STUDY PROTOCOL

Official Title: Development and Validation of a Hemorrhage Risk Prediction Model for Moyamoya Disease Based on 5T MRI and Metabolomic Features: A Prospective Cohort Study

NCT Number: [To be inserted upon registration]

Document Date: November 17, 2025

English Translation Date: August 19, 2026

Ethics Approval Date: January 13, 2026

Sponsor: Beijing Hospital, Beijing, China

Funding: Qihang Special Project of the Central High-Level Hospital Clinical Research Project, Beijing Hospital (Grant No.: BJ-2025-222)

Principal Investigator: Shaosen Zhang, MD, Attending Neurosurgeon, Department of Neurosurgery, Beijing Hospital

Document Control: This document is a faithful English translation prepared on August 19, 2026 for ClinicalTrials.gov from the Chinese study protocol Version 2.0 dated November 17, 2025. The source protocol was approved by the Beijing Hospital Ethics Committee on January 13, 2026 (Approval No. 2025BJYYEC-KY400-02). Translation and file generation do not constitute a protocol amendment.

1. Study Background

Moyamoya disease (MMD) is a rare cerebrovascular disease characterized by stenosis or occlusion of the distal bifurcation of the internal carotid artery and the formation of collateral networks at the base of the brain. In China, the annual incidence between 2016 and 2018 was estimated at 1.14 per 100,000 population, showing a significant upward trend over the years. MMD presents predominantly as ischemic or hemorrhagic manifestations; hemorrhagic-type MMD has an annual hemorrhage rate of 17% and a mortality between 6.8% and 20%. Its prognosis is worse than that of the ischemic type, and it can lead to severe disability and even death. At present, the etiology of MMD-related hemorrhage remains unclear. Autopsy findings and imaging studies suggest that dilatation and rupture of collateral vessels are the likely cause of hemorrhage in MMD; it is therefore essential to clarify the mechanisms underlying the formation, expansion, and rupture of collateral vessels. Some studies suggest that metabolic changes in patients with MMD — such as alterations in amino acid metabolism involving lysine, L-methionine, L-glutamate, β-alanine, and O-phosphoserine — may participate in the development and progression of MMD. However, no study has analyzed the association between metabolic alterations and the hemorrhagic subtype of MMD; whether these metabolic changes are causally related to hemorrhagic outcomes, and whether they can serve as disease markers for the early identification of hemorrhagic-type MMD, remain unclear. Meanwhile, collateral vessel morphology, hemodynamics, and vessel wall signal changes have been shown to be significantly associated with hemorrhage risk in MMD, yet quantitative methods for assessing collateral vessel morphology, blood flow, and vessel wall characteristics are still lacking, preventing an in-depth exploration of the significance of collateral vessels in MMD-related hemorrhage events. The investigators' preliminary work has shown that glycine metabolism promotes angiogenesis and is associated with the formation and dilatation of moyamoya collateral vessels. In summary, this study proposes the following hypothesis: metabolic changes in patients with MMD, by regulating angiogenesis, drive the morphological changes — formation, dilatation, and rupture — of collateral vessels, thereby altering the hemodynamic state and leading to hemorrhage. Therefore, the purpose of this study is to establish a prospective clinical cohort of hemorrhagic-type MMD, to collect metabolic markers potentially related to the formation and alteration of collateral circulation in hemorrhagic-type MMD, and to simultaneously collect imaging features and hemodynamics of the collateral circulation based on 5.0-tesla magnetic resonance imaging (5T MRI). The study will then explore and elucidate the association between metabolic markers and imaging features in hemorrhagic-type MMD, and ultimately establish a long-term hemorrhage risk prediction model for MMD that integrates 5T MRI features and multi-omics metabolic features, enabling hemorrhage risk stratification of patients with MMD, identification of high-risk patients, and clinical translation.

