Application research of pulse signal physiology and pathology feature mining in the field of disease diagnosis

Lin Fan¹²³, Jin Cheng Zhang¹²³, Zhongmin Wang¹²³, Xiaokang Zhang¹, Ruiling Yao¹ and Yan Li¹

¹School of Computer Science and Technology, Xi’an University of Posts and Telecommunications, Xi’an, Shaanxi, China; ²Shaanxi Key Laboratory of Network Data Analysis and Intelligent Processing, Xi’an University of Posts and Telecommunications, Xi’an, Shaanxi, China; ³Xi’an Key Laboratory of Big Data and Intelligent Computing, Xi’an 710121, Shaanxi, China

ABSTRACT

This experiment is based on the principle of traditional Chinese medicine (TCM) pulse diagnosis, the human pulse signal collected by the sensor is organized into a dataset, and the algorithms are designed to apply feature extraction. After denoising, smoothing and eliminating baseline drift of the photoelectric sensors pulse data of several groups of subjects, we designed three algorithms to describe the difference between the two-dimensional images of the pulse data of normal people and patients with chronic diseases. Convert the calculated feature values into multi-dimensional arrays, enter the decision tree (DT) to balance the differences in human physiological conditions, then train in the support vector machine kernel method (SVM-KM) classifier. Experimental results show that the application of these feature mining algorithms to disease detection greatly improves the reliability of TCM diagnosis.

ARTICLE HISTORY

Received 6 August 2021
Accepted 1 November 2021

KEYWORDS

Pulse diagnosis; signal processing; decision tree; SVM; kernel method; pulse wave analysis

1. Introduction

Before the popularization of modern medicine, pulse diagnosis is a common method of TCM diagnosis of diseases and has accumulated rich clinical experience. In some developing countries such as China, the proportion of people over 60 years old is expected to increase from 12% to 40% within the first half of the 21st century (Fang et al. 2015). This will bring challenges to the construction of the medical system.

Pulse can reflect the physiological and pathological information of the human body, this information can be used to judge the physiological state of the subject. A study by the Indian Institute of Science proposed a pulse analysis model to identify the physical state of the subject by extracting the waveform characteristics of the pulse signal. This model can accurately identify a person’s state before and after meals, before and after exercise. The classification accuracy rates of sports and lunches were 99.71% and 99.94%, respectively (Rangaprakash and Dutt 2015). At the same time, certain diseases, such as cardiovascular disease (CVD), including coronary heart disease (CHD), can also be identified by biological signals (Wang et al. 2007). However, some specific chronic diseases are difficult to diagnose through physical signals, and the scope of application of this model is correspondingly limited (Shu and Sun 2007; Chen et al. 2009; Kurdyak et al. 2015; Wang et al. 2016). In other words, for chronic diseases that are not easy to find in patients, such as coronary heart disease, hypertension, heart rate imbalance, intestinal ulcers, etc. Without considering the individual’s physiological conditions, it is difficult to have an accurate judgment result.

Pulse diagnosis, as the main disease detection method of TCM diagnosis, can theoretically collect a large amount of information about the health status of patients, and the information collected can be more extensive compare with auscultation in modern medicine (Rangaprakash and Dutt 2015; Chu et al. 2018). At present, it is the main way to detect human pulse signals with the sensor-computer model (Laurent et al. 2001; Chung et al. 2013; Wang et al. 2019), and many automatic pulse diagnoses technology using signal processing and pattern recognition techniques (Chen et al. 2011; Liu et al. 2012; Yu-Feng et al. 2013; Rangaprakash 2014; Rangaprakash and Dutt 2014; Boutry et al. 2019). Some experiments have shown that when multiple parameters are used for supervised learning and classifiers, principal
component analysis (PCA) is not required, and the accuracy of any classification for the corresponding disease can reach more than 90%. It is also feasible in the SVM classification model to detect the physiological state of the human body by using the pulse time-domain feature data image (Nanyue et al. 2015; Tsai et al. 2019). Stable, ultra-low power consumption and high sensitivity sensors are of great significance for the miniaturization of wearable physiological signal monitoring systems and the assistance of TCM diagnosis. The characteristic index collection method is used to analyze the pulse signals of volunteers, and the detection of a variety of chronic cardiovascular diseases can be simple, low-cost, and efficient (Ouyang et al. 2017). Existing literature often mentions the idea of recording pulse electrical signal data for a period of time as a data set, using this algorithm for disease detection, but it should be noted that the pulse image shape and structure of sensors with different principles are also different (Jaafar et al. 2015).

In the existing research on pulse measurement for disease detection, it has been proved to be an effective experimental means to classify and determine diseases by using the SVM classification algorithm and DT technology in machine learning (Luo et al. 2016; Vardoulis et al. 2016). SVM soft margin classifiers are important learning algorithms for classification problems, especially suitable for huge amounts of samples without classification standards (Wu and Zhou 2005). Because the SVM method is suitable for large-scale multi-feature data classification, according to the data sample sampling and disease feature labeling methods of this study, the current use of SVM is still the most effective disease method of this model. This property can obtain effective classification results when processing a large amount of pulse signal data, as the amount of sample data increases, the stability and accuracy also increase, which is suitable for machine learning of large-scale high-dimensional data of pulse records (Ruping 2001). After extracting the pulse wave feature data and generating a two-dimensional atlas, normal people and patients are classified under the framework of SVM, which can achieve high disease recognition accuracy (Shafri and Ramle 2008; Jia et al. 2014). Because of the different physiological characteristics of each person, the pulse waveform will vary with age, gender, and activity. For example, the index of enhancement (augmentation index) and the index of enhancement pressure (augmentation pressure) extracted from the pulse wave of the wrist or right carotid artery is different between genders (Hayward and Kelly 1997; Gatzka et al. 2001; Weber et al. 2004; Cumberknop et al. 2011; Lee et al. 2019). Therefore, when we convert the pulse data into data sets and process them, we need to design a processing algorithm that can eliminate the superposition of physiological factors and pathological factors, to complete the accurate judgment of the patients’ diseases.

The main goal of this research is to design a method that uses sensors to record the subject’s pulse and uses a well-designed computer algorithm to monitor the patient’s disease. And hope to use this paper model to design an instrument for detecting diseases through pulse diagnosis, and a brand-new disease monitoring system.

