A Classification Method of Acupoints and Non-acupoints based on Traditional Features and Wavelet Features

Xin Zhang*, Xing Zhang†, Xin’an Wang, Qiuping Li, Changpei Qiu and Zhong Liu
The Key Laboratory of Integrated Microsystems, Peking University Shenzhen Graduate School, Shenzhen, China

e-mail: 1801213337@pku.edu.cn; zhx@pku.edu.cn*

Abstract. Meridians and acupoints are the basis of TCM theory and play an important role in disease diagnosis and acupuncture treatment. There are still many problems in current research on electrical signals of acupoints. On the one hand, most of the studies did not consider the integrity of the meridian, but only based on a few acupoints. On the other hand, the lack of targeted feature extraction and classification methods leads to unsatisfactory classification results. Considering the above problems, a method combining traditional features and wavelet features is proposed to classify acupoints and non-acupoints. Based on the integrity of the meridians, we first collect the body surface electrical signals of some acupoints and non-acupoints on the twelve meridians of the human body, and then extract traditional and wavelet features from the measured signals. Finally, SVM and XGBoost are used to classify acupoints and non-acupoints respectively. The experimental results show that this method can effectively improve the classification performance of acupoints and non-acupoints, and for the feature vectors constructed in this paper, XGBoost has better classification capabilities.

1. Introduction

In recent years, the problems existing in the western medicine therapy, such as chemical drug resistance and increase in the cost of medicare, cannot be neglected any more. Traditional Chinese Medicine (TCM) has attracted more and more attention because of its safety, convenience and effectiveness. However, compared with the western medicine, the application of TCM is constrained in many areas due to the lack of objective and scientific basis in meridian theory [1] and the absence of a clear and independent anatomical structure.

Acupuncture points are the basis of traditional Chinese medicine theory, and their states can reflect the physical health of the human body. Therefore, conducting research from the perspective of electricity, analyzing their electrical characteristics, and distinguishing them from non-acupuncture points, can help to achieve precise positioning of acupuncture points. And it can also provide similar research ideas for disease diagnosis and treatment through acupoint electrical signals, which has certain significance for supporting the theory of Chinese medicine.

The current research on the electrical characteristics of meridian points is mainly divided into two directions: one is to find out the different characteristics between acupoints and non-acupoints under the condition of applying electrical stimulation, and some results have been achieved. For example, a large number of experimental results show that the acupuncture points have low impedance and high capacitance characteristics [2]. The other is to study the potential characteristics of acupoints without...
stimulus, and it has been concluded that acupuncture points have higher potential characteristics than non-acupuncture points [3]. Compared with the method of applying electrical stimulation, the study of acupoint potential without stimulation directly collects the electrical signals spontaneously generated by the human body for analysis, which avoids the influence of external stimulation, so it may be closer to the essence of the electrical characteristics of acupoints [4]. However, there are still many problems in current research on acupoint potential, such as:

1) Most studies are only based on the potential characteristics of a few acupoints and non-acupuncture points on one meridian. They have not studied the similarities and differences between acupuncture points and non-acupuncture points on multiple meridians based on the integrity of the meridian.

2) The lack of targeted feature extraction and classification methods results in low classification accuracy.

3) There are relatively few studies on the electrical signals of acupuncture points and non-acupuncture points without stimulation.

In order to solve the above problems, this paper collects the electrical signals of some acupoints and non-acupuncture points on the twelve meridians of the human body, based on the integrity of the meridian, considers both the global and local features of the signals, and finally uses machine learning methods to classify acupoints and non-acupoints.

The rest of this article is organized as follows. Section 2 briefly introduces the overall plan of this research. Section 3 mainly introduces the data processing process. Section 4 shows the experiments and results. Conclusion and future work are given in Section 5.