MMD-related intracranial hemorrhage usually occurs in specific brain regions. According to one study, periventricular white matter is the most common site of cerebral hemorrhage in MMD, accounting for 63.3% of cases, and the basal ganglia/thalamus is the second most common site, accounting for 20%. The choroidal arteries are frequently regarded as the vessels responsible for periventricular hemorrhage in MMD. Early studies have shown that dilatation of the anterior choroidal and posterior communicating arteries predisposes individuals to hemorrhage in MMD. The development of collateral anastomoses among the lenticulostriate arteries, choroidal or perforating arteries, and medullary arteries is significantly associated with the hemorrhagic manifestation of MMD. One study evaluated the choroidal arteries and hemorrhage locations using time-of-flight MR angiography (TOF-MRA) and revealed a correspondence between the distribution of

choroidal collaterals and rebleeding sites. These findings indicate that dilatation and rupture of overcompensating collateral vessels are the main cause of cerebral hemorrhage in MMD. Autopsy results of hemorrhagic-type MMD show microaneurysms within collateral vessels, indicating that the collateral vessels formed as compensation in MMD are fragile, dilated, and thin-walled. It can therefore be inferred that pathological overcompensating collateral vessels and their fragility are related to hemorrhage risk in MMD. Reduced cerebral blood flow and impaired cerebrovascular reserve are risk factors for ischemic symptoms in MMD. By contrast, normal perfusion and focal hyperperfusion are more common in hemorrhagic cases. CT perfusion imaging shows that hemorrhagic-type MMD exhibits normal or even excessive perfusion, whereas ischemic-type MMD is characterized by hypoperfusion, indicating that blood flow in the collateral vessels of hemorrhagic cases is abundant and can ensure blood-flow compensation to ischemic brain regions. At the same time, other studies have reported that hemorrhagic-type MMD can also lead to cerebral ischemia, with lower cerebral blood flow and higher time-to-peak and mean transit time in the hematoma-bearing hemisphere and surrounding lateral temporal regions. One study has shown that hemodynamic failure is an independent risk factor for recurrent hemorrhage in hemorrhagic-type MMD. Taken together, hemodynamic studies in hemorrhagic-type MMD remain controversial, but hemodynamic changes are clearly associated with MMD-related hemorrhage. Intracranial high-resolution vessel wall imaging (HR-VWI) allows direct observation of the vessel wall, going beyond conventional angiography, which only shows the vessel lumen. It can detect small lesions in the vessel wall that may lead to hemorrhage, even before they cause changes in luminal morphology. Studies have shown that the “periventricular collateral enhancement sign” on vessel wall imaging is associated with a high risk of hemorrhagic stroke in adult MMD; however, that study was retrospective, with a small sample size and no follow-up. The associations of collateral vessel morphology, blood flow, and vessel wall characteristics with hemorrhage in MMD remain incompletely understood, and the current imaging assessment methods are relatively outdated. Limited by spatial resolution and signal-to-noise ratio (SNR), conventional 1.5T or 3.0T MRA has limited capability for accurately assessing the abnormal collateral circulation in patients with MMD. In addition, because vessel visualization in TOF MRA depends largely on the bottom-up inflow effect of unsaturated blood, this method poorly displays non-bottom-up or slow-flowing blood and cannot accurately characterize the collateral vessel morphology and hemodynamic changes of MMD. With advances in imaging technology, higher field strengths enhance SNR, enabling high-resolution imaging and better blood-tissue contrast. 5.0T MRI is an intermediate field strength that provides better B1+ field uniformity, lower radiofrequency energy deposition, and fewer specific-absorption-rate-related safety concerns, while still significantly outperforming 3.0T MRI in SNR. Recent studies have shown that the image quality of 5.0T TOF-MRA is comparable to that of 7.0T TOF-MRA and superior to that of 3.0T TOF-MRA. 3D arterial spin labeling (ASL) can provide information on brain perfusion downstream of the stenosis, and 4D imaging techniques can directly reveal leptomeningeal collateral vessels in patients with MMD. These technological advances make precise multi-dimensional assessment of collateral vessel morphology, blood flow, and vessel wall in MMD possible. Accordingly, the first key research focus of this study is to achieve quantitative analysis of collateral circulation formation in patients with MMD based on 5T MRI and to mine key imaging features.