There is pulse waveform analysis (PWA) and disease diagnosis methods based on machine learning and pattern recognition technology in the existing pulse diagnosis research. Although these methods can usually achieve a high success rate, their disease classification results are only applicable to specific instruments. And in this type of research, there is no specific standard for determining a person with a specific disease, and no factors affecting the results are considered. The factors that affect the characteristic value of the pulse waveform, such as the measurement process, the subject’s gender, age, biological clock, physiological state, and disease, are often not considered separately. For this kind of diagnostic method that requires high-precision measurement, these shortcomings obviously cannot be ignored. Due to the influence of many of the above factors, it is difficult to use the features extracted from the existing pulse signal as a basis for disease diagnosis. To discuss how individual physiological characteristics are reflected in the specific value of the pulse condition of the subject, and what influence they have on the internal characteristics of the pulse condition data, we often need to explore these factors when completing the PWA disease diagnosis. If these factors that interfere with the diagnosis result are not considered, the results obtained by directly using the SVM to segment the pulse data of normal people and chronic diseases are unreliable. But we tried to create a method to identify the characteristics of pathological data in human pulse images by excluding all factors that affect the results to determine the subject’s disease.

2. Pulse data acquisition methods reference

The experiment is mainly to improve the existing methods of using computers for pulse diagnosis, and to try to create a more reliable, effective, and comprehensive diagnosis model for evaluating the patient’s state. While providing a brand-new disease diagnosis method, the
pulse image differences between different people found in the experiment are also listed one by one.

At present, there are three main kinds of sensors existing for measuring pulse, which are electric pulse sensor, piezoelectric sensor, and photoelectric sensor. The electric pulse sensor is generally used in ECG measurement, and its measurement principle is roughly the same. A photoelectric sensor is a common device in the market, which can detect the pulse beat sensitively in different measuring positions, so it is widely used in wearable sporting heart rate detection equipment. However, although the above two devices can accurately detect the human pulse signal and cycle, they are unable to obtain more detailed changes of wrists’ pulse pressure. The piezoelectric sensor can collect the pressure change of the artery, which is consistent with the principle of pulse diagnosis but requires exact measurement position and higher accuracy. This experiment tried to find the pathological characteristics reflected by the two-dimensional pulse condition through analysis, discover the internal connection between the waveform characteristics of the human pulse condition and the disease and try to describe it with an algorithm, trying to establish a method for detecting specific diseases using photoelectric signals.

Another aspect is that the shape of the two-dimensional data collected by the piezoelectric sensor is also different from that of the electric pulse and photoelectric sensor (as Figure 1 describes). The starting point of the pulse cycle data collected by the piezoelectric sensor is the lowest value, which rises and rebounds with the fluctuation of the heart, and finally returns to the lowest value. However, other types of sensors are often data of current oscillation, which is unable to reflect the real pressure changes of arterial blood flow, and it is difficult to distinguish the physiological and pathological differences of subjects. The former is usually used in practical experiments (Figure 1).

The data of this experiment is based on the photoelectric sensor. The measurement of the human pulse by the photoelectric sensor is based on TCM theory combined with machine learning technology. There are three acupoints Cun, Guan, and Chi in the human wrist pulse from top to bottom, which correspond to several organs of the human body (as Figure 2 shows, human body information corresponding to different acupoints in TCM). The abnormal pressure changes between acupoints at the wrist are the response to the changes of physiological and pathological conditions of the human body. It should be noted that the deviation of measurement position by the piezoelectric sensor and the wrist movement of the subject need to be avoided, and the bending angle of the wrist also needs to set a standard. When we get the data that accurately reflect all kinds of signals, we can identify the signal source through machine learning algorithms such as SVM and DT, and evaluate the health status of the subjects, and finally, make the diagnosis of the diseases (Figure 2).

Finally, it should be noted that the sampling device designed based on the current technology is composed of electrical signals and external embedded devices, and the data transferred from the sensor to the computer may be affected by some hardware factors. Therefore, these external embedded devices for signal processing need to set the appropriate sampling frequency and delay according to the characteristics of the sensor. The clarity of the pulse waveform will increase as the sampling frequency increases, but once the threshold is exceeded, the signal will be distorted due to excessive noise. The sampling delay of the device will affect the sampling frequency. Setting the cycle too long will cause redundant data to be generated, otherwise, the recording of the signal content will be lost (Wang et al. 2016).

The data set used in this experiment is from the pulse data collected by the 211 Hospital of the People’s Liberation Army. It is from 60 groups of photoelectric sensor signals of different ages. Half of all subjects are healthy people, and half are patients with pancreatitis, appendicitis, acute appendicitis, and duodenal ulcer.
3. Data processing technological process

After effective measurement, the pulse waveforms image will show obvious periodicity, which is equal to the subject’s heart rate. Whether the pulse data is periodic or not and whether it is completely synchronized with the heart rate of the subjects are the basic criteria to determine whether the pulse data can be used in the experiment.

The other aspect is the experimental interference, in the process of pulse data acquisition, except for the abnormal factors caused by the movement of the measurers and the vibration of the surrounding environment, the interference waves, and noises of the instrument circuit will also lead to the abnormal enhancement the intensity of pulse images. Usually, simply reducing these abnormal peaks to the average height of the normal pulse cycle can simulate the original signal index of this part of the pulse waves image without interferences, these anomalies are caused by the hardware defects of sensors and embedded devices, which need to be eliminated manually. Excessive signal noise is usually caused by the abnormal processing of the circuit current value by the amplifier module. When processing data, we should use filtering algorithms to remove errors.

After the pulse waveform image is denoised, it is not easy to observe a clear pulse waveform image. The reason is that when the sampling frequency is set high enough, the pulse image will have up and down amplitude due to the electric pulse oscillation caused by noise interference during current conduction (as Figure 3a image described, the original waveform will appear longitudinal oscillation phenomenon). Therefore, the typical pulse characteristics after noise reduction can observe a more accurate and clear human pulse waveform.

