

DeepHemorrhageNet

-- An Imbalance Aware Deep Learning Model for Clinical Haemorrhagic Stroke Detection

Muchen Wang*

Shenzhen Foreign Languages School (Senior High Division), Shenzhen, Guangdong 518083, China

*Corresponding author: wmc090807@icloud.com

Abstract

Notable for its high mortality, and substantial clinical costs, Hemorrhagic stroke poses a severe global health threat. Our study proposes DeepHemorrhageNet, an imbalance-aware deep learning framework for early detection and risk assessment of hemorrhagic stroke, addressing key challenges including class imbalance in clinical data, subtle lesion detection, and integration of multimodal information. The framework consists of three core components: (1) A hybrid CNN-Transformer encoder-decoder architecture with Attention Gate (AG) mechanism, designed to extract multi-scale spatial features and model global context for precise segmentation of hemorrhagic lesions and classification of hemorrhage types; (2) An imbalance-aware module integrating Class-Balanced Focal Loss, Online Hard Example Mining (OHEM), and feature-space oversampling, which mitigates the impact of long-tailed data distribution; (3) An Enhanced Video Masked Autoencoder (E-VMAE) and Early Predictive Behavioral Recognition with Physiological Signals (EPBR-PS) algorithm for real-time detection of stroke-related abnormal behaviors by fusing temporal masking, local convolutional attention, and physiological signals. We also use a multimodal clinical dataset and behavioral video datasets for model training and validation. To prepare the data, preprocessing steps included Borderline SMOTE for class imbalance mitigation, VIF-based multicollinearity reduction, and standardized CT/image/video processing. Consequently, Experimental results showed that DeepHemorrhageNet outperformed state-of-the-art models on public datasets, achieving 83.2% mAP on THUMOS14, 54.6% mAP on ActivityNet1.3, and high accuracy (91.45% on test set) for behavioral recognition. For hematoma expansion, the prediction of logistic regression model achieved 84.8% accuracy, 82.1% recall, and 0.91 AUC on the test set. Additionally, a web-based prototype system integrating real-time monitoring, heart rate estimation, and risk assessment was developed, providing clinical decision support by alerting on high-risk events. In conclusion, DeepHemorrhageNet effectively combines multimodal data to enhance early detection and risk assessment of hemorrhagic stroke, demonstrating strong robustness, generalization, and clinical utility for improving patient outcomes and reducing healthcare burdens.

Keywords

Hemorrhagic stroke detection, deep learning, class imbalance, CNN-Transformer hybrid architecture, DeepHemorrhageNet, multimodal data fusion, medical image segmentation.