Patients with MMD show potential metabolic biomarker alterations that are widely involved in the pathophysiological processes of the disease. One study using liquid chromatography–mass spectrometry (LC-MS) metabolomics identified 129 significantly different metabolites in the cerebrospinal fluid of patients with MMD compared with healthy controls, including metabolites related to purine and pyrimidine metabolic pathways. Gas chromatography–mass spectrometry (GC-MS)-based metabolomic studies

identified 25 different serum metabolic biomarkers in patients with MMD compared with healthy controls, including altered levels of amino acids, lipids, carbohydrates, and related metabolites. Targeted metabolomic analysis of serum amino acids showed that 12 amino acids were significantly different in patients with MMD, of which 4 potential amino acid biomarkers (L-methionine, L-glutamate, β-alanine, and O-phosphoserine) showed high diagnostic value for MMD. In summary, multiple metabolomic studies on different biological fluid samples (cerebrospinal fluid and serum) have consistently shown that patients with MMD have distinct metabolic profiles and biomarker signatures compared with controls. These metabolic alterations involve various pathways, such as purine/pyrimidine metabolism, amino acid metabolism, lipid metabolism, and inflammatory processes, indicating their participation in the pathogenesis and progression of MMD. Recent studies have proposed metabolomic signatures and biomarkers specifically related to pathological angiogenesis and inflammation in MMD, involving metabolic pathways related to cholesterol, vitamins, lipids, and immunomodulation, many of which participate in inflammatory and immunometabolic pathways that affect collateral vessel development and stability, indicating potential links between multiple metabolites and collateral vessels. However, no study has analyzed the association between metabolomic alterations and the hemorrhagic subtype of MMD, and whether these metabolic changes are causally related to hemorrhagic outcomes, or whether they can serve as disease markers for the early identification of hemorrhagic-type MMD, remains unclear. Accordingly, the second key research focus of this study is to identify, through metabolomic methods, the core markers that influence collateral circulation, and to establish their correspondence with imaging manifestations on 5T MRI.

Deep learning and machine learning techniques have been widely applied in the development of intracranial hemorrhage risk prediction models. For the task of predicting cerebral infarction in intracranial large-artery occlusive disease, the investigators previously constructed a deep learning model combining thin-slice non-contrast CT and CT angiography (CTA); by analyzing imaging data from 372 patients, the model achieved an area under the receiver operating characteristic curve (AUC-ROC) of 0.918 and a classification accuracy of 0.9032, demonstrating the potential of deep learning models in prognostic prediction for occlusive diseases. In addition, through previous work, the investigators built, for the first time, a cerebral hemorrhage prediction model based on CTA radiomics features and ensemble learning algorithms, achieving a ROC-AUC of 0.982 in the training set and 0.893 in the validation set — among the highest levels reported internationally for comparable prediction models — demonstrating the potential of machine learning methods in cerebral hemorrhage prediction. Accordingly, the third key research focus of this study is to construct, using deep learning and machine learning methods, a hemorrhage risk prediction model for MMD that jointly incorporates 5T MRI features and metabolomic features.

In summary, exploring disease markers of hemorrhagic-type MMD and establishing a highly predictive MMD hemorrhage risk prediction model carry important social and scientific significance. This study will build a prospective MMD cohort based on 5T MRI and metabolic multi-omics, analyze the imaging and metabolomic characteristics of hemorrhagic-type MMD, cross-validate metabolomic features with imaging features, screen core novel markers of hemorrhagic-type MMD, and construct a hemorrhage risk prediction model for MMD. Through this study, patients with MMD can be risk-stratified, and high-risk patients can be identified early so that surgical treatment can be undertaken in time to reduce the hemorrhage rate, thereby substantially reducing the disability and mortality of MMD, improving patient outcomes and quality of life, and carrying major significance for the comprehensive prevention and treatment of stroke in China and for reducing the family and social burden of hemorrhagic stroke.

2. Study Objectives

1. To extract the morphological and hemodynamic features of moyamoya collateral vessels based on 5T MRI, cross-validate them with metabolic features, and elucidate the core mechanisms by which metabolic markers relate to collateral vessel morphology and hemodynamic changes.
2. To screen and validate novel biomarkers of hemorrhagic-type MMD using metabolomic techniques, perform association analysis with imaging features, and cross-validate the biomarkers in superficial temporal artery (STA) specimens.
3. To construct and validate a hemorrhage risk prediction model for MMD by integrating core biological and imaging markers using deep learning and machine learning methods, and to establish a clinical risk stratification system.