However, due to the electromagnetic wave oscillation phenomenon of the high-precision wearable test sensor, the pulse characteristics cannot be accurately extracted after removing the noise in the original pulse signal data. Therefore, we need to obtain a smooth waveform to extract appropriate data features and pulse waveform for analysis. When the smoothing algorithm is applied to the data processing, the amplitude and shape of each position of the pulse waveform can be observed more clearly, which can be used as the data analysis and eigenvalue extraction. Figure 3b described an example of the effect of the pulse smoothing algorithm.

Another case that needs to be dealt with is the longitudinal offset of the baseline of the pulse wave image due to the measurement movement or the current conduction change of the instrument caused by external factors, the criterion of this phenomenon is that there is a deviation in the height of the fixed position of each pulse cycle. The principle of the baseline elimination algorithm is to calculate the average time length of the pulse cycle of a whole set of pulse data and separate each pulse cycle. Taking the position of the initiation point of the whole pulse data as the reference point, the positive and negative difference between the first data point and the initiation point of each subsequent pulse cycle is calculated, and the whole pulse cycle is moved longitudinally according to the difference (As Figure 4 described).

However, before designing the physiological characteristic algorithm, the physiological differences between people reflected in the pulse image also need to be taken into account in the pulse design algorithm, because there is very significant that the pulse waves images difference between different people that due to age, gender, physical fitness, and other conditions, as shown in Figure 5. For example, the elasticity of human blood vessels gradually decreases with age, so the
boundary between the systolic and diastolic pulses in the pulse image will also become inconspicuous. This situation is reflected in the pulse images of elderly subjects (Yokota et al. 2016; Lee et al. 2019). In addition, according to the information reflected in our data, the heartbeat difference between the sexes will also reflect different pulse waveform states. Women’s hearts are more inclined to ‘shrink’, men are more inclined to ‘beat.’ Specifically, the ratio of systolic and diastolic blood pressure to the entire pulse cycle in women’s pulse images is higher than that in men (Gatzka et al. 2001; Weber et al. 2004; Nanyue et al. 2015), which can also be reflected in the data marked in red in Table 2.

In addition, another commonly used pulse signal processing method is the filtering algorithm, which can more accurately reflect the pulse waveform. To observe the distribution of the subject’s pulse waveform, the entire pulse cycle of the subject can be divided according to the principle of sampling points in the baseline drift elimination algorithm. The Butterworth bandpass filtering algorithm is widely used to remove the noise of the collected data (Bianchi and Sorrentino 2007). Therefore, after the previous pulse has passed the noise reduction algorithm, the filtering algorithm can be used again to make the pulse waveform smoother.

Through the above three kinds of data processing, we can get the pulse waveform which can observe the physiological characteristics of the subject’s pulse. These data are transformed into two-dimensional images by several one-dimensional values, and the image changes very smoothly and accurately after processing. Human pulses images usually tend to show periodicity. The physiological and pathological characteristics of the subject can be obtained by calculating the average of several cycles of these data and drawing the images. The final pulse image (after three processing methods) can present a specific waveform structure.

**Figure 6** is the standard pulse waveform of the human body after processing. It is composed of one heartbeat of the human body and is divided into two stages: systole and diastole. As shown in Figure 6, the pressure change monitored by the sensor also conforms to this curve.

The standard pulse waveform image measured by a piezoelectric sensor usually presents a standard double-peak structure (some chronic pathological diseases, such as pancreatitis, will appear three-peak structure), and several data eigenvalues can be extracted according to these data, which can distinguish the individual’s physiological conditions. These eigenvalues will be affected by the subject’s age, gender, heart rate along other physiological and pathological conditions. These factors are like physiological fingerprints of subjects, which can not only distinguish different subjects according to the differences between these characteristic values of pulse images of different subjects but also analyze the physiological conditions of subjects according to these eigenvalues.

4. Classification algorithm design and practice of disease determination

When the subjects could not know that they had a certain chronic disease, then judging the chronic diseases of the subjects couldn’t rely on simple disease classification accuracy to experiment. Because the standard pulse images of subjects and normal people can be distinguished with very high accuracy, the decisive factor is not only determined by the pathological characteristics of some specific chronic diseases, but also some physiological characteristics. For example, with the increase of age and the decrease of vascular elasticity, the dicrotic notch point of the pulse image will gradually disappear, and the trough position will be difficult to determine. Besides, when the two genders are different and the other health conditions are the same, the pulse intensity of men will be higher than that of women, and the height of the pulse sampling point of men will also be higher than women. Therefore, we hope to try to collect the physiological information of the subjects and let the subjects exclude the physiological characteristics through DT classification through DT algorithm. Finally, the pathological characteristics can be distinguished between the subjects and normal people by the SVM classification algorithm.
Under the same physiological conditions (such as the same age, gender, height, weight, and health), the shape of normal people’s photoelectric pulse data image is roughly the same, which will show the same waveform structure. As shown in the figure above, the pulse image (black curve) of normal people will show an obvious double-peak structure, and the height of the waveform is the highest in all States. The pulse image of patients with pancreatitis (red curve) will show the structure of three peaks, which is significantly different from the second peak of normal people. The goal of distinguishing normal people from patients...
with pancreatitis is to calculate the shadow area (definite integral of the considered region) according to the pulse waveform of the subjects and set a threshold value to determine whether the subjects have pancreatitis or not. This model is one of the important experimental ideas of this project. The formula in Figure 7 (corresponding Formula (1) and Formula (2)) records the change rate of the diastole phase of each pulse wave image of the same tester. \( D_t = \sum_{n=0}^{N} \left( y_{t+n} - A \right) \)

\[
A = \frac{1}{N} \sum_{i=0}^{N} y_{t+i} \quad D_t = \sum_{n=0}^{N} (y_{t+n} - A)
\]

\[
n \in S_F, \quad Y \in \{ (y_1), (y_{t+n}), (y_{t+2n}), \ldots, (y_{t+N}) \}
\]

\[
t = \max(Y), \quad T = \max(C_t)
\]

Calculated and recorded the changes in the diastolic wave of each pulse wave line chart for each subject, and as a dataset divided by SVM classification. Select the perceptron strategy to divide the two kinds of subjects into SVM classification.

Generally, these methods are suitable for extracting the feature values that can be used in machine learning classification methods, but whether it can also be directly used to determine the subject has a specific disease, there is no specific standard for the threshold.

