Classification of Stroke Disease Assessment based on Body Surface Electrical Signals at Acupuncture Points

Hu Wang*, Xin’an Wang**, Xing Zhang, Changpei Qiu, Qiuping Li and Tianxia Zhao

The Key Laboratory of Integrated Microsystems, Peking University Shenzhen Graduate School, Shenzhen, China

*whsece@pku.edu.cn; **anxinwang@pku.edu.cn

Abstract. Stroke is a common brain disease with high sudden onset, high lethality, and a single means of rehabilitation assessment. In this paper, we designed a novel stroke condition rehabilitation classification experiment using a 4-channel body surface electrical signal acquisition circuit independently developed by the Integrated Microsystems Laboratory of Peking University to acquire body surface electrical signals from acupoints on both sides of the body meridians of 8 healthy volunteers and 16 stroke patients with different degrees of rehabilitation. The signals were filtered and wavelet transformed to extract features, which were fed into LR, RR, XGBoost, and SVM models for condition assessment and classification (types: healthy, mild hemiplegia, and severe hemiplegia). The results showed that the experimental scheme had strong classification ability for stroke condition recognition, with F1-score of 0.75 for LR, 0.72 for RR, 0.81 for XGBoost, and 0.79 for SVM. Furthermore, the “anti-interference ability of 50Hz” feature was separately introduced for classification, and a better classification effect was obtained. Among them, the F1 score of stroke evaluation and classification based on the SVM model increased by 2.65%, which is helpful to realize the intelligent prediction of stroke disease in clinical practice Diagnostic function.

1. Introduction

Stroke is a disease caused by impaired cerebral blood circulation and is divided into ischemic stroke and hemorrhagic stroke depending on the pathogenesis [1]. Ischemic stroke is a brain dysfunction caused by insufficient blood supply due to narrowing of cerebral blood vessels, also known as cerebral infarction, which accounts for more than 70% of all stroke causes. Hemorrhagic stroke is a brain dysfunction caused by rupture and bleeding of cerebral blood vessels, also known as cerebral hemorrhage [2]. According to statistics, the mortality rate of stroke among Chinese residents accounted for 22% of the national resident morbidity and mortality rate in 2018, and stroke became the leading cause of death and disability among adult patients in China [3].

Classification of stroke patients and assessment of rehabilitation efficacy are important components of modern rehabilitation medicine research, but there are still many problems in clinical application. The long-used assessment methods include the modified Ashworth evaluation method and the freehand muscle strength grading evaluation, but because the assessment results are mostly based on the empirical observations of physicians, and even the diagnoses given by different experts may differ, the traditional assessment methods are questionable and limited in the classification of disease levels. How to make the assessment results have real-time detectability and objectivity, provide a quantitative stroke
classification scheme for patients and doctors, and improve the accuracy of stroke diagnosis is of great significance in accelerating the development of stroke intelligent diagnosis.

Figure 1. The conceptual structure of the digitally controlled oscillators

The acquisition and processing of physiological data is the key technology of wearable intelligent diagnostic system. Body surface electrical signals such as ECG and EEG have advantages such as flexible and non-invasive acquisition methods. Studies have shown that human acupuncture points have low electrical impedance and meridians represent channels with low impedance [4]. Considering that stroke patients often have normal motor function on one side of the body and weak motor function or even loss on the other side, the assessment results are highly correlated with the selection of body parts. Therefore, in this paper, the body surface electrical signals of "Sanjian" and "Zhouliao" points marked with red dots on the left and right limbs of the patient were collected to evaluate the rehabilitation effect of stroke patients, as shown in Figure 1. After filtering the signals, wavelet decomposition was used to obtain wavelet features in different sub-bands, and the wavelet features were combined with time-domain frequency-domain features to predict the rehabilitation degree of stroke patients using four algorithms: LR, RR, XGBoost and SVM, which were classified into three levels (healthy, mild hemiplegia and severe hemiplegia). It was also found that the "anti-interference ability of 50Hz" improved the performance of stroke patients, providing a new way to assess the degree of recovery of stroke patients.

The main novelties of this paper include: (1) The first algorithm based on acupuncture point body surface electrical signal acquisition and analysis was proposed as a new index for stroke rehabilitation classification and efficacy. (2) We found that the "anti-interference ability of 50Hz" was more effective for assessing stroke patients' disease grade, and the classification performance of all models improved after introducing this feature. This finding provides a new idea for stroke condition assessment index.