3. Study Contents

(1) Prospective cohort study of hemorrhage risk factors in moyamoya disease

Relying on the large-sample MMD clinical research platform of Beijing Hospital, this study will use routine preoperative imaging examinations for MMD — including MRA, CTA/CTP, and DSA — together with conventional clinical assessments, to collect traditional risk factors of MMD, including demographic characteristics, biochemical indices, and assessments of intra- and extracranial large-vessel stenosis, collateral vessel compensation, and intracranial perfusion indicated by imaging. Based on the initial symptoms, patients with MMD will be classified into hemorrhagic-type and ischemic-type groups. The clinical and basic imaging features described above will serve as baseline data. No fewer than 500 patients with MMD will be consecutively enrolled, and their symptom changes will be followed up 1 year after study entry, using the participants' complaints and head CT or MRI findings at 1 year as the basis of assessment, with the cerebral hemisphere as the unit of analysis and new-onset hemorrhage as the endpoint event, to construct a prospective cohort of hemorrhage risk in MMD. Patients will be divided at a 7:3 ratio into a training set and a validation set.

(2) Exploration of 5T MRI-based imaging markers of hemorrhagic-type moyamoya disease

This study will construct an imaging database containing multimodal radiomics based on 5T MRI, including 5.0T high-resolution MRA, 4D MRA, high-resolution vessel wall MRI, 3D ASL, 3D T1WI, and conventional MRI (DWI, T2WI, SWI) sequences, and will perform preliminary image quality assessment and grading of the scanned data to build a complete imaging dataset. Based on 5.0T high-resolution MRA and 4D MRA, a semi-quantitative scoring method will first be established to score intracranial vessels and collateral vessels, including the internal carotid artery, anterior cerebral artery, middle cerebral artery, posterior cerebral artery, leptomeningeal collateral vessels, and moyamoya vessels. Because manual delineation of collateral vessels is difficult and difficult to standardize, a deep learning algorithm for extracting the vascular skeleton will be designed to quantitatively evaluate the number of distal vessels and collateral vessels and to compute the SNR. In addition, based on high-resolution vessel wall MRI, qualitative and quantitative assessments will be performed: qualitative indices include stenosis site, stenosis distribution, and vascular remodeling pattern; quantitative indices include stenosis degree, stenotic lumen length, maximum lumen cross-sectional area, wall thickness, and remodeling index. Combining the above, comprehensive assessment of intracranial large vessels, distal vessels, and collateral vessels will be achieved.

Furthermore, based on 3D ASL and conventional MRI sequences of the 5.0T system, the correlations of 5.0T vascular imaging features with brain perfusion and brain structure will be evaluated; perfusion and structural parameters include cerebral blood flow, cerebral blood volume, cortical thickness, cortical surface area, brain tissue volume, and deep nuclei volume. Baseline data and follow-up outcomes will be combined to explore the correlations of vascular features–structure and perfusion features with hemorrhage risk.

To screen the features most effective for predicting hemorrhage, feature selection and dimensionality reduction are required. This study will use L1 regularization for dimensionality reduction of the standardized training-set data, with the best ROC-AUC as the primary tuning objective. In this way, core imaging features related to MMD hemorrhage will be explored and refined, their associations with hemorrhage analyzed, and they will be used for subsequent model construction.

(3) Development and identification of novel metabolomics-based biomarkers of hemorrhagic-type moyamoya disease

This study will collect readily available preoperative clinical samples from patients with MMD, including blood and urine, to construct an MMD biobank. High-throughput metabolomic measurement techniques will be used to measure metabolic molecules in the preoperative clinical samples, establish a metabolic molecular profile of patients with MMD, and — combined with follow-up outcomes — screen novel biomarkers of hemorrhagic-type MMD, achieving metabolic subtyping of patients with MMD based on symptom phenotype.

In addition, clinical data will be used to analyze the associations between the identified biomarkers and imaging features, exploring their influence on collateral vessel morphology, blood flow, and vessel wall characteristics, to further validate the value of the biomarkers. Using STA specimens trimmed during MMD surgery, together with immunohistochemistry and related basic experiments, the mechanisms by which the biomarkers promote MMD-related hemorrhage will be elucidated and potential intervention targets provided. Rapid screening methods for readily available samples will be developed; novel biomarkers with potential value for hemorrhage risk assessment and prognosis will be summarized from the training set and further used in the subsequent construction of the MMD hemorrhage risk prediction model.

(4) Establishment and validation of the moyamoya disease hemorrhage risk prediction model

Based on the training set of the hemorrhagic-type MMD cohort described above, and combining clinical baseline data, radiomics features, and metabolomic markers, a deep learning network will be designed around the screened core metabolomic and imaging markers. Model construction will first test the performance of basic machine learning models; based on the analysis of basic machine learning models, the architecture of an ensemble learning model will be designed and constructed. Ultimately, a hemorrhage risk prediction model for MMD based on the prospective cohort and integrating metabolomic and imaging features will be established.