Figure 7. (a) Diagram about distinguishing healthy patients from pancreatitis patients, there will be some pathological differences in the pulse images of patients with pancreatitis, obvious deviation in the definite integral values of the two in the areas with considered region (b) Algorithm design of stability index and schematic diagram.
However, it is more stable and robust to use the calculated eigenvalues as SVM classification multidimensional datasets, because if a manually designed algorithm is used as the judgment basis of suffering from a certain disease, its accuracy will greatly depend on the quality of the designed algorithm, and there is no fixed standard. If it is one of the eigenvalues of machine learning classification algorithms, we can design a variety of algorithms and use a variety of methods to verify, and improve the accuracy of feature classification as much as possible. If it is one of the eigenvalues of machine learning classification algorithms, we can design a variety of algorithms and use a variety of methods to verify, and improve the accuracy of feature classification as much as possible.

However, the two feature extraction algorithms in Figure 7 are designed based on the pathological features of five diseases (pancreatitis, appendicitis, acute appendicitis, duodenal ulcer, and hypertension) in the pulse image. The purpose is to extract these pathological features that are used as the basis for judging whether the subject has the disease. However, in order to use an algorithm to maximize the difference in pulse images between individuals, a new algorithm needs to be designed. Because even the two differences mentioned in Figure 7 still cannot reflect all the data samples, not all the pulse images of the case samples strictly meet these two characteristics, so the most specific pulse images are extracted from all individual pulse images. Strong pathological characteristics are still the challenge of this classification model. Therefore, if you want to use a set of data to represent the most detailed difference between a chronically ill patient and a normal person while maximizing the difference between the subject and the standard image, the Multiple Empirical Gini Coefficient (MEGC) could achieve this goal, as shown in Figure 8.

$$H(Y) = -\sum_{n=1}^{N} p_n \log p_n, \quad Y \in \{y_1, y_2, y_3, \ldots, y_n\}$$

(3)

Among them, a in the logarithmic function ‘$\log a \ p_n$’ in formula (3) is a parameter determined by the amplitude, and for the pulse function image of each coordinate axis established,

$$Gini(y) = \sum_{n=1}^{N} p_n(1 - p_n) = 1 - \sum_{n=1}^{N} p_n^2$$

(4)

$$Gini(D) = 1 - \sum_{n=1}^{N} \left( \frac{|Dn|}{|D|} \right)^2$$

(5)

In order to explain that the above formula can measure the uncertainty of the data, and the function image properties in the inflection point of the logarithmic image of the information entropy function (the random variable takes y and the probability of each class is equal, that is, y is completely random), so you can proof

$$P_1 = P_2 = P_3 = \ldots = P_n = \frac{1}{N}$$

(6)

Gini gets the maximum value of 1 - $\frac{1}{n}$; When $n^*$ exists so that $P_{n^*} = 1$, Gini($y$) = 0. When $n = 2$, you can export:

$$Gini(y) = 2p(1 - p)$$

(7)

When any value in Y obtains the ordinate value of the inflection point, there is:

$$Gini(y) = X_n - P^2(y \in y_{n-1}) - P^2(y \in X_n)$$

(8)
For the function, take any point in y as the origin (take the abscissa and ordinate of \( y_1, y_2, y_3, \ldots, y_n \) as 0), establish a set of coordinates, and integrate the x-axis with the intersection of the two functions as

\[
\pm \int_0^{X_n} \text{Gini}(y) \, dx = \pm \int_0^{X_n} X_n - P^2(y \in y_{n-1}) - P^2(y \in X_n) \, dx \quad (9)
\]

As shown in Figure 8, each area marked with black dots can be represented by the above formula; The absolute value of the difference between the standard pulse image of normal people and the integral of patients with chronic diseases can be calculated by bringing the pulse logarithmic function of normal people and patients with chronic diseases into Formula (9), as shown in Figure 8. The absolute value of the integral difference between the standard function image of normal people and the function image of patients with chronic diseases. This group of MEGC data can reflect the difference between any patients with chronic diseases and the standard pulse image to the greatest extent.

For the specific human pulses data to establish a standard classification model, we chose 60 subjects, half healthy people and half people with chronic diseases. We measured the pulse data of each subject many times, and grouped and recorded the data of each subject as a data set, and converted it into a single pulse image to extract the characteristic values (as Figure 9 shows). And these data are drawn into a chart, through comparison, the specific eigenvalues are substituted into SVM as a multidimensional array for classification, which can achieve very high classification accuracy (more than 95% generally).

However, only three algorithms are used to extract the values of specific pathological features, so that disease diagnosis is not accurate enough. To analyze the pulse waveform of the subjects more accurately, we chose to list the mean and range of all the characteristic values, and find out those specific characteristics that can distinguish the individual. We used the pulse feature classification model in Figure 7 and grouped the data of all subjects according to different standards and then summarized them into a table. We judged by observing some data items with obvious differences between different groups. Whether the characteristic value has specificity. The two tables we have established are to analyze more intuitively, which feature values cause the physiological and pathological differences between people.

Table 1 can help us find out the specific pathological characteristic value changes of some specific chronic diseases. Taking the data of normal people’s pulse image as the standard, by comparing the differences between the characteristic values of the subjects’ pulse and the standard normal people’s pulse images, and the SVM algorithm is used for classification distinguish.

However, although this method can achieve high classification accuracy, it is not sufficient to use this method as the basis of disease diagnosis. It should be considered that some eigenvalues are not only affected by specific diseases but also have specificity due to some physiological characteristics, such as gender and ages a shown in To diagnose the diseases, we need to exclude the physiological differences between individuals, so the classification results could have high reliability. The values of the feature data marked in red in the above table have obvious differences. When we use the disease features of pulse image to determine, we can multiply these by different weights to eliminate the physiological differences of individuals, to improve the accuracy of disease ... ... ... classification (as Figure 11 shows, the weighted eigenvalues can balance the factors of individual differences).