2. Experimental methods and signal pre-processing

2.1 Signal Acquisition
In this experiment, a 4-channel acquisition circuit independently developed by the Integrated Microsystem Science Engineering Laboratory of Peking University was used. Collect the body surface electrical signals at the "Sanjian" acupoints and "Zhouliao" acupoints on the hand-yangming large intestine meridian on the left and right arms of the subject. Each subject obtained the electrical signals of the biological surface at 4 acupoints at the same time, which was used to compare the rehabilitation efficacy of hemiplegic patients.

2.2 Experimenters
In this experiment, 8 healthy volunteers from Peking University Shenzhen Graduate School (4 males, 4 females, aged 20-29 years) and 16 stroke patients (8 males, females) from the Acupuncture
Rehabilitation Department of Shenzhen Hospital of Traditional Chinese Medicine 8 cases, aged 32-64 years old) were used as experimental subjects. All volunteers signed an informed consent form before the experiment, and all subjects were tested at 1 hour after a meal without strenuous exercise. During the measurement, all subjects kept quiet, dropped metal objects and turned off all electronic devices. The collection process is shown in Figure 2. When collecting data, the common electrode of differential input is located on the right knee, the reference electrode of common ground is located on the left elbow, the sampling frequency is 1000hz, and the acquisition time is 5 minutes.

![Figure 2. Data collection process of healthy volunteers, mild hemiplegia patients and severe hemiplegia patients](image)

### 2.3 Signal preprocessing

Considering the length of the original data segment, the original data was first cut into a sample set of 3000 data points per segment before the experiment. For the 4-lead acupoint body surface electrical signal, a sample set corresponds to 3000 columns and 4 rows of data.

Next, we filter the collected electrical signals of the acupuncture points and body surfaces, and use a notch filter to remove 50Hz power frequency interference and 100Hz and 150Hz Gaussian noise. Furthermore, we filter out signals below 1Hz to eliminate baseline drift.

For the filtered data, we use the wavelet packet decomposition method to refine and analyze the signal in multiple scales to improve the stability of the low-frequency time domain signal and the high-frequency domain signal.

![Figure 3. Comparison of filtering effects of time-domain waveforms](image)

Figure 3 shows the original time-domain waveform and filtered waveform comparison of the body surface electrical signal at the "Sanjian" acupoint. Figure 4 shows the original frequency domain
waveform and the filtered waveform comparison of the body surface electrical signal at the "Sanjian" acupoint.

![Original frequency domain diagram](image1)

![Frequency domain diagram after filtering](image2)

**Figure 4.** Comparison of filtering effects in frequency domain

It can be seen from the figure that because the power supply frequency in my country is 50Hz, the power frequency interference caused by it is relatively large, and the Gaussian noise at 100Hz also interferes greatly with the signal. The signal after filtering and denoising has higher fidelity and smoothness than the original signal, which restores the waveform of the original signal well, and improves the accuracy of subsequent feature extraction.

3. Feature extraction and algorithm introduction

3.1 Time domain features

Time domain features can characterize the contour shape and dimensions of the signal waveform through time as a dimension, using statistical laws [5]. The time domain features used in this paper include: maximum value, mean value, variance, standard deviation, root mean square, kurtosis, over-zero rate, skewness, etc.

**Variance (Var):** It reflects the extent to which the sample deviates from the mean. The larger the variance, the greater the difference in waveform amplitude. Its formula is as follows.

$$\text{Var} = \frac{\sum_{i=0}^{N} (x_i - \bar{X})^2}{N}$$  \hspace{1cm} (1)

**Mean value (MEAN):** It provides a benchmark for signal waveform evaluation and reflects the basic size of the signal waveform. Its formula is as follows.

$$\text{MEAN} = \frac{\sum_{i=1}^{N} x_i}{N}$$  \hspace{1cm} (2)

**Zero crossing rate (ZR):** It can indicate the number of signal waveforms crossing the zero point per unit time, and the larger the zero crossing rate is, the faster the signal waveform changes. Its formula is as follows.