1. Introduction

Hemorrhagic stroke, which occurs when blood vessels rupture and bleed in the brain, silently causes the deaths of 3 million people worldwide each year (Krishnamurthi RV, et al., 2014) [1], with the acute case fatality rate ranging from 34 % to 50 % within 30 days (Halminen O, et al., 2025) [2] [3]. Furthermore, most of the survivors experience severe complications, including sensory loss, seizures, and cognitive impairment, which affect all aspects of daily life (Han Y, et al., 2022) [4] [5]. Compared to patients with other diseases, patients with hemorrhagic stroke require long-term care and incur high medical expenses, resulting in alarmingly high clinical costs. According to Centers for Disease Control and Prevention (CDC) (Hall M J., 2012) [6], the average cost exceeds \$30,000 per admission and can reach \$100,000 for total lifetime care in America. Mainly triggered by Brain tissue damage, the expansion of hematoma could lead to an increase in intracranial pressure, resulting in a poor prognosis. Therefore, early identification of hematoma expansion and edema occurrence can better guide care decisions and improve outcome predictions. In recent years, AI has been widely applied in the medical field, including detecting the expansion of hematoma and perihematoma edema. Renowned for its robustness and high accuracy, AI has the potential to eliminate human subjectivity and improve the efficiency of diagnosis and treatment. (Nazari-Farsani, et al., 2020) [7] introduced a method that can segment stroke lesions automatically, which can be applied into routine clinical diagnosis. Zhang et al. investigated the use of MRI machine learning models to classify the time of stroke onset, aiding treatment decisions (Man F Y, et al., 2022) [8]. Weber et al. utilized deep learning-based software to analyze rodent stroke recovery process (Tackenberg C, et al., 2020) [9]. Hung et al. proposed a new model by using deep Neural Network in predicting stroke's mortality based on the data collected from haemorrhagic and ischaemic stroke patients (Hung C Y, et al., 2017) [10]. Although modern medicine has become quite advanced, traditional diagnostic and monitoring approaches still face limitations, including subtle lesion presentation, imbalance in clinical data, and inefficient integration of multimodal information. In response to these issues, three major innovations are introduced in the proposed framework. At the imaging level, a CNN-Transformer hybrid encoder-decoder with Attention Gate mechanisms is incorporated to enable accurate lesion segmentation and hemorrhage type classification. To address the imbalance in clinical data, Class-Balanced Focal Loss, Online Hard Example Mining, and feature-space oversampling are integrated to enhance sensitivity to rare but clinically significant events. In addition, an Enhanced Video Masked Autoencoder (E-VM AE) is combined with an Early Predictive Behavioral Recognition using Physiological Signals (EPBR-PS) algorithm to detect early stroke-related abnormal behaviors. Specifically, through the integration of a Multi-Head Attention mechanism and a Local Convolutional Attention Feature Extraction Module, the algorithm captures fine-grained local dynamics while preserving contextual relationships. Building on these components, we further develop a multimodal prototype system that fuses visual cues with physiological signals to enable real-time monitoring and risk alerting. Experimental results show that the proposed framework outperforms other advanced baseline methods, thereby demonstrating its clinical significance.

2. Data Processing

This study proposes DeepHemorrhageNet, a hybrid encoder-decoder architecture that integrates Convolutional Neural Networks (CNNs) for local feature extraction with Transformers for global contextual modeling. Designed to achieve high-precision, robust detection in stroke CT scans, the model specifically improves sensitivity to small, low-contrast, and often missed hemorrhagic lesions. As illustrated in Figure 1, the architecture consists of four key components:

Let the input CT image be denoted as $X \in \mathbb{R}^{H \times W \times 1}$, where $H = W = 256$ is the preprocessed resolution. The image is first processed through a lightweight CNN backbone (in this study, ResNet-34) for initial feature extraction:

$$F_0 = \text{ConvBlock0}(X), F_0 \in \mathbb{R}^{H/4 \times W/4 \times C_0}, C_0 = 64 \quad (1)$$

Four residual blocks then generate a pyramid of multi-scale feature maps $\{F_1, F_2, F_3, F_4\}$, with spatial resolution progressively downsampled to $H/32 \times W/32$.