In order to solve this problem, we put forward a method to classify and distinguish physiological differences by DT and eliminate them. The analysis of the eigenvalues shows that some physiological and pathological characteristics can affect the eigenvalues of the pulse, and the disease diagnosis needs a decision-making algorithm to ignore the physiological differences as much as possible. We can set up a standard model for reference, such as the pulse values of 30-year-old healthy medium build men, the eigenvalue differences of subjects due to physiological differences were multiplied by a fixed weight (in this
Table 1. Averaged eigenvalues grouped by specific chronic diseases.

| Spatial feature              | Normal                  | Pancreatitis | Appendicitis | Acute Appendicitis | Duodenal Ulcer |
|-----------------------------|-------------------------|--------------|--------------|--------------------|----------------|
| Percussion peak (h1)        | 111.1 (73.3 – 126.0)    | 106.2 (72.3 – 131.0) | 112.1 (72.8 – 135.0) | 112.2 (72.5 – 136.9) | 118.6 (76.2 – 136.5) |
| Dicrotic notch height       | 9.9 (–25.4 – 38.6)      | 10.6 (1.5 – 17.2)    | 13.1 (–8.1 – 21.3)   | 14.0 (–11.8 – 18.4) | 9.9 (–10.6 – 19.1)  |
| Dicrotic peak (h2)          | 36.2 (25.0 – 52.1)      | 25.5 (20.2 – 42.1)   | 28.4 (7.3 – 57.5)    | 29.8 (10.9 – 37.7)  | 42.3 (26.0 – 56.7)  |
| Pulse period (T/ms)         | 818 (659 – 1029)        | 909 (890 – 998)      | 814 (658 – 940)      | 821 (640 – 989)     | 901 (715 – 990)     |
| Systolic phase time (T1/ms) | 287 (230 – 350)         | 299 (240 – 420)      | 242 (215 – 318)      | 276 (219 – 425)     | 347 (218 – 475)     |
| Diastolic phase time (T2/ms)| 531 (287 – 689)         | 610 (493 – 668)      | 571 (360 – 740)      | 545 (348 – 724)     | 554 (392 – 725)     |
| Systolic index (T1/T)       | 0.35 (0.29 – 0.44)      | 0.33 (0.27 – 0.47)   | 0.30 (0.23 – 0.46)   | 0.34 (0.27 – 0.60)  | 0.39 (0.22 – 0.66)  |
| Diastolic index (T2/T)      | 0.64 (0.53 – 0.70)      | 0.66 (0.54 – 0.69)   | 0.70 (0.38 – 0.79)   | 0.64 (0.29 – 0.40)  | 0.61 (0.55 – 0.73)  |
| Vasoconstriction range      | 108.5 (69.0 – 148.9)    | 94.3 (70.0 – 120.2)  | 97.5 (68.2 – 134.5)  | 104.4 (89.6 – 118.0)| 104.5 (84.0 – 136.8)|
| Dicrotic notch index (h2/VR)| 0.33 (0.29 – 0.45)      | 0.27 (0.18 – 0.39)   | 0.29 (0.25 – 0.52)   | 0.29 (0.27 – 0.35)  | 0.40 (0.29 – 0.73)  |
| Vasoconstriction index (VR/h1) | 0.98 (0.70 – 1.18) | 0.89 (0.81 – 1.00)   | 0.87 (0.65 – 0.98)   | 0.93 (0.73 – 1.10)  | 0.88 (0.63 – 1.00)  |
| Stability index             | 132.53 (58.53 – 443.65) | 205.03 (120.08 – 675.69) | 198.31 (60.23 – 582.56) | 136.78 (59.36 – 401.23) | 225.78 (61.15 – 455.23) |
| Three-peak structure index  | (–187.40 – 25.38)       | (–38.40 – 104.72)    | (–126.45 – 108.12)   | (–89.11 – 94.23)    | (–8.45)          |

The article, Table 2 only lists gender and age. If more physiological characteristics are listed in the table to classify, more differences can be found. Moreover, Figure 10 is a view of the pulse eigenvalue box summarized according to Tables 1 and 2. The differences of different eigenvalues in different subjects will also show different degrees.

The analysis of the eigenvalues shows that some physiological and pathological characteristics can affect the eigenvalues of the pulse, and the disease diagnosis needs a decision-making algorithm to ignore the physiological differences as much as possible. We can set up a standard model for reference, such as the pulse values of 30-year-old healthy medium build men, the eigenvalue differences of subjects due to physiological differences were multiplied by a fixed weight (Refer to Table 2, only gender and age are listed above, we can put more physiological characteristics on the table) (Figure 11).

The basic idea is to set a standard pulse image (appropriate state, age, BMI index, exercise, and living habits) as the reference pulse image, and extract each physiological factor that affects the normal person’s pulse image features that fit with the standard image. Through several physiological characteristics to determine the parameters of pulse image factors, from large to small as the DT from top to bottom nodes, and step by step down to determine, until the characteristics left only pathological factors. First, the characteristics of which are physiological factors and which are pathological factors are determined by statistics. After excluding the physiological factors, the pathological features were classified with SVM. If the physiological factors and pathological factors overlap, the algorithm is used to offset and divide them according to pathological factors. We can set a weight on the image to offset the impact of physiological factors on the pulse image. For instance, under the same conditions, the pulse intensity of men will be 12% higher than that of women. We can multiply the pulse parameters of women by 1.2 to reach the balance. Before the subject’s pulse waveform data enters the nuclear method classification model, the data is first sent to the DT, and following the reference standard, each feature value is multiplied by the corresponding weight, so that each feature is closer to that of a normal person, which means before and after comparison with 28 control groups, the method can improve the accuracy by 5.74% on average. However, to determine these weights, we need to find out the physiological characteristics that determine the characteristic value of pulse image through big data (characteristic values marked red in Table 2) and determine the accurate weight value through various comparisons. But due to the three-peak structure index being too dependent on individuals, so it is only a reference for disease.

After multiplying all twelve indexes by corresponding weights, the SVM-KM is used to classify and calculate their accuracy. If we select the characteristic value with strong specificity for some chronic diseases according to the previous table and use it as the parameter of SVM classification, after the weight calculation of the DT, if the classification accuracy is much higher than the average value of healthy normal people, it can be determined that the subject has this specific disease. If the accuracy is lower than or similar to that of healthy people, it is judged that the subject does not have the disease. We use the existing data, according to this method for some classification attempts. DT combined with weight algorithm is mainly used to solve the problem of people who are too old or too young and have special physical
conditions and overcome the SVM judgment of pathological factors caused by physiological differences (Table 3).