Internal validation of the prediction model will be performed by cross-validation, and external validation will be performed using the validation cohort. The primary performance metric of the final model is the ROC-AUC, with accuracy, recall, precision, and F1 score used for auxiliary performance assessment. Combining the core biological and imaging markers screened by multi-omics analysis, the clinical value of the core markers will be validated in the validation set, and a marker validation platform capable of hemorrhage risk stratification and prognostic value assessment in MMD will be built. The novel markers screened in this study will undergo whole-life-cycle clinical value assessment — covering onset risk, progression prediction,

risk stratification, treatment prognosis, and recurrence probability — and multicenter clinical promotion studies.

4. Study Design and Implementation

4.1 Study Population

4.1.1 Prospective cohort population selection

Five hundred patients with MMD will be consecutively enrolled and classified into hemorrhagic-type and ischemic-type groups according to their chief complaints and previous examinations. All patients will undergo routine multimodal head CT, head MRI, and DSA examinations, complete preoperative 5T MRI, and provide preoperative blood and urine samples. Symptom changes will be followed up within 1 year after study entry, with the cerebral hemisphere as the unit of analysis and hemorrhage as the endpoint event; the participant's complaints and head CT or MRI at the time of hemorrhage will serve as the basis for adjudicating hemorrhage.

4.1.2 Inclusion and exclusion criteria

Inclusion criteria:

1. Moyamoya disease or moyamoya syndrome diagnosed by digital subtraction angiography (DSA).
2. Able to undergo 5.0T magnetic resonance imaging (MRI).

Exclusion criteria:

1. Claustrophobia or other conditions preventing completion of imaging acquisition.
2. Revascularization surgery already performed in both cerebral hemispheres.
3. Any hemorrhage event within 1 month before enrollment.

4.1.3 Sample size calculation

Given the known annual hemorrhage rate of 17%, and assuming 5 predictor variables (potential markers) with 10 events (hemorrhage cases) required per variable, $5 \times 10 = 50$ events are needed. The total sample size N = required positive events / incidence rate = 50 / 17% ≈ 294. Assuming a 10% loss to follow-up, the adjusted sample size $N = 294 / (1 - 0.10) \approx 327$. With a 7:3 group split, the total sample size $N = 327 / 0.7 \approx 466$; therefore, approximately 500 patients with MMD are required.

4.2 Study Design

This study adopts standard machine learning model development and validation methods to establish a prospective cohort for hemorrhage risk in MMD, divided at a 7:3 ratio into a training cohort and a validation cohort. The training cohort will retrospectively include 325 eligible cases for model training and internal validation, and the validation cohort will prospectively include 175 eligible cases for external validation of the model.

4.3 Study Procedures and Follow-up Plan

4.3.1 Patient screening

Patients will be screened according to the inclusion and exclusion criteria described above.

4.3.2 Informed consent

The attending physician will explain to the patient and their relatives the potential risks and benefits of participating in this study, and will inform them that, whether or not they agree to participate, the patient will receive the "best available treatment." If the patient and their family agree to participate, the informed consent form will be signed. The above process should be witnessed by another physician or nurse.

4.3.3 Baseline imaging assessment and database construction

After enrollment, eligible patients will first undergo routine preoperative imaging assessment, including CTA/CTP, multimodal multi-sequence MRI, and DSA, to preliminarily evaluate the involved brain regions and the course and perfusion of intra- and extracranial vessels in MMD, construct the clinical imaging database of MMD, and upload the original images to the imaging platform for unified management.

