and sign an informed consent form.
(2)Exclusion criteria:
1)Patients with cancer; 2) Patients with cardiogenic infarction, other causes of cerebral infarction, or stroke of unknown cause; 3) Patients with hemorrhagic stroke, mixed stroke, and tumor stroke; 4) Patients with severe heart, lung, or liver system diseases; 5) Patients judged by researchers to have poor compliance and unable to complete long-term follow-up; 6) Patients currently participating in any clinical trial of investigational drugs or medical devices.
7. Observation Indicators:

Demographic data, physical examination, medical history, imaging images, medication information, scale scores, CYP2C19 gene test results, laboratory examinations, etc., mRS scores, recurrence situation (whether recurrent, number of recurrences, time of recurrence, type of stroke recurrence, etc.), and adverse events (type of adverse event, severity, start time, stop time, etc.) of ischemic stroke patients within 1, 3, 6, and 12 months after discharge.

8. Statistical Analysis Methods:

This study uses SPSS25.0 software for statistical analysis. Measurement data that conforms to the normal distribution are expressed as mean ± standard deviation ( ); Measurement data that do not conform to the normal distribution are expressed as median (quartile) [M(Q1, Q3)]; Count data are expressed as frequency and percentage (%). The comparison between count data groups uses the chi-square test or Fisher's exact probability method; the comparison between measurement data groups uses the t-test or Mann-Whitney U test. $P<0.05$ is considered statistically significant.

9. Modeling Methods:

CNN, Logistic, SVM, RF, XGBoost model construction is implemented using PyCharm (professional v2018.3) software under the Anaconda Python3.6 environment. Data preprocessing uses the numpy (1.20) and Pandas (1.2.0) packages, CNN image processing model construction uses the tensorflow (2.1.1) architecture, data analysis and model construction use Scikit-learn (0.23.1).

Brain imaging feature extraction: Introducing deep convolutional neural network models such as UNet++, FCN, AlexNet, etc., to process brain imaging images, while considering the natural 3D attributes of brain imaging images, using 3D convolution, pooling to capture changes in brain imaging between continuous slices; comparing the effects of different deep convolutional neural network models, selecting the optimal model as the method for extracting brain imaging features.

Prediction model construction: The variables screened by Cox regression with statistical differences are fused with the imaging features extracted by CNN. The fused data is divided into training sets, validation sets, and test sets, using a ten-fold cross-validation training method, respectively establishing machine learning-based ischemic stroke recurrence risk prediction models (research models: XGBoost; control models: Logistic, SVM, RF) within 1, 3, 6, and 12 months after discharge, continuously optimizing the model effects on the validation set until the optimal.

Model Evaluation: After the model effect reaches the ideal effect, compare the performance evaluation indicators of the research model and the control model on the test set (AUC value, sensitivity, time efficiency, etc.). AUC, sensitivity, and time efficiency are used to evaluate the comprehensive predictive ability, accuracy rate, and prediction efficiency of the model, respectively.

"Research on the Ischemic Stroke Recurrence Risk Prediction Model Using XGBoost Combined with Convolutional Neural Network Algorithm"

Informed Consent Form

Part One Subject Instructions

Research Introduction:

You are invited to participate in the project titled "Research on the Predictive Model of Recurrent Risk of Ischemic Stroke Using XGBoost Combined with Convolutional Neural Networks." You have been invited to join this study because you meet the research criteria.

This informed consent form provides you with some information to help you decide whether to participate in this clinical study. Your participation in this study is voluntary. If you agree to join this study, please read the following instructions.

Please read carefully, and if you have any questions, please ask the researcher responsible for the study.