Judging from the classification results of our existing multi-group samples, if the SVM-KM classification error rate is less than 18%, it is the standard to determine the disease. As shown in Table 4, from a single test result, the success rate of this classifier in identifying patients with four chronic diseases is higher than 85%, and the misdiagnosis rate of other diseases except for acute appendicitis (AA) is very low. The results show that the classification method is feasible in the existing photoelectric sensor pulse data samples for disease prediction results. As long as other diseases are monitored multiple times, the diagnosis success rate can almost reach 100%.

5. Discussion

Although the classification result is based on the traditional SVM-KM method, compared with others such as the Physical meaning design algorithm and BP Neural Network, this mode combines feature mining and classification algorithms (Rangaprakash and Dutt 2015; Chen et al. 2020; Yuesheng 2020), and its results are more robust. And this mode combined with more highly sensitive and information-gathering optical sensors (such as ultrasensitive pulse sensors

![Figure 10. Numerical ranges of different types of eigenvalues are represented by block maps. Taking the data of healthy people as the reference model, we can conclude by comparing the data of a certain subject with that of normal people. Box data refers to the highest value, the highest sample mean, the average value, the lowest sample mean value, and the extremely low value from the top to the bottom (data of systolic phase and diastolic phase had been compressed by half).](image_url)

**Table 2.** Averaged eigenvalues grouped by gender and age.

| Spatial feature          | Men                  | Women                | Youths               | Elders               |
|-------------------------|----------------------|----------------------|----------------------|----------------------|
| Percussion peak (h1)    | 110.5 (72.5 – 136.9) | 97.9 (72.3 – 130.8)  | 120.1 (72.8 – 136.9) | 103.9 (72.3 – 125.0) |
| Dicrotic notch height   | 12.9 (7.3 – 38.6)    | 9.4 (7.3 – 34.5)     | 12.3 (7.3 – 38.6)    | 10.0 (7.3 – 27.7)    |
| Dicrotic peak (h2)      | 26.0 (7.3 – 56.7)    | 29.2 (10.9 – 57.5)   | 24.1 (7.3 – 57.5)    | 36.4 (26.0 – 52.1)   |
| Pulse period (T/ms)     | 352 (638 – 1029)     | 735 (640 – 948)      | 824 (640 – 1029)     | 861 (742 – 990)      |
| Systolic phase time (T1/ms) | 299 (218 – 420) | 379 (215 – 475)     | 291 (215 – 475)     | 338 (267 – 420)     |
| Diastolic phase time (T2/ms) | 531 (287 – 740) | 356 (308 – 443)     | 532 (287 – 740)     | 522 (392 – 525)     |
| Systolic index (T1/T)   | 0.36 (0.22 – 0.66)   | 0.52 (0.23 – 0.60)   | 0.35 (0.22 – 0.66)   | 0.39 (0.26 – 0.57)   |
| Diastolic index (T2/T)  | 0.64 (0.23 – 0.60)   | 0.48 (0.23 – 0.60)   | 0.65 (0.29 – 0.79)   | 0.61 (0.55 – 0.70)   |
| Vasoconstriction range  | 1083 (48.0 – 148.9)  | 95.5 (70.0 – 118.0)  | 116.5 (48.0 – 148.9) | 100.3 (70.0 – 120.2) |
| Dicrotic notch index (h2/VR) | 0.25 (0.18 – 0.73) | 0.31 (0.21 – 0.52)   | 0.21 (0.18 – 0.73)   | 0.36 (0.25 – 0.50)   |
| Vasoconstriction index (VR/h1) | 0.98 (0.63 – 1.18) | 0.97 (0.81 – 1.10) | 0.97 (0.63 – 1.18) | 0.96 (0.81 – 1.01) |
| Stability index         | 149.45 (58.53 – 675.69) | 159.52 (59.36 – 582.56) | 152.1 (58.53 – 675.69) | 152.98 (78.76 – 625.4) |
| Three-peak structure index | -95.30 (-217.69 – 108.12) | -13.65 (-126.45 – 100.01) | -72.19 (-197.23 – 108.12) | -36.27 (-217.69 – 107.55) |
Of course, the three algorithms designed in this paper are all used to calculate the sampling data of the pulse photoelectric sensor. If the two-dimensional image of the pulse data is applied to the computer vision algorithm, the clustering results can be obtained through the fusion of similarity maps and features (Li et al. 2018b, 2019). The adaptive probabilistic neighborhood learning process is used to restore the block diagonal affinity matrix of the ideal graph. At the same time, through a flexible embedding scheme, the inherent clustering structure is revealed in the low-dimensional subspace, which effectively suppresses the irrelevant information and noise in the high-dimensional data (Li et al. 2018a). This may also replace the limitations of the SVM classification method, and can also process the pulse image, but the corresponding algorithm needs to be modified.

6. Conclusion

After selecting the standard reference pulse model, we collected the characteristics of each subject’s pulse data. We label the characteristic values of each subject’s pulse data, and adjust them by a DT, and select the specific characteristic values for classification according to the corresponding chronic diseases. If the subject’s pulse characteristic value has pathological abnormalities, then we can use the same standard pulse model for the SVM classification algorithm application, and take the accuracy of the algorithm model as the judgment standard of suffering from specific diseases.

The results show that the success rate of SVM kernel classification between healthy subjects and the pulse standard model is 52% to 92%. If the subjects suspected to have a certain chronic disease are judged according to the eigenvalue data, they are classified with the standard model, the accuracy is much higher than the averages of the former, it can be determined that it has the disease. According to the data in Table 1, when we use the average value of normal people as the reference sample group, then find out those characteristic values that are specific, and selected characters of dicrotic notch height, stability index, and

![Figure 11. DT used to exclude individual physiological differences, from root to leaf node to judge individual physiological characteristics, health status, and physiological state.](image-url)

### Table 3. The classification success rate of the four diseases based on the linear classification algorithm, the SVM-KM, and the KM weighted by the DT based on the two indexes.

| Spatial feature | Linear Classify | SVM-KM | KM after DT |
|-----------------|----------------|--------|-------------|
| Pancreatitis    | 80.17%         | 87.07% | 92.24%      |
| Appendicitis    | 72.41%         | 83.62% | 84.48%      |
| Acute Appendicitis | 67.59%     | 77.93% | 79.31%      |
| Duodenal Ulcer  | 63.01%         | 84.95% | 84.01%      |