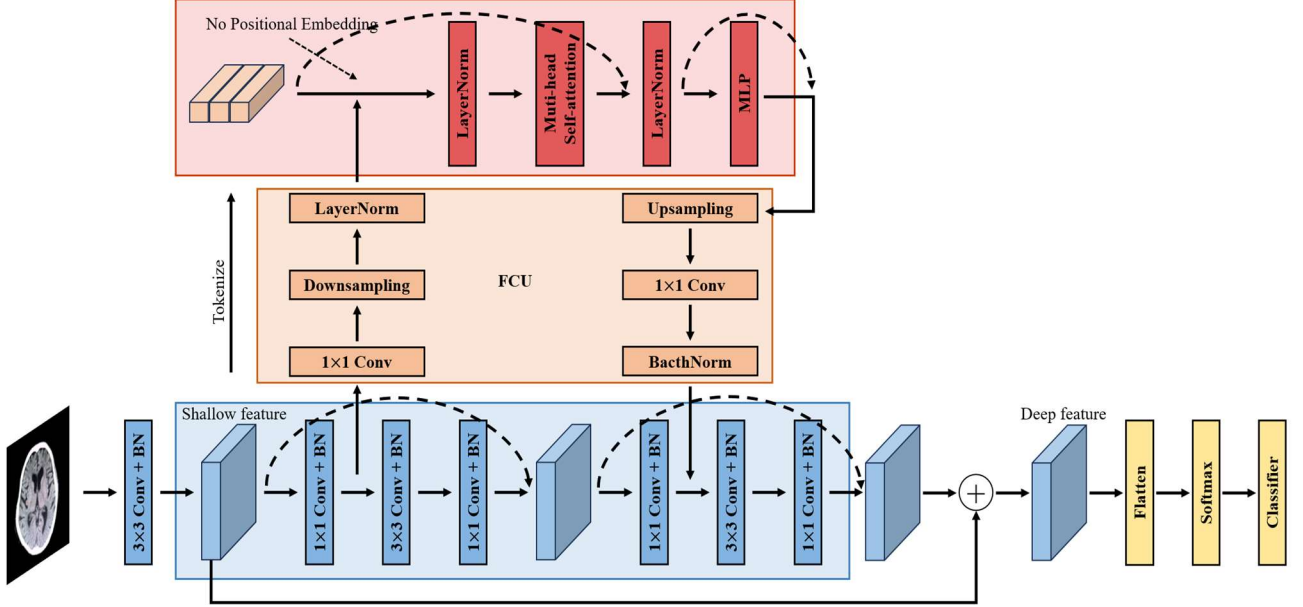


Figure 1. The architecture consists of four key components

To incorporate global contextual information, a Vision Transformer (ViT) encoder is integrated after the fourth stage. The deepest feature map $F_4 \in \mathbb{R}^{h \times w \times d}$ (with $h = w = 8, d = 512$) is flattened into a sequence:

$$Z_0 = [f_1, f_2, \dots, f_{hw}] + E_{\text{pos}} \in \mathbb{R}^{(hw+1) \times d} \quad (2)$$

where $f_i \in \mathbb{R}^d$ represents the embedding of the i -th patch, and $E_{\text{pos}} \in \mathbb{R}^{(hw+1) \times d}$ is a learnable positional encoding. An additional [cls] token is appended for the subsequent classification task. The ViT encoder comprises L standard Transformer encoder layers:

$$Z_i = \text{TransformerBlock}_i(Z_{i-1}), i = 1, 2, \dots, L \quad (3)$$

Traditional skip connections often propagate background noise in sparse lesion scenarios. To address this, we introduce an Attention Gate (AG) mechanism that dynamically focuses on potential hemorrhage regions during decoding.

Let $F_i \in \mathbb{R}^{H_i \times W_i \times C_i}$ be the encoder feature at level i , and $U \in \mathbb{R}^{H_u \times W_u \times C_u}$ the current high-level decoder feature. First, U is upsampled to $U' \in \mathbb{R}^{H_i \times W_i \times C_i}$, and channel dimensions are aligned via 1×1 convolutions:

$$\mathbf{g} = \mathbf{W}_g * \mathbf{U}'_{\geq}, \mathbf{x} = \mathbf{W}_x * \mathbf{F} \quad (4)$$

where σ denotes the Sigmoid function, $[\cdot]$ indicates element-wise addition, and $*$ represents convolution. The final gated feature is computed as:

$$F_{att} = \alpha \odot F_i \quad (5)$$

In hemorrhagic stroke analysis using medical images and behavioral videos, most data belong to common categories such as intraparenchymal hemorrhage or frequent actions like “sneezing,” while events such as intraventricular hemorrhage, small hematomas (< 5, mL), or critical behaviors like “falling” are rare. This severe class imbalance causes models to be biased toward majority classes and results in poor performance in detecting rare cases.

To address this challenge, we introduce an imbalance-aware module that enhances the model's sensitivity to minority classes. The module integrates class-balanced loss functions, dynamic sampling strategies, and feature-level reweighting mechanisms.