4.3.4 5.0T MRI acquisition protocol

All MRI examinations will be performed on a 5T whole-body MR scanner (Siemens Healthcare, Erlangen, Germany). All scans will use a birdcage transmit coil and a 32-channel receive head coil (Nova Medical, MA, USA). The following data will be collected from each participant: T1-weighted magnetization-prepared rapid gradient echo (T1W MPRAGE), TOF-MRA, 3D ASL, and HR-VWI. Two TOF-MRA protocols were optimized to visualize fine collateral vessels. To assess the reproducibility of the method, a whole-brain isotropic TOF-MRA at 0.40 mm isotropic resolution (Gval) will be acquired with the following parameters: field of view (FOV) = 205 × 175 × 107 mm^3, 6 slabs, repetition time (TR) = 18 ms, echo time (TE) = 4.03 ms, flip angle (FA) = 22°, bandwidth (BW) = 158 Hz/Px, generalized autocalibrating partially parallel acquisitions (GRAPPA) factor = 3, acquisition time (TA) = 9 min 28 s. For Gapp, a single-slab TOF-MRA was optimized to shorten the acquisition time. The imaging slab will be placed in an oblique axial orientation, covering the middle cerebral artery (MCA) territory and the base of the basal ganglia region. Detailed parameters are: FOV = 180 × 135 × 52 mm^3, resolution = 0.23 × 0.23 × 0.36 mm^3, TR = 15 ms, TE = 3.57 ms, FA = 20°, BW = 151 Hz/Px, GRAPPA factor = 2, TA = 8 min 20 s. Structural images will be obtained with T1W MPRAGE using the following parameters: FOV = 224 × 224 × 179 mm^3, resolution = 0.70 × 0.70 × 0.70 mm^3, TR = 3000 ms, TE = 3.23 ms, inversion time = 1050 ms, FA = 8°, BW = 320 Hz/Px, GRAPPA factor = 3, TA = 5 min 54 s.

4.3.5 Biospecimen collection

Before surgery, 5 mL of venous blood will be drawn from each patient and centrifuged within 20 minutes to separate the plasma; 10 mL of urine will also be collected from each patient before surgery. All biospecimens will be frozen at −80 °C for long-term storage to construct the MMD biobank, for exploring novel biomarkers and performing clinical validation.

If a patient undergoes direct bypass surgery, the STA trimmed during the operation will be retained and stored in formalin for the long term. After paraffin embedding and sectioning, the tissue will be used for

immunostaining to cross-validate the biomarkers.

4.3.6 Extraction of collateral vessel morphology, hemodynamics, and vessel wall features from 5T MRI

This part of the study will use the TOF sequences of 5T MRI. Before network training, image data will be standardized by min–max normalization as data preprocessing. In the segmentation stage, the VT-UNet network will be introduced to reconstruct the three-dimensional structure of the collateral arteries; this process mainly completes skeleton extraction of the vessels. The Swin Transformer network architecture will then be applied to extract vessel features and generate predictions.

(1) Automatic collateral vessel segmentation and feature extraction

Vessel skeleton segmentation is mainly performed with a 3D semantic segmentation model adopting the VT-UNet network architecture. The model input is the TOF-MRA voxel data, with input dimensions of D × H × W × C. The inputs are fed into the downsampling modules of the UNet structure; the features are then fed into the VT encoder for feature extraction and merging; the VT encoder then feeds the features into the upsampling modules of the UNet; finally, the vascular skeleton with a 1 × 1 × 1 output dimension is produced.

The VT encoder adopts an enhanced Transformer multi-head attention module, composed of an input layer, an attention layer, and an output layer. First, an input layer composed of multiple linear layers (LN) collects features; this part draws on the characteristics of the ResNet structure, extracting features separately from the results of each downsampling step, feeding the features into the multi-head attention layer through matrix computation, and then outputting to an output layer composed of linear layers and a multilayer perceptron (MLP); the final output is added to the downsampling layers of the UNet structure.

After the above process, the entire skeleton extraction is completed and used for further feature analysis.

The feature extraction module is designed with reference to the Swin Transformer architecture. Two types of windowing operations are proposed: regular windows (W-MSA) and shifted windows (SW-MSA). Both use windowed attention layers followed by a two-layer multilayer perceptron with a Gaussian error linear unit (GELU) nonlinear activation function. Layer normalization is used before each MSA and MLP module, and residual connections are applied after each module. Windowing allows the model to inject inductive bias while modeling long-range dependencies between tokens; attention between tokens within a window facilitates representation learning, so features within a window can be extended to other windows.

A 3D patch-merging module is then used to generate a feature hierarchy in the encoder of VT-UNet. Such a hierarchy is crucial for producing finer output details in dense prediction tasks. After each VT encoder block, adjacent tokens are merged along the spatial axes in a non-overlapping manner to generate new tokens. To this end, the features of each group of 2 × 2 adjacent tokens are first concatenated; the resulting vector is then projected through a linear mapping into a space where the channel dimension of the tokens is doubled. The benefit of patch merging is not limited to the feature hierarchy: the computational complexity of self-attention is proportional to the square of the number of tokens, and therefore, after each VT encoder block, 3D patch merging reduces the FLOPs of VT-UNet by a factor of 16.