### Table 4. Confusion matrix of the recognition success rate of multiple sets of pulse data for four diseases.

| Disease Condition | P Healthy | A Healthy | AA Chronic Disease | DBU Chronic Disease |
|-------------------|-----------|-----------|-------------------|---------------------|
|                   | Yes | No | Yes | No | Yes | No | Yes | No |
| Healthy            | 0   | 90 | 1   | 89 | 6   | 84 | 2   | 88 |
| Chronic Disease    | 16  | 2  | 33  | 3  | 19  | 5  | 17  | 1  |

(Xu et al. 2021)), its accuracy is expected to be higher. If this mode is applied to wearable devices equipped with this instrument, combined with related machine learning algorithms, it can even solve the problem of using smart devices to monitor diseases and use smart devices to monitor the wearer suffering from pancreatitis, appendicitis, hypertension, and duodenum. The condition of the disease, even the physiological state of the subject, such as pregnancy, before and after meals, and before and after exercise, etc. (Rangaprakash and Dutt 2015; Chen et al. 2020). If these measuring and recording instruments are fully improved and applied, combined with similar feature extraction algorithms, we will obtain a faster, more accurate, and low-cost disease monitoring method too.
three-peak structure index from appendicitis patients as the multidimensional samples for kernel method classification, for example, it can reach an average accuracy of over 98%. For the patients with acute appendicitis and duodenal ulcer, it can also reach an average accuracy of 95.4 ± 4.1% and 97.7 ± 1.9%. The selection of eigenvalues and the feature optimization algorithm designed according to different target diseases determine the classification accuracy of the SVM-KM.

DT algorithm is used to introduce the accuracy of the intelligence factor of the algorithm, so it can also improve the classification accuracy to a certain extent and increase the credibility of the result. To verify the accuracy of the DT algorithm, we randomly select several pulse cycles in the four groups of diseases. As shown in the decision Table 3, after the DT method algorithm, the classification accuracy reached 92.24%, 84.48%, 84.01%, and 79.31% respectively.

At present, based on the existing data sets, this method has a certain degree of efficiency and accuracy, which can be used as a theoretical reference for auxiliary diagnosis with the help of biological signals. It is also necessary to overcome the problem that for some special physiological characteristics of the subjects (professional athletes, obese people, the elderly, people with arrhythmia and congenital defects), whether they can change their eigenvalue data through DT weight so that they can carry out disease diagnosis. Or design a more effective DT algorithm to process the data. Moreover, the samples we used are limited, and it is worth further exploring whether the specific eigenvalues are universal.

Acknowledgment

Thanks for the help of the open data set provided by 211 Hospital of PLA. The work of this thesis is supported by the special fund construction project for key disciplines of ordinary colleges and universities in Shaanxi Province.

Disclosure statement

There are no conflicts of interest for all the authors.

Funding

Project supported by National Natural Science Foundation of China (61373116), General Project in the industrial field of Shaanxi Province (2018gy-013), Project of Xianyang Science and Technology Bureau (2017k01-25-2); The work of this paper is supported by the special fund construction project for key disciplines of colleges and universities in Shaanxi Province.

ORCID

Jin Cheng Zhang http://orcid.org/0000-0002-2040-0496

Data availability statement

Any data is provided as a supplementary file for the paper, it is used to verify the reliability of the experimental results.

References

Bianchi G, Sorrentino R. 2007. Electronic filter simulation & design. New York: McGraw-Hill Professional; p. 17–20.
Boutry CM, Beker L, Kaizawa Y, Vassos C, Tran H, Hinckley AC, Pfattner R, Niu S, Li J, Claverie J, et al. 2019. Biodegradable and flexible arterial-pulse sensor for the wireless monitoring of blood flow. Nat Biomed Eng. 3(1):47–57.
Chen J, Huang H, Hao W, Xu J. 2020. A machine learning method correlating pulse pressure wave data with pregnancy. Int J Numer Method Biomed Eng. 36(1):e3272.
Chen Y, Zhang L, Zhang D, Zhang D. 2009. Wrist pulse signal diagnosis using modified Gaussian models and Fuzzy C-Means classification. Med Eng Phys. 31(10):1283–1289.
Chen Y, Zhang L, Zhang D, Zhang D. 2011. Computerized wrist pulse signal diagnosis using modified auto-regressive models. J Med Syst. 35(3):321–328.
Chu Y, Zhong J, Liu H, Ma Y, Liu N, Song Y, Liang J, Shao Z, Sun Y, Dong Y, et al. 2018. Human pulse diagnosis for medical assessments using a wearable piezoelectric sensing system. Adv Funct Mater. 28(40):1803413.
Chung Y-F, Chu Y-W, Chung C-Y, Hu C-S, Luo C-H, Yeh C-C. 2013. New vision of the pulse conditions using biosensing pulse diagnosis instrument. 2013 1st International Conference on Orange Technologies (ICOT); pp. 5–8.
Cymbeknopp L, Bia D, Zocalo Y, Farro I, Armentano RL. 2011. Genderrelated differences in the excess pressure component of central aortic pressure waveform of healthy young. Annual International Conference of the IEEE Engineering in Medicine and society 6090037.
Fang EF, Scheibye-Knudsen M, Jahn HJ, Li J, Ling L, Guo H, Zhu X, Preedy V, Lu H, Bohr VA, et al. 2015. A research agenda for aging in China in the 21st century. Ageing Res Rev. 24(Pt B):197–205.
Gatzka CD, Kingwell BA, Cameron JD, Berry KL, Liang YL, Dewar EM, Reid CM, Jennings GL, Dart AM. 2001. Gender differences in the timing of arterial wave reflection beyond differences in body height. J Hypertens. 19:2197–2203.
Hayward CS, Kelly RP. 1997. Gender-related differences in the central arterial pressure waveform. J. Am. Coll. Cardiol. 30(7):1863–1871.
Jaafar NAL, Sidek KA, Azam SNAM. 2015. Acceleration Plethysmogram Based Biometric Identification. International Conference on BioSignal Analysis, Processing and Systems.
Jia D, Zhang D, Li N. 2014. Pulse waveform classification using support vector machine with Gaussian time warp edit distance kernel. Comput Math Methods Med. 2014:947254.