2. The overall plan

The overall plan of this article is shown in Figure 1. Firstly, we collect the raw signals of acupoints and non-acupoints on the twelve meridians of the human body, and then denoise the raw signals. Next, we extract traditional features and wavelet features from the de-noised signals. Traditional features consider the overall information of the signals and describe the overall characteristics of the signals, while wavelet features consider the local information of the signals and describe the local characteristics of different sub-band signals. Finally, machine learning methods are used to classify acupoints and non-acupoints.

![Diagram](image)

**Figure 1.** The process of the classification of acupoints and non-acupoints based on traditional features and wavelet features.

3. Data processing

3.1. Data collection

In this experiment, we used an 8-channel acquisition circuit independently developed by the Life and Health Group of The Key Laboratory of Integrated Microsystems Peking University. Some acupuncture
points and some non-meridian non-acupuncture points 2cm near the acupuncture points on the twelve meridians were collected sequentially. All measured points are on the left body of the subject. At each measured point, we collected a 1000hz, 150s-long signal. A group of subjects includes 7 healthy volunteers (20-29 years old, 3 males, 4 females). For a single subject, we obtained the raw signals of 47 acupuncture points and 38 non-acupuncture points.

3.2. Data pre-processing
For a single 150s long signal, we cut it every 3s as an acupoint or non-acupoint signal data. Among the 7 subjects, we selected the collected signals of 6 subjects (3 males, 3 females) as the training set data, and took the collected signals of the last subject as the test set data. In the training set data and the test set data, the total amount of acupuncture point and non-acupoint point samples is balanced.

We first observed the frequency domain of the collected signal and found that the 50hz power frequency and its 100hz harmonic interference in the collected signal frequency domain were relatively large, so we used the notch filter to remove the 50hz and 100hz power frequency interference [5]. For the signal whose power frequency interference has been removed, we used wavelet decomposition filtering techniques [6] to remove other noises. As show in figure 2, comparing with the original signal, the denoised signal eliminate useless noise on the basis of retaining the useful signal to the maximum extent, which is very beneficial for feature extraction in the following work.

![Figure 2. Raw signal and de-noised signal: (a) Raw acupoint signal and de-noised acupoint signal in Shàoshāng(LU11). (b) Raw non-acupoint signal and de-noised non-acupoint signal near Tiānfǔ(LU3).](image)

3.3. Traditional features
Traditional features include time domain features, frequency domain features, and nonlinear features. For a denoising signal, this paper extracts 42-dimensional traditional features.

3.3.1. Time domain features. We perform various operations on the amplitude of the signals to extract the time domain features which can intuitively measure the change of the measured signal's amplitude and shape. The statistical characteristics used in this paper are: count value, average value, maximum, minimum, range, variance, skewness, kurtosis, etc. The method for calculating the variance is given by equation (1), where $X_i$ is the value at each time, $N$ is the length of the sequence and $\langle x \rangle$ is the mean of the sequence.

$$\text{VAR} = \frac{\sum_{i=1}^{N}(X_i - \langle x \rangle)^2}{N-1} \quad (1)$$

The calculation method of the kurtosis shown in equation (2) can characterize the sharpness of the data distribution.

$$K = \frac{\sum_{i=1}^{N}(X_i - \langle x \rangle)^4}{\left[ \sum_{i=1}^{N}(X_i - \langle x \rangle)^2 \right]^{3/2}} \quad (2)$$
3.3.2. **Frequency domain features.** Due to the randomness of the acupoint signal, the time-domain analysis has a certain instability, so we further obtain the signal’s frequency-domain characteristics, which further reflect the nature of the body surface electrical signal.