(2) Hemodynamic feature calculation

After the vascular skeleton is extracted by the above model, the three-dimensional skeleton model will be converted into a tetrahedral mesh, and simulation techniques will be used to compute parameters including

wall shear stress (WSS), wall shear stress gradient (WSSG), oscillatory shear index (OSI), and relative residence time (RTT).

(3) Vessel wall feature extraction

This part of the study mainly uses the high-resolution vessel wall imaging sequence of 5T MRI. To extract the vessel wall edge information as accurately as possible, manual annotation will be used in this part. Annotation will be performed on the high-resolution vessel wall imaging sequence using Slicer software (version 4.11; https://www.slicer.org). To ensure annotation reliability, all lesions will be annotated synchronously by the principal investigator and another auxiliary researcher in a blinded, independent ("back-to-back") manner. After annotation is completed, the intraclass correlation coefficient (ICC) will be used to compare the differences between the two sets of annotations for each feature; an ICC greater than 0.8 is considered reliable, and an ICC below 0.8 requires necessary corrections to the annotation. To ensure the stability of the basic study data, once the ICC is satisfactory, the principal investigator's annotations will be used as the study data, while two experts with more than 20 years of clinical experience in cerebrovascular disease and imaging diagnosis will perform annotation quality checks and make necessary corrections before the data are used in subsequent research.

Based on the generated segmentation masks, the PyRadiomics package in Python will be used to extract radiomic features. Within the group, the final segmentation results of the two researchers will be compared; features will be included in subsequent modeling only when the ICC exceeds 0.8. Features that are not applicable in the actual comparison process — for example, meaningless features whose values are all 1 or 0 — will be removed. Based on previous experience, 1,790 features are expected to be extracted, including 744 wavelet features. These features comprise 356 first-order features; 14 three-dimensional shape features; 432 gray-level co-occurrence matrix (GLCM) features; 312 gray-level size zone matrix (GLSZM) features; 312 gray-level run-length matrix (GLRLM) features; 90 neighborhood gray-tone difference matrix (NGTDM) features; and 274 gray-level dependence matrix (GLDM) features.

4.3.7 Follow-up plan

Patients in the prospective MMD hemorrhage risk validation cohort will be followed up 1 year after enrollment, with the cerebral hemisphere as the unit of analysis and hemorrhage as the endpoint event. Based on the patient's chief complaints and head CT or MRI examination, the occurrence of new cerebral hemorrhage will be adjudicated, and the time and location of each new hemorrhage will be recorded. For patients who undergo cerebral revascularization surgery during the follow-up period, the reason, side, and timing of the revascularization surgery will be recorded in detail.

4.3.8 Model development and validation

After multi-omics data extraction is completed, the data will be jointly constructed into a complete, high-throughput dataset containing multiple dimensions and multiple scales. The entire process includes data preprocessing, feature screening, dataset splitting, and model training.

(1) Data preprocessing and dimensionality reduction

All data require preprocessing, which will be performed in the Python environment. The data preprocessing process includes the following parts: normalization and standardization, which differ according to variable type. Categorical variables will be normalized by conversion to one-hot encoding; continuous variables will be transformed to normally distributed variables using the Box-Cox transformation. After normalization, all

data will be standardized by the z-score method. After standardization, dimensionality reduction will be performed through L1 regularization, and the resulting data will be included in the final model computation.

(2) Model training

To construct an interpretable machine learning model, this study will first test several basic algorithms, including logistic regression, naïve Bayes, support vector machine (SVM), neural network, k-nearest neighbor (KNN), and decision tree. If the results are unsatisfactory, the performance of ensemble learning algorithms on this task will be further tested. Hyperparameter tuning for all models will use grid search to enumerate and identify the optimal parameter combinations.

4.4 Study Evaluation Measures

4.4.1 Segmentation model evaluation

The main metrics used for model evaluation in this study are the loss value and the Dice coefficient.