Kurdyak P, Lin E, Green D, Vigod S. 2015. Validation of a population-based algorithm to detect chronic psychotic illness. Can J Psychiatry. 60(8):362–368.

Laurent S, Boutouyrie P, Asmar R, Gautier I, Laloux B, Guize L, Ducimetiere P, Benetos A. 2001. Aortic stiffness is an independent predictor of all-cause and cardiovascular mortality in hypertensive patients. Hypertension. 37(5):1236–1241.

Lee BJ, Jeon YJ, Bae J-H, Yim MH, Kim JY. 2019. Gender differences in arterial pulse wave and anatomical properties in healthy Korean adults. Eur J Integr Med. 25:41–48.

Li Z, Nie F, Chang X, Nie L, Zhang H, Yang Y. 2018a. Rank-constrained spectral clustering with flexible embedding. IEEE Trans Neural Netw Learn Syst. 29(12):6073–6082.

Li Z, Nie F, Chang X, Yang Y, Zhang C, Sebe N. 2018b. Dynamic affinity graph construction for spectral clustering using multiple features. IEEE Trans Neural Netw Learn Syst. 29(12):6323–6332.

Li Z, Yao L, Chang X, Zhan K, Sun J, Zhang H. 2019. Zero-shot event detection via event-adaptive concept relevance mining. Pattern Recognit. 88:595–603.

Liu L, Zuo W, Zhang D, Li N, Zhang H. 2012. Combination of heterogeneous features for wrist pulse blood flow signal diagnosis via multiple kernel learning. IEEE Trans Inf Technol Biomed. 16(4):598–606.

Luo C-H, Su C-J, Huang T-Y, Chung C-Y. 2016. Non-invasive holistic health measurements using pulse diagnosis: I. Validation by three-dimensional pulse mapping. Eur J Integr Med. 8(6):921–925.

Nanyue W, Youhua Y, Dawei H, Bin X, Jia L, Tongda L, Liyuan X, Zengyu S, Yanping C, Jia W. 2015. Pulse diagnosis signals analysis of fatty liver disease and cirrhosis patients by using machine learning. Sci World J. 2015:859192.

Ouyang H, Tian J, Sun G, Zou Y, Liu Z, Li H, Zhao L, Shi B, Fan Y, Fan Y, et al. 2017. Self-powered pulse sensor for antidiastole of cardiovascular disease. Adv Mater. 29(40):1703456.

Rangaprakash D, Dutt DN. 2014. Analysis of wrist pulse signals using spatial features in time domain. In: IEEE International Conference on Communications and Signal Processing. Melmaruvathur (India) April 2014; pp. 345–348.

Rangaprakash D, Dutt DN. 2015. Study of wrist pulse signals using time domain spatial features. Comput Electr Eng. 45:100–107.

Rangaprakash D. 2014. Statistical analysis of wrist pulse signals obtained under different food intake conditions. In: Proceedings of IEEE International Conference on Communications and Signal Processing. Melmaruvathur (India); pp. 1410–1411.

Ruping S. 2001. Incremental learning with support vector machines. Proceedings 2001 IEEE International Conference on Data Mining 10.1109/ICDM.2001. pp. 989589.

Shafri HZM, Ramle FSH. 2008. A comparison of support vector machine and decision tree classifications using satellite data of Langkawi Island. Inf Technol J. 8(1):64–70.

Shu J-J, Sun Y. 2007. Developing classification indices for Chinese pulse diagnosis. Complement Ther Med. 15(3):190–198.

Tsai Y-N, Chang Y-H, Huang Y-C, Jui-Shan Lin S, Lee S-M, Cheng Y-Y, Su Y-C. 2019. The use of time-domain analysis on the choice of measurement location for pulse diagnosis research: A pilot study. J Chin Med Assoc. 82(1):78–79.

Vardoulis O, Saponas TS, Morris D, Villar N, Smith G, Patel S, Tan D. 2016. In vivo evaluation of a novel, wrist-mounted arterial pressure sensing device versus the traditional hand-held tonometer. Med Eng Phys. 38(10):1063–1069.

Wang D, Zhang D, Lu G. 2016. A robust signal preprocessing framework for wrist pulse analysis. Biomed Signal Process Control. 23:62–75.

Wang J, Liu K, Sun Q, Ni X, Ai F, Wang S, Yan Z, Liu D. 2019. Diaphragm-based optical fiber sensor for pulse wave monitoring and cardiovascular diseases diagnosis. J Biophotonics. 12(10):20190084.

Wang TJ, Gona P, Larson MG, Toftler GH, Levy D, Newton-Cheh C, Jacques PF, Rifai N, Selhub J, Robins SJ, et al. 2007. Multiple biomarkers for the prediction of first major cardiovascular events and death. J Vasc Surg. 45(3):638.

Weber T, Auer J, O’Rourke MF, Kvas E, Lassnig E, Berent R, Eber B. 2004. Arterial stiffness, wave reflections, and the risk of coronary artery disease. Circulation. 109(2):184–189.

Wu QW, Zhou D-X. 2005. SVM Soft margin classifiers: linear programming versus quadratic programming. Neural Computation. 17(5):1160–1187.

Xu L, Zhang Z, Gao F, Zhao X, Xun X, Kang Z, Liao Q, Zhang Y. 2021. Self-powered ultrasensitive pulse sensors for noninvasive multi-indicators cardiovascular monitoring. Nano Energy. 81:105614.

Yokota T, Zalar P, Kaltenbrunner M, Jinno H, Matsuhisa N, Kitanosako H, Tachibana Y, Yukita W, Koizumi M, Someya T. 2016. Ultra-flexible organic photonic skin. Sci Adv. 2(10):e15.

Yuesheng L. 2020. Physical meaning of pulse signal characteristics in TCM and the latest research progress of its recognition methods. 2nd International Symposium on the Frontiers of Biotechnology and Bioengineering. Yu-Feng C, Chung-Shing H, Cheng-Chang Y, Ching Hsing L. 2013. How to standardize the pulse-taking method of traditional Chinese medicine pulse diagnosis. Comput Biol Med. 43(4):342–349.