In the frequency domain analysis, the measured data need to convert from the time domain to the frequency domain by FFT(Fast Fourier Transform), and then extract the corresponding features. The features used are: Shannon Entropy, Mean power, Total power, signal-to-noise ratio(SNR), etc. The Shannon entropy can represent the degree of uncertainty of the sequence. The calculation method of total power shown in equation (3) can represent the sum of the power spectral density, where \( p_i \) is the power spectral density.

\[
TTP = \sum_{i=1}^{N} p_i
\]

The calculation method of signal-to-noise ratio shown in equation (4) can be used to characterize the proportion of valid information in the signals, where \( s \) is the signal's total power and \( n \) is the noise’s total power.

\[
SNR = s/n
\]

3.3.3. **Nonlinear domain features.** These features are extracted based on the nonlinear analytical method. The features in this paper include coefficient of variation(COV), median absolute deviation(MAD), autocorrelation coefficient(AC), the number of turning points (NT) [7], etc. The calculation method of COV shown in equation (5) can characterize the degree of dispersion of the signal, where \( \text{std}(X) \) is the variance of the sequence.

\[
COV = \text{std}(X)/\langle x \rangle
\]

NT counts the number of changes in the sign of the slope, in other words, the number of signal peaks, and the calculation method is shown in equation (9), where \( u(x) \) indicates a unit-step function.

\[
NT = \sum_{i=1}^{N-2} u[(x_{i+1} - x_i)(x_{i+1} - x_{i+2})]
\]

3.4. **Wavelet features**

Wavelet analysis transforms a one-dimensional signal in the time domain into a two-dimensional time-scale space. The time-domain signal is projected and decomposed at multiple scales in the wavelet transform domain to obtain signals of different sub-bands. Different sub-band signals contain different local information and describe the dimensions of the signal more finely [8]. The basic principles of discrete wavelet transform are as follows:

Assuming that there is a continuous signal \( f(t) \), we perform discrete wavelet transform on it. The definitions of discrete wavelet transform and inverse transform are shown in equation (7) and equation (8):

\[
C_{j,k}(f,\psi_{j,k}) = 2^j \sum_{n=-\infty}^{\infty} f(t) \psi(2^j t-k) \quad j,k \in \mathbb{Z}
\]

\[
f(t) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} C_{j,k}(f,\psi_{j,k})(t) = \sum_{j=-\infty}^{\infty} f_j(t) \quad j,k \in \mathbb{Z}
\]

where \( \psi_{j,k}(t)=2^{j/2} \psi(2^{j/2} t-k) \) is a wavelet sequence, and \( \psi(t) \) is a wavelet basis function that satisfies certain conditions. \( f_j(t) \) represents the component of the signal \( f(t) \) at a certain scale \( 2^j \). \( j \) and \( k \) respectively represent frequency resolution and time offset.

By using the mallat algorithm, we decompose the signal \( f(t) \) in a finite layer, then \( f(t) \) can be expressed as equation (9):

\[
f(t) = f_0(t) + \sum_{j=1}^{L} f_j(t) = A_L + \sum_{j=1}^{L} D_j = A_1 + D_1 = A_2 + D_2 + D_1 = \cdots
\]

In equation (9), \( L \) is the number of decomposition levels, \( D_j \) is the detail component at different scales, and \( A_L \) is the low-pass approximation component, more specific equation, \( A_L = f_0^L(n) \) and \( D_j = f_j^L(n) \). The entire frequency band of the signal is divided into multiple sub-bands. The approximation coefficients and wavelet layer coefficients corresponding to these sub-bands are \( cA_L, cD_L, cD_{L-1}, \ldots, cD_1 \). Let us define \( f(n) \) as the sampled signal of continuous signal \( f(t) \), then \( f(n) \) can be regarded as an approximate value when the scale \( j=1 \), that is, \( A_0(n)=f(n) \). As shown in Figure
3, the discrete signal is decomposed by wavelet with scale $j=2,3,4$, and we get the approximation coefficient $cA_3$ and wavelet coefficients $cD_3, cD_2, cD_1$.

\[
\begin{align*}
& j = 1 \\
& j = 2 \\
& j = 3 \\
& j = 4
\end{align*}
\]

\[c(n)\]

\[cA1, cD1, cA2, cD2, cA3, cD3, cD1, cD1\]

**Figure 3.** Signal wavelet decomposition with three layers.

Due to the lack of prior knowledge of acupoint signals, a thorough decomposition of acupoint signals may be more helpful to extract useful local information. In this paper, we use the Daubechies (DB7) wavelet to decompose the signals in 7 levels ($j=8$), and get the wavelet decomposition coefficient vector $C$. The structure of the coefficient vector $C$ is as follows:

\[
C = \{cA_7, cD_7, cD_6, cD_5, cD_4, cD_3, cD_2, cD_1\}
\]

where $cA_j$ is the approximation coefficient, and $cD_7, cD_6, \ldots, cD_1$ are the wavelet coefficients.