The feature information extracted by the two-stream I3D model from the video only integrates local spatiotemporal information, failing to consider the relationships between adjacent frames, which affects the performance of temporal action detection methods. Thus, we propose a global feature encoder based on an enhanced local temporal relationship attention mechanism (ELTRAMHA) that can promote feature information interaction between neighboring frames. At its core, multi-head attention mechanism is employed to enhance local temporal relationships.

The feature matrix F goes through one-dimensional convolution to obtain feature matrices Q , K , and V . Then, the feature matrix K is added to a linearly transformed version of K , which enhances the global feature information carried by the K matrix. Subsequently, it performs matrix inner product operations with Q to obtain the feature interaction matrix A . The formulas are shown in equations (6), (7), (8).

$$Q = \text{Conv1d}^q(F) \quad (6)$$

$$K = \text{Conv1d}^k(F) \quad (7)$$

$$V = \text{Conv1d}^v(F) \quad (8)$$

Here, Conv1d^q , Conv1d^k , Conv1d^v represent one-dimensional convolutions with a kernel size of 5 and stride of 2. And Norm , Linear , and SoftMax correspond to batch normalization, linear transformation, and the softmax function for normalization, respectively. Next, the feature matrix V is added to a depthwise separable convolution processed matrix V , enhancing the local feature information carried by V , resulting in a rich V' matrix containing global feature information.

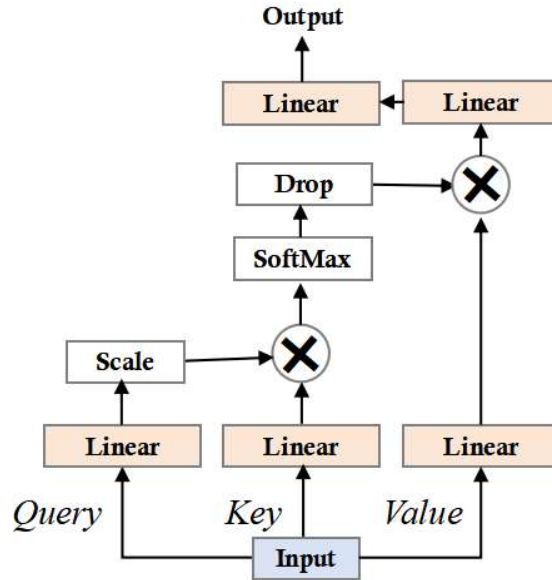


Figure 2. The advanced temporal relationship attention mechanism-based global feature encoder

3. Analysis of Empirical Results

The occurrence of hematoma expansion events is considered as the dependent variable, while patient personal history, disease history, onset and treatment-related information, and the results of the first imaging examination are the independent variables. Among these, gender is a categorical variable, and blood pressure needs to be divided into systolic and diastolic pressure. We use one-hot encoding to convert categorical variables into numerical form, where “1” represents male and “0” represents female. The results of the preprocessing are shown in Table 1.

Table 1. Preprocessing Mapping Table

Patient ID	Gender	Blood Pressure	Gender	Systolic BP	Diastolic BP
sub001	Female	180/90	0	180	90
sub009	Male	199/122	1	199	122
...
sub159	Male	183/122	1	183	122
sub160	Female	208/95	0	208	95

Data on the imaging examination times and edema volumes for patients from sub001 to sub100 was collected³², resulting 450 data, as shown in Table 2.

Table 2. Imaging Examination Times and Edema Volumes

Patient ID	Imaging Examination Count	Imaging Examination Time (h)	Edema Volume (mL)
sub001	1	2.5	48.919
	2	8.27	57.898
	3	132.11	81.747
	4	259.74	107.793
	5	425.54	126.558
...
sub100	1	0.5	7.367
	2	34.73	6.033
	3	197.24	3.191

First, a curve fitting on the imaging examination times and edema volumes using cubic and quartic polynomial functions is performed. The results are shown in Figure 3.