The loss value measures the discrepancy between the model's predicted output and the actual labels. One of the goals of neural network training is to minimize the loss value so that the model better fits the data and improves its predictive ability. The loss function of this study is mainly adjusted on the basis of the cross-entropy function. The basic formula of the cross-entropy function is shown in Equation (1). The loss value is mainly used to evaluate model convergence and is computed on the training set.

Equation (1): $L(G, Y) = -(1/I) \times \Sigma_{i=1..I} \Sigma_{j=1..J} G(i,j) \cdot \log Y(i,j)$

The Dice coefficient is a metric for evaluating image segmentation quality, comparing the similarity between model segmentation and manual segmentation — that is, the similarity between the prediction and the ground-truth label, as shown in Equation (2), where $|A|$ denotes the number of voxels in the prediction and $|B|$ the number of voxels in the ground-truth label. The Dice coefficient is mainly used to observe model performance and generalization ability, and is therefore computed on the test set; a soft Dice loss is also used as part of the loss function to guide model training.

Equation (2): $Dice(A, B) = 2|A \cap B| / (|A| + |B|)$

During neural network training, the total number of epochs at model convergence will also be recorded. In each epoch, the neural network traverses the entire training dataset, performing forward and backward propagation for each sample to update the network parameters. Completion of an epoch means that the entire training dataset has been used for training once. The number of epochs required for convergence can be used to infer whether the model depth is appropriate.

4.4.2 Metabolomic analysis of preoperative blood and urine samples

High-throughput measurement of metabolic indices, mainly including amino acids.

4.4.3 Hemorrhage events within 1 year

With the cerebral hemisphere as the unit of analysis, the time of occurrence, location, and hematoma size of hemorrhage events within 1 year will be recorded during follow-up.

4.4.4 Hemorrhage prediction model evaluation

The model evaluation process includes internal validation and external validation. Internal validation will use 5 repetitions of 5-fold cross-validation (5 × 5-fold cross-validation), with the best ROC-AUC as the tuning endpoint of the grid search. External validation will be based primarily on the ROC-AUC, with accuracy, recall, precision, and F1 score as auxiliary evaluation parameters.

4.5 Statistical Analysis Methods

Statistical analyses will be performed using SAS 9.1 (SAS Institute, Cary, NC), SPSS 17.0 (SPSS Inc., Chicago, IL), and MedCalc 12.3 statistical software packages. Continuous variables will be presented as mean ± standard deviation or median (interquartile range); categorical variables will be presented as percentages. Comparisons of continuous variables between two groups will use the independent-samples t-test or the Mann-Whitney test. Comparisons of means among more than two groups will use one-way analysis of variance; comparisons of rates will use the chi-square test or Fisher's exact test. Correlation analysis will use the Pearson correlation coefficient. Multivariable regression will use the generalized estimating equation (GEE) model or logistic regression. The area under the receiver operating characteristic curve (AUROC) and the Hosmer-Lemeshow goodness-of-fit test will be used to evaluate imaging assessment performance.

4.6 Subject Withdrawal, Study Termination, and Risk Management

Subject withdrawal criteria:

1. The subject refuses to participate in the study.
2. The subject refuses follow-up.
3. The observed hemisphere receives craniotomy treatment during the follow-up period.

Study termination criteria:

1. The subject develops major dysfunction of other important organs during follow-up.
2. The subject has poor compliance and cannot adhere to follow-up.
3. The subject requires other surgical treatment due to worsening of the condition.

Risk management:

Study risks: (1) leakage of patient personal privacy; (2) trauma and infection caused by venous blood sampling.

Risk mitigation measures: (1) Except for study information published in an orderly arrangement by the research group, all personal information of patients will be kept confidential. Names, telephone numbers, contact information, and other basic information will not be disclosed to non-research personnel, organizations, unrelated third parties, or media. (2) Venous blood sampling will be performed in the treatment room of the hospital ward with properly disinfected medical dressings and instruments.

5. Ethical Considerations

This study will strictly follow the Declaration of Helsinki and the relevant laws and regulations on ethical review in China, and will be carried out strictly in accordance with the study protocol. Before the study

begins, the relevant materials must be submitted to the hospital ethics committee for review, and the study may proceed only after review and approval. Subjects must sign the informed consent form before participating in the study. During the study, any modification of the study protocol, informed consent form, or other documents must be re-submitted to the hospital ethics committee for review and may be implemented only after approval. When study information and data are published in scientific meetings or journals, the identity of the subjects will not be disclosed.