$$ZR = \frac{\sum_{i=1}^{N-1} f(i)}{T}, \quad f(i) = \begin{cases} -1, & x_i x_{i+1} < 0 \quad |x_i - x_{i-1}| > th \\ 0, & \text{else} \end{cases}$$  \hspace{1cm} (3)
Skew: It reflects the rotation deviation angle of the signal waveform and can evaluate the signal symmetry. Its formula is as follows:

$$\text{Skew} = E \left( \frac{x_{i-(x)}}{\sigma} \right)^3$$ \hspace{1cm} (4)

Kurtosis: It reflects the sharpness of the peaks of the signal waveform and describes the statistics of high or low rate of change of the distribution pattern of all values in the overall. Its formula is as follows:

$$\text{Kurtosis} = E \left( \frac{x_{i-(x)}}{\sigma} \right)^4$$ \hspace{1cm} (5)

3.2 Frequency domain features

In order to extract more comprehensive information about the electrical signal of the cavity body surface, we Fourier transform the signal to obtain its energy spectrum distribution (frequency domain features), which in turn serves as an in-depth complement to the time domain features [6]. Common statistical indicators are frequency band power, Shannon entropy, signal-to-noise ratio, etc.

High Frequency Ratio (HFR): It reflects the percentage of the high frequency part of the signal in the frequency domain. Its formula is as follows:

$$HFR = \frac{\sum_{i=n}^{m} P_{\text{high}}}{\sum_{i=1}^{N-1} p_i}$$ \hspace{1cm} (6)

Low frequency ratio (LFR): It reflects the percentage of the low frequency part of the signal in the frequency domain. Its formula is as follows:

$$LFR = \frac{\sum_{i=1}^{k} P_{\text{low}}}{\sum_{i=1}^{N-1} p_i}$$ \hspace{1cm} (7)

Shannon entropy (SE): It is often used to express the uncertainty of the information in the signal, and the larger the Shannon entropy is, the greater the uncertainty of the information contained in the signal. Its formula is as follows:

$$SE = -\sum_{i=1}^{N} p(X_i) \log_2 p(X_i)$$ \hspace{1cm} (8)

Signal-to-noise ratio (SNR): It is often used to characterize the ratio of useful information to noise in a signal. Its formula is as follows:

$$SNR = \frac{s}{n}$$ \hspace{1cm} (9)

3.3 Wavelet characteristics

Wavelet analysis is known as a "mathematical microscope" because it can amplify, reduce and translate the signal [7]. Wavelet analysis can also project the signal in multiple dimensions to obtain different sub-bands of the signal, which is of high significance for refining the signal and improving the classification accuracy [8].

To extract the electrical signal from the body surface of the acupuncture point, we perform wavelet decomposition on the filtered signal and calculate the wavelet coefficients of the corresponding nodes.
In this paper, we choose to perform 4-layer wavelet decomposition on the signal to obtain the wavelet coefficient vector \( S \), whose structure is 
\[
S = \{ sA_1, sA_2, sA_3, sD_1, sD_2, sD_3, sD_4 \}
\]
where \( sA_i \) represents the approximation coefficient, \( sD_i \) represents the wavelet coefficient.

The next step to extract the indicators corresponding to different wavelet coefficients, including energy entropy, high frequency ratio(HFR), medium frequency ratio(MFR), low frequency ratio(LFR), higuchi fractal a total of five statistical features, so each original signal contains a total of 40 dimensional wavelet features, the specific structure is shown in Figure 5.

**Figure 5.** The four-layer wavelet decomposition structure chosen in this paper

### 3.4 Algorithm Introduction

Logistic regression (LR) is a generalized linear regression analysis model that is commonly used to solve classification problems and can directly model classification probabilities without prior assumptions about the data distribution, reducing the problems caused by inaccurate assumptions about the distribution [9]. Logistic regression does not only predict the categories, but also yields approximate probability predictions. The goal of logistic regression is to maximize the probability that each sample belongs to its true marker by estimating the model parameters by the method of great likelihood.

Ridge regression (RR) is a biased estimation regression method, which is essentially a modified least squares estimation method. The overfitting problem that occurs in linear regression is solved by giving up the unbiased nature of the least squares method, losing some information and reducing the accuracy at the cost of obtaining more realistic regression coefficients [10].

XGBoost is an integrated learning model framework based on the gradient boosting algorithm. It is optimized on the basis of Adaboost and GBDT algorithms to improve the accuracy and computational speed of the algorithm [11].