According to the graph, the curve fluctuates more greatly when the image examination time increases. Thus, common methods such as the boxplot method and the 3-sigma rule are applied to identify and exclude outlier samples with extended examination times for further reduction of the influence on the model. To verify the normality of the data-a prerequisite for applying the 3-sigma rule-we analyzed the data distribution and its probability density function, as shown in Figure 4.

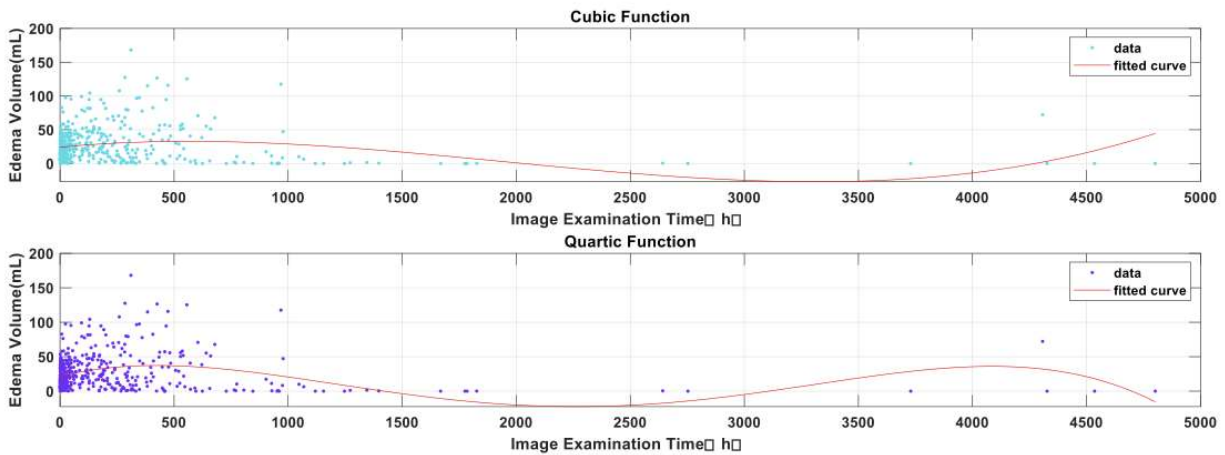


Figure 3. Curve Fitting Results of Unprocessed Data

Examination times for further reduction of the influence on the model. To verify the normality of the data-a prerequisite for applying the 3-sigma rule-we analyzed the data distribution and its probability density function, as shown in Figure 4.

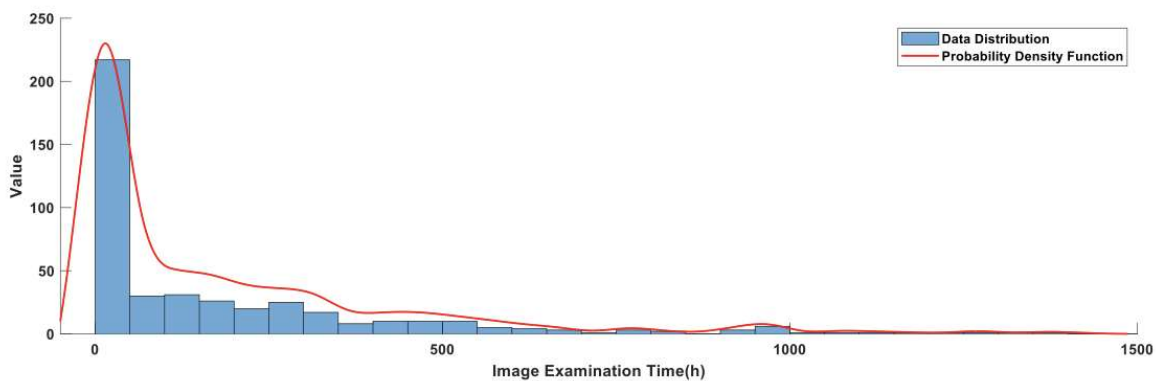


Figure 4. Data Distribution and Probability Density Function

The data distribution of the imaging examination times is primarily concentrated within 24 hours. Since image examination times are not normally distributed, we considered using a boxplot method to detect outliers in Figure 5. Raw CT images often suffer from non-uniform intensity distribution, noise interference, and inter-scanner variability, which may significantly impair the model's ability to extract meaningful features. This study designed a standardized medical image preprocessing pipeline in response to the question. First, a windowing operation was applied to enhance the visibility of hemorrhagic regions.