For the approximation coefficient and the wavelet coefficients of each band, six statistical features which include maximum, minimum, mean, median, standard deviation, and average energy are extracted respectively. Therefore, an original signal can obtain 48-dimensional wavelet features [9]. These wavelet features describe different local information of the original signals.

### 4. Experiments and Results

In this paper, we obtained 7112 1000hz 3s-long signals from the original signals of 6 subjects, which include 3920 acupoint signals and 3192 non-meridian non-acupoint signals. These signals are used as training set data.

In order to evaluate the classification ability of the model more accurately. We randomly selected 12 acupuncture points and 12 non-meridian non-acupuncture point signals from all the measured signals of the seventh subject as a test data set. We randomly selected 11 times and got 11 different test data sets. Each test data set has 672 1000hz 3s-long original signals, which include 336 acupoint signals and 336 non-meridian non-acupoint signals.

We take traditional features as the first type of features, and the combination of traditional features and wavelet features as the second type of features.

Firstly, we extract the first type of features from training data and test data. In the training stage, we use the training feature data to train SVM based on the RBF kernel [10] and XGBoost based on the CART regression tree [11], optimize the parameters by the grid search method, and finally save the trained model. After that, we use 11 test data sets that have the first type of features to verify the classification effects of different models. F1-score is used as the evaluation standard for the classification ability of our models. Then we extract the second type of features from the training data and test data, and repeat the above process.
4.1. Model accuracy

![Figure 4. F1-score on 11 test data sets.](image)

| Features                      | Method       | Average of F1-score | Standard deviation of F1-score |
|-------------------------------|--------------|---------------------|-------------------------------|
| Traditional features         | SVM          | 0.6981              | 0.0271                        |
| Traditional features + Wavelet features | SVM          | 0.7373              | 0.0271                        |
| Traditional features         | XGBoost      | 0.7436              | 0.0234                        |
| Traditional features + Wavelet features | XGBoost      | 0.7728              | 0.0309                        |

As we have seen in Figure 4 and Table 1, from the perspective of the features, when using the second type of features that combine traditional features and wavelet features, the average F1-score of SVM in the 11 test sets is 0.7373, which has an increase of 0.0392 compared to using only traditional features, the average F1-score of XGBoost in the 11 test sets is 0.7728, which has an increase of 0.0292 compared to only using traditional features. The experimental results show that, compared with the first type of feature vectors that only extract the global features of the signals, the second type of feature vectors integrate the global and local features of the signals, have higher separability, and improve the model’s ability to classify acupoint and non-acupoint potential signals.

From the perspective of the models, using the feature vectors constructed in this paper, the average F1-score of the SVM and XGBoost classification models on the 11 test sets are all above 0.73, which can achieve the purpose of classification for acupoints and non-acupoints well. And when using the same type of feature vectors proposed in this paper, the classification performance of the XGBoost is better than SVM, so it has a better generalization ability.

4.2. Feature importance

In order to further explore the contribution of traditional features and wavelet features in the second type of features, We calculate the importance of all the features in the XGBoost model, shown in Table 2, which show Top10, Top20, Top40, Top60 important features proportion.

| Features                      | Total | Top10 | Top20 | Top40 | Top60 |
|-------------------------------|-------|-------|-------|-------|-------|
| Traditional features         | 42    | 6     | 12    | 23    | 35    |
| Wavelet features             | 48    | 4     | 8     | 17    | 25    |

We can find that the traditional features are effective and they represent the global features of the signals. Meanwhile, the wavelet features are also indispensable, they represent the local characteristics of the signals. Combining global features with local features can describe the regular differences between the feature spaces of acupoints and non-acupoint signals more comprehensively, thereby
improving the classification performance of the models.