Support vector machine (SVM) is a classifier that classifies data according to a supervised learning approach. Its decision boundary is a maximum margin hyperplane solved for the learned samples, and the goal of the SVM is to make the hyperplane as large as possible from the boundary where the support vector is located. Regular problems are often complex and nonlinear, so kernel functions are often invoked to map the features to a high-dimensional space and then solve them. Common kernel functions include linear kernels, Gaussian kernels, polynomial kernels, sigmoid kernels, etc[12].

### 4. Discussion of results

Among the 24 subjects collected in this experiment, we acquired a total of 9742 acupuncture point body surface electrical signals of 3s duration, including 3212 data from healthy volunteers, labeled as healthy; 3302 data from mild stroke patients, labeled as mild hemiplegia, and 3228 data from severe stroke patients, labeled as severe hemiplegia.
The above data were filtered and features were extracted, and the standard library functions in scikit-learn were called to feed the features into logistic regression model (LR), Ridge regression model (RR), XGBoost model, and SVM model, respectively.

The parameter search optimization was performed using the table search method, and finally the accuracy, recall, and F1-score were used as the model training effect criteria using K-fold cross-validation, and the classification results are shown in Figure 6 and Table 1.

**Table 1.** Comparison of LR, RR, XGBoost and SVM classification performance.

|       | Accuracy | F1-score |
|-------|----------|----------|
| LR    | 0.673    | 0.75     |
| RR    | 0.752    | 0.72     |
| XGBoost | 0.801   | 0.81     |
| SVM   | 0.753    | 0.79     |

**Figure 6.** Comparison of F1-score and accuracy of various classification models

By comparison, we found that the XGBoost model had the highest F1-score in this study, reaching 0.81, which was higher than LF, RF, and SVM. Therefore, we recommend using the XGBoost model for stroke rehabilitation classification based on acupuncture point body surface electrical signals in the future.

During the experiment, we found that there is a certain difference in the anti-interference ability of 50Hz power frequency interference between the hemiplegic side and the non-hemiplegic side of stroke patients. The specific effect is shown in Figure 7.

In order to further explore the contribution of this feature to the improvement of the recognition rate of conventional machine learning models, we send the composite feature set containing the “anti-interference ability of 50Hz” feature to logistic regression (LR), Ridge regression (RR), XGBoost, and SVM models for training. And compared with the original results, the results are shown in Table 2.
Table 2. Comparison of LR, RR, XGBoost and SVM classification performance.

|                | LR   | RR   | XGBoost | SVM  |
|----------------|------|------|---------|------|
| Original features | 0.751| 0.723| 0.818   | 0.792|
| Composite features | 0.762| 0.741| 0.831   | 0.812|
| Increased percentage | 1.46%| 2.49%| 1.59%   | 2.65%|

Figure 7. Comparison of filtered 4-channel frequency domain diagrams of healthy volunteers and stroke patients

The results of the comparison showed that the addition of “anti-interference ability of 50Hz” feature had a better effect on solving the classification of stroke patients’ conditions, with the F1-score of the SVM algorithm increasing the most, reaching 2.65%.

The reason for this is that the bioimpedance and electrical potential of the hemiplegic side of the stroke patient are different from the normal side due to the lack of exercise and long-term lack of real-time control of the central brain.

5. Conclusion

In this paper, we propose a novel way to classify stroke conditions by collecting body surface electrical signals at the “Sanjian” and “Zhouliao” acupoints on the hemiplegic and non-hemiplegic sides of the hand-yangming large intestine meridian in healthy subjects and stroke patients.

We filtered and decomposed the above signals, extracted the features, and fed the feature set into logistic regression (LR), Ridge regression (RR), XGBoost, and SVM models, and achieved good prediction results for stroke patient classification. The F1score of XGBoost model reached 0.81, which provides a new solution to the problems of "single test scheme" and "strong subjectivity in judgment" in the assessment of stroke patients.

Further, we tried to add “anti-interference ability of 50Hz” feature to the feature set fed into the model and found that the performance of each model for stroke patient condition classification was improved to different degrees, with the F1score of the SVM model improved by 2.65%, and the experimental results also validated the conclusion that “anti-interference ability of 50Hz” is more sensitive to the classification of stroke patient condition based on acupuncture point body surface electrical signals in existing studies. The experimental results also validate the conclusion that “anti-interference ability of 50Hz” is more sensitive to the classification of stroke patients based on body surface electrical signals.