Research Purpose:

Cerebrovascular disease is a serious illness that leads to death and disability, especially ischemic stroke, which has become a severe threat to human health, significantly affecting the quality of life of patients and imposing a heavy burden on families and society. The preferred treatment for preventing neurological dysfunction and recurrence of cerebrovascular disease, leading to further disability, is antiplatelet therapy. However, about 25%-60% of patients exhibit resistance to aspirin or clopidogrel, rendering treatment ineffective. Previous studies have found that CYP2C19 gene polymorphism and urinary 11dhTxB2 can effectively reflect resistance to commonly used clinical antiplatelet drugs and have a certain predictive effect on the occurrence and development of the disease. This study aims to collect various clinical data from subjects, including demographic information (age, gender, marital status, etc.), past medical history (previous stroke, hypertension, diabetes, dyslipidemia, etc.), laboratory tests (complete blood count, liver and kidney function, lipid profile, etc.), imaging studies (CT, MRI, etc.), clinical medication (aspirin, clopidogrel, etc.), clinical diagnosis (pulmonary infection, urinary tract infection, etc.), antiplatelet drug resistance testing (CYP2C19 gene and urinary 11-dhTxB2) results, neurological function assessment (NIHSS/mRS scores), follow-up data (adverse events, recurrence), etc., to explore the risk of recurrence in ischemic stroke and construct corresponding predictive models. This can provide new methods for predicting ischemic stroke recurrence and assisting in clinical decision-making for major chronic diseases.

Research process and methods:

Since you have been diagnosed with ischemic stroke, we invite you to participate and further verify whether you meet the inclusion criteria. Researchers will obtain your basic and disease-related information during your hospital stay and schedule follow-ups for you at 1 month, 3 months, 6 months, and 12 months after discharge to

understand your treatment and changes in your condition. If you agree to participate in this study, we will assign a number to each participant and establish a medical record.

Possible benefits of the study:

If you agree to participate in the study and meet the inclusion criteria of the research protocol, analyzing your medical records will help provide important basis for your individualized treatment, and also provide useful information for the study of the disease, ultimately bringing blessings to you and other patients in similar situations.

Research on risk and discomfort:

Privacy issues:

If you decide to participate in this study, your participation and personal information during the trial will be kept confidential. Information that can identify you will not be disclosed to anyone outside the research team without your permission. All members of the research team and the research sponsors are required to keep your identity confidential. To ensure that the study is conducted in accordance with regulations, members of the ethics review committee may, when necessary, review your personal information at our institution. When the results of this study are published, no personal information about you will be disclosed.

Free Exit:

As a participant, you can at any time access information and progress related to this study, and voluntarily decide whether to continue participating or not. After joining, regardless of whether any harm occurs or how severe it is, you can choose to notify the researcher at any time to request to withdraw from the study. Your data will not be included in the study results, and your medical treatment and rights will not be affected by this. If continuing to participate in the study would cause you serious harm, the researcher would also stop the study.

However, during the study period, please provide accurate information about your medical history and current health status: inform the study doctor whether you have recently participated in other studies or are currently participating in other studies. If you do not comply with the study plan, or if any injury related to the study occurs, or for any other reason, the study physician can terminate your participation in this study.

Contact information:

If you have any questions related to this study, or if you experience any discomfort or injury during the research, or if you have any concerns about the rights of participants in this study, you can contact the study doctor.

The trial protocol has been approved by the hospital ethics committee for implementation, and if there is any violation of the research protocol during the trial, participants can directly lodge a complaint with the hospital ethics committee. Contact number: 0791-86209562.

Part Two Informed Consent Signature

I have read this informed consent form, understood the purpose, content, risks, and benefits of the study, and I am aware of this clinical research. I voluntarily agree

to participate in this study.

Subject's name:_____Subject signature:_____

Signature date:_____Contact phone number:_____

Legal representative's name:_____Relationship to the subject:_____

Legal representative's signature:_____

Signature date:_____Contact phone number:_____

Researchers declare:

I have explained to the patient the details of this study, including their rights as well as potential benefits and risks, and have given them a copy of the signed informed consent form.

Researcher's name:_____Researcher's signature:_____