5. Conclusion

Based on the integrity of the meridians, this paper obtains the electrical signals of the acupoints and non-acupoints on the twelve meridians of the human body. Compared with previous studies that only focused on a few acupuncture points, the experimental data are more sufficient, diverse and ensure the balance of the number of different types of samples. In addition, this paper proposes a method that combines traditional features and wavelet features to classify acupoints and non-acupoints. This method can effectively improve the ability of SVM and XGBoost to classify acupoints and non-acupoints. It provides new ideas for the selection of feature vectors for meridian potential signal classification, the precise positioning of acupoints and non-acupoints. And it also helps to further explore the characteristics of meridian acupoints. For the acupoint and non-acupoint classification tasks, the experimental results show that the classification performance of XGBoost is better than that of SVM, and it has a better generalization ability. As for the future work, we will apply this method to classify the same acupoint between healthy people and people with specific diseases and provide new ideas for intelligent disease diagnosis based on meridian.

Acknowledgments

This work was supported by the Special Fund for the Development of Shenzhen (China) Strategic New Industry (JCYJ20170818085946418) and the Shenzhen (China) Science and Technology Research and Development Fund (JCYJ20180503182125190).

References

[1] Li L X, Meng Z H, Fan X N, et al. 2012 Progress of research on specificity of meridian acupoint efficacy. J. World Journal of Acupuncture – Moxibustion 22(2) 55-59
[2] Ahn A C and Martinsen Ø G 2007 Electrical characterization of acupuncture points: technical issues and challenges. J. The journal of alternative and complementary medicine 13(8) 817-824
[3] Lee Y H, et al. 2005 Effects of acupuncture on potential along meridians of healthy subjects and patients with gastric disease. J. The American Journal of Chinese Medicine 33(06) 879-885
[4] Jabbari A, Grimnes S and Martinsen Ø G 2007 Electrodermal response–correlation between potential and conductance. 13th Int. Conf. on electrical bioimpedance and the 8th Conf. on electrical impedance tomography (Springer, Berlin, Heidelberg) pp 747-750
[5] Luo S and Johnston P 2010 A review of electrocardiogram filtering. J. Journal of electrocardiology 43(6) 486-496
[6] Liu Q, Ma L, Ai Q, Chen K and Meng W 2018 Knee joint angle prediction based on muscle synergy theory and generalized regression neural network. In 2018 IEEE/ASME Int. Conf. on Advanced Intelligent Mechatronics (AIM) pp 28-32
[7] Xi X, Tang M, Miran S M and Luo Z 2017 Evaluation of feature extraction and recognition for activity monitoring and fall detection based on wearable sEMG sensors. J. Sensors 17(6) 1229
[8] Tuncer T, Dogan S and Subasi A 2020 Surface EMG signal classification using ternary pattern and discrete wavelet transform based feature extraction for hand movement recognition. J. Biomedical Signal Processing and Control 58 101872
[9] Sravani C, Bajaj V, Tararan S and Sengur A 2020 Flexible analytic wavelet transform based features for physical action identification using sEMG signals. J. IRBM 41(1) 18-22
[10] You Y, Bai L, Xue T, Zhong C, Liu Z and Tian J 2011 Differential spatial activity patterns of acupuncture by a machine learning based analysis. In Medical Imaging 2011: Biomedical Applications in Molecular, Structural, and Functional Imaging Vol. 7965 pp 79651Q
[11] Chen T and Guestrin C 2016 Xgboost: A scalable tree boosting system. In Proc. of the 22nd acm sigkdd Int. Conf. on knowledge discovery and data mining pp 785-794