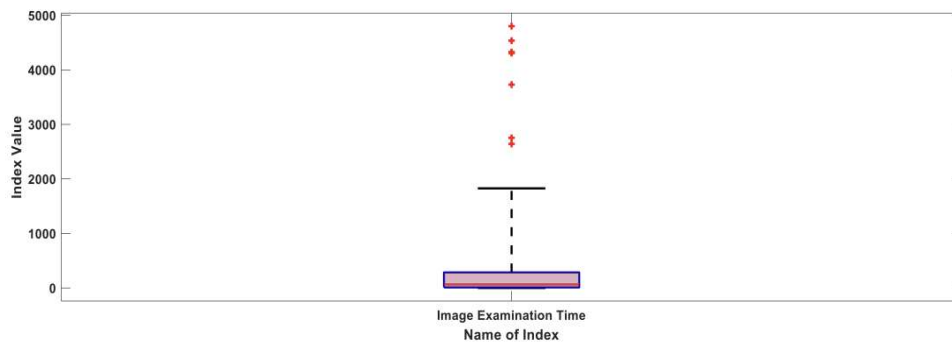


Figure 5. OutlierDetectionUsingBoxplot

To address the need for early stroke risk assessment, this chapter introduces a prototype system that integrates heart rate and visual features. The system includes several key modules: real-time monitoring, validation functionality, heart rate estimation, risk assessment, and user information storage.

The monitoring module continuously tracks user behavior using an already proposed Enhanced Predictive Behavioral Recognition (EPBR-PS), an enhanced WMAE-based method, to determine abnormal behaviors. Meanwhile, the validation functionality module ensures the reliability and stability of the system by verifying the anomalies detected by the EPBR-PS model. Additionally, the validation functionality module can also utilize the fusion temporal relationship and contextual information from the Early Temporal Relationship Detection (ETRD) method proposed. By offering the outputs of the start and end boundaries of detected behaviors, the method provides a basis for calculating the duration of these behaviors, which is crucial for subsequent early stroke risk assessments.

Moreover, the user information module stores users' basic details and physiological signals, which offers the comprehensive data support from heart rate and behavioral vision perspectives for more accurate early stroke risk assessments. Below, we investigate the various information modules of the early stroke risk assessment prototype system based on heart rate and visual features.

The real-time monitoring module captures user behaviors in real time. We then classify the acquired video actions by applying the improved VideoMAE-based method for early predictable behavior recognition of stroke. The classification results are then stored in a MySQL database and displayed on users' interface , providing a data foundation for the system to analyze user behaviors. Specifically, OpenCV captures real-time video frames, the EPBR-PS model processes them to generate classification results, and these results are saved in MySQL. Meanwhile, using Python's Flask backend, Ajax technology, and HTML technology, the recognized action results are sent to the front-end and displayed in a table on the webpage. The real-time monitoring module captures user behaviors in real time. We then classify the acquired video actions by applying the improved VideoMAE-based method for early predictable behavior recognition of stroke. The classification results are then stored in a MySQL database and displayed on users' interface , providing a data foundation for the system to analyze user behaviors. Meanwhile, using Python's Flask backend, Ajax technology, and HTML technology, the recognized action results are sent to the front-end and displayed in a table on the webpage.