Based on the current findings of this paper, we will optimize three aspects in the future.

From the experimental aspect, we will optimize the experimental protocol and consider using devices with higher conductance numbers for whole-body meridian acupuncture point body surface electrical signal extraction for more comprehensive signal acquisition.
From the feature level, we will remove the features with small contribution, streamline the feature dimension, and continue to explore the effective features, for example, we can introduce neural network features for feature fusion.

From the model level, in the future, we can try to use recurrent neural networks (RNN) that deal with temporal information, as well as models such as LSTM [13] that retain important temporal information based on the traditional RNN model by setting up a feedback loop mechanism between adjacent temporal units.

**Acknowledgments**

This work was supported by the Special Fund for the Shenzhen (China) Science and Technology Research and Development Fund (JCYJ20180503182125190) and Shenzhen (China) Science and Technology Research and Development Fund (JCYJ20200109120404043).

**References**

[1] G. A. Donnan, M. Fisher, M. Macleod, and S. M. Davis, "Stroke,", Lancet, vol. 371, no. 9624, pp. 1612-23, May 2008.

[2] Yang X, Yu H, Zhang T, et al. The Effects of Jin’s Three-needle Acupuncture Therapy on EEG Alpha Rhythm of Stroke Patients[J]. 2019, 25(7):1-5

[3] Bivard, A., Churilo, L. & Parsons, M. Artificial intelligence for decision support in acute stroke — current roles and potential. Nat Rev Neurol 16, 575–585 (2020).

[4] Charles Shang. “Electrophysiology of growth control and acupuncture,” Life Sciences, vol.68, pp. 1333-1342, 2001

[5] S. A. Shufani and M. Y. Mashor, "ECG signals classification based on discrete wavelet transform, time domain and frequency domain features," 2015 2nd International Conference on Biomedical Engineering (ICoBE), Penang, 2015, pp. 1-6, doi: 10.1109/ICoBE.2015.7235914.

[6] A. Sethi, S. Arora and A. Ballaney, "Frequency Domain Analysis of ECG Signals using Auto-Associative Neural Networks," 2006 International Conference on Biomedical and Pharmaceutical Engineering, Singapore, 2006, pp. 531-536.

[7] A. M. Ciupe, R. V. Ciupa, A. I. Roman and N. M. Roman, "Study of continuous wavelet transform used in ECG pre-diagnostic," 2015 E-Health and Bioengineering Conference (EHB), Iasi, 2015, pp. 1-4, doi: 10.1109/EHB.2015.7391509.

[8] Ö. F. Karaaslan and G. Bilgin, "ECG classification with empirical mode decomposition denoised by wavelet transform," 2014 22nd Signal Processing and Communications Applications Conference (SIU), Trabzon, 2014, pp. 694-697, doi: 10.1109/SIU.2014.6830324.

[9] X. Zou, Y. Hu, Z. Tian and K. Shen, "Logistic Regression Model Optimization and Case Analysis," 2019 IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT), Dalian, China, 2019, pp. 135-139, doi: 10.1109/ICCSNT47585.2019.8962457.

[10] C. Yu, F. Gao and Q. Wen, "An improved quantum algorithm for ridge regression," in IEEE Transactions on Knowledge and Data Engineering, doi: 10.1109/TKDE.2019.2937491.

[11] J. Bao, "Multi-features Based Arrhythmia Diagnosis Algorithm Using Xgboost," 2020 International Conference on Computing and Data Science (CDS), Stanford, CA, 2020, pp. 454-457, doi: 10.1109/CDS49703.2020.00095.

[12] Y. Bhatia, A. Mittal, S. Athavale, T. Mohanani and G. Bhatia, "Classification of Cardiac Arrhythmia using Kernelized SVM," 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184), Tirunelveli, India, 2020, pp. 922-926, doi: 10.1109/ICOEI48184.2020.9143000.

[13] S. Saadatnejad, M. Oveis, and M. Hashemi, "LSTM-Based ECG Classification for Continuous Monitoring on Personal Wearable Devices," in IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 2, pp. 515-523, Feb. 2020, doi: 10.1109/JBHI.2019.2911367.