The verification module can validate abnormal behavior videos stored in the MySQL database. The system provides two model algorithms: the EPBR-PS behavior recognition model and the ETRD behavior detection model. The EPBR- PS behavior recognition model performs data processing, encoding-decoding, and classification on the video to output action classification results, which enables the system to review abnormal behavior videos and improve system

robustness. The ETRD behavior detection model performs temporal relationship modeling and action localization on video frames, providing the action classification as well as the start and end boundaries of the action, which are used to calculate its duration. The workflow proceeds as follows: First, the user uploads a video and the system acquires the original video. Second, the user selects different models to process the video. Finally, the system outputs the final result, and if any abnormal behavior is detected, the video is then stored in the database.

4. Conclusion

In conclusion, this study proposes the DeepHemorrhageNet for the early detection and risk assessment of hemorrhagic stroke. We propose DeepHemorrhageNet to address the critical challenges of hematoma expansion and perihematomal edema progression, and incorporate the E-VMAE and EPBR-PS algorithms for early stroke behavior recognition. Using methods such as VIF and Borderline-SMOTE, we preprocess the data and enhance the performance of the logistic regression model. A comprehensive risk assessment prototype system that combines heart rate monitoring (through rPPG and wearable sensors) with behavioral analysis is also developed. The system integrates multiple modules, including monitoring, verification, heart rate estimation, and risk assessment, into a unified web-based platform. When tested on public datasets, DeepHemorrhageNet outperforms state-of-the-art models in both segmentation accuracy and behavioral recognition, confirming its robustness. In summary, DeepHemorrhageNet is a reliable, efficient solution for early hemorrhagic stroke detection and risk assessment, with the potential to lower mortality and reduce clinical costs.

References

- [1] Krishnamurthi R V, Moran A E, Forouzanfar M H, et al. The global burden of hemorrhagic stroke: a summary of findings from the GBD 2010 study[J]. *Global heart*, 2014, 9(1): 101-106.
- [2] Tiili P, Lehto M, Halminen O, et al. Hemorrhagic stroke in atrial fibrillation: Trends in incidence, case fatality, and prior oral anticoagulation[J]. *Journal of the American Heart Association*, 2025, 14(12):
- [3] Sacco S, Marini C, Toni D, et al. Incidence and 10-year survival of intracerebral hemorrhage in a population-based registry[J]. *Stroke*, 2009, 40(2): 394-399.
- [4] Guo X, Zhong R, Han Y, et al. Incidence and relevant factors for seizures after spontaneous intracerebral hemorrhage: a systematic review and meta-analysis[J]. *Seizure: European Journal of Epilepsy*, 2022, 101: 30-38.
- [5] Kazim S F, Ogulnick J V, Robinson M B, et al. Cognitive impairment after intracerebral hemorrhage: a systematic review and meta-analysis[J]. *World neurosurgery*, 2021, 148: 141-162.
- [6] Hall M J. Hospitalization for stroke in US hospitals, 1989-2009[M]. US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics, 2012.
- [7] Nazari-Farsani S, Nyman M, Karjalainen T, Bucci M, Isoja...rvi J, Nummenmaa L. Automated segmentation of acute stroke lesions using a data-driven anomaly detection on diffusion weighted MRI[J]. *Journal of Neuroscience Methods*, 2020, 333: 108575.
- [8] Zhang Y Q, Liu A F, Man F Y, et al. MRI radiomic features-based machine learning approach to classify ischemic stroke onset time[J]. *Journal of Neurology*, 2022, 269(1): 350-360.
- [9] Weber R Z, Mulders G, Kaiser J, Tackenberg C, Rust R. Deep learning-based behavioral profiling of rodent stroke recovery[J]. *BMC Biology*, 2022, 20(1): 232. doi:10.1186/s12915-022-01434-9.
- [10] Hung C Y, Chen W C, Lai P T, et al. Comparing deep neural network and other machine learning algorithms for stroke prediction in a large-scale population-based electronic medical claims database[C]//2017 39th annual international conference of the IEEE engineering in medicine and biology society (EMBC). IEEE, 2017: 3110-3113.