Proceeding of the IEEE
International Conference on Robotics and Biomimetics
Dali, China, December 2019

Continuous Blood Pressure Estimation from Two-Channel PPG Parameters by XGBoost

Xinze Che, Min Li, Wenbo Kang, Fujun Lai and Jiaxin Wang

Department of Mechatronic Engineering and Automation
Shanghai University

Room 421, Mechanical Building, No. 99, Shangda Road, Baoshan District, Shanghai, China

xinzeche@shu.edu.cn & min_li@shu.edu.cn

***Abstract* – Cuff-less and non-invasive continuous blood pressure (BP) measurement is essential for the monitoring of cardiovascular disease. It is widely recognized that the photoplethysmography (PPG) signal can track BP by extracting PPG feature points. In this study, the PPG signal of the two sensors fixed at the hand was extracted, include the time-domain features and the pulse transit time (PTT) features extracted by the two-channel PPG signal. This study proposes a novel feature selection algorithm to ensure continuous extraction of feature points from a low-quality PPG signal. The blood pressure estimator was trained by the XGBoost model, which can solve the problem of missing features. The data was collected from 20 subjects, and a total of 80 sets of data were used to train and validate our proposed model. The results attained the accuracy of 1.56 ± 3.39 mmHg for SBP and 3.11 ± 3.78 mmHg for DBP.**

***Index Terms – BP estimate, PPG, PTT,* time-domain features, XGBoost**

I. Introduction

Cardiovascular diseases (CVD) are the leading cause of death in the world. For the diagnosis and monitoring of many kinds of CVDs, continuous blood pressure monitoring is of great medical significance. Generally, only invasive devices can achieve accurate continuous blood pressure monitoring, while normal no-invasive blood pressure measuring equipment is mainly mercury column or cuff type electronic blood pressure monitors which cannot achieve continuous blood pressure measurement, and these devices can cause discomfort in patients. Therefore, a large number of scholars are engaged in the research in this field. the development of continuous blood pressure measuring equipment is also changing with each passing day.

PPG contains rich blood pressure information. Many studies use the PPG signal to analyze changes in blood pressure. PTT is the time delay for the pressure wave to travel through two arterial sites, also called "PAT" which based on the time difference between the R-peak of ECG and a characteristic point of PPG peak. There are also methods which get the PTT by calculating the time difference in PPG signal's peak point in different locations on the same artery. There is an apparent linear relationship between BP and PTT, which has been widely recognized. In C.C.Y. Poon's research [1], the relationship between BP and PTT was well established by mathematical models. In addition to the method of building mathematical models, more researchers use the method of establishing linear or nonlinear models to predict blood pressure by using the PTT feature [2-6]. In the study of Sharma [2], more than 10 kinds of linear formula methods have been validated. In their research, it is concluded that it is possible to predict the trends of BP by a PTT/PAT based mode. Based on the methods mentioned above, additional calibration is required for different subjects.

Different from traditional linear or mathematical models, many studies focus on using a machine learning method to predict BP values. In Mengyang Liu's study [7], their method combined the proposed 14 SDPPG features with the conventional 21 time-scale PPG features [8] to train the Support Vector Regression (SVR) model to achieve blood pressure estimator. It attained an accuracy of 8.54 ± 10.9 mmHg for systolic blood pressure (SBP) and 4.34 ± 5.8 mmHg for diastolic blood pressure (DBP). Nowadays, with the rise of neural network research, a large number of studies have begun to use different neural network models to predict BP [9-13]. There is study [9] using a typical-structure feed-forward ANN to estimate BP, achieving an accuracy of 4.02 ± 5.21 mmHg for SBP, 2.27 ± 1.82 mmHg for DBP. LSTM neural network is also a model used to estimate BP by some researches [10]. When using the root-mean-squared error (RMSE) to evaluate blood pressure predictions, the LSTM neural network model attained an accuracy of 2.751 mmHg for SBP and 1.604 mmHg for DBP, results are more accurate than the SVR model (12.38 mmHg for SBP and 6.34 for DBP) and PTT-based model (8.88 mmHg for SBP and 5.97 for DBP). Feature loss is a prevalent problem because the subject is difficult to ensure the stability of the body during the measurement process. Solving the problem of a feature missing is an important focus point for machine learning models and neural network models. Such problems make it difficult for the above models to estimate the correct results. For such cases, most machine learning models and neural network models are difficult to adapt to such situations.

In this paper, we proposed a novel continuous blood pressure estimation method using the eXtreme Gradient Boosting (XGBoost) model [15] as an estimator which can guarantee high estimation accuracy. The BP value could be estimated by XGBoost based model in the case of missing feature extraction caused by PPG waveform instability. Two-channel PPG signal was extracted from the hand, PTT parameters and other time-domain features were extracted as

978-1-7281-6321-5/19//$31.00 ©2019 IEEE

input parameters of the XGBoost model. All data were obtained from the two-channel PPG sensors fixed at the head. A feature selection algorithm was used to select waveform features for processing unstable PPG signals to ensure the accuracy of feature points. Based on this novel method proposed by this paper, the continuity of BP value estimation can be signicantly improved. In the second section, the methods are explained including data collection and pre-processing feature selection and model training. The proposed method is verified and concluded in the third section.

II. METHODOLOGY

A. Data collection and Pre-process

Data acquisition using two sets of pulse wave sensors named Max30102. A Max30102 sensor includes internal LEDs, photodetectors, optical elements, and low-noise electronics with ambient light rejection. Sensors are used to acquire the required information about blood motion in the form of the PPG signal. It integrates Red and IR LED drivers to modulate LED pulses for PPG measurements. Sensors are fixed at the palm position and the first phalanx position of the index finger both the same hand. Sensors are fixed as shown in Fig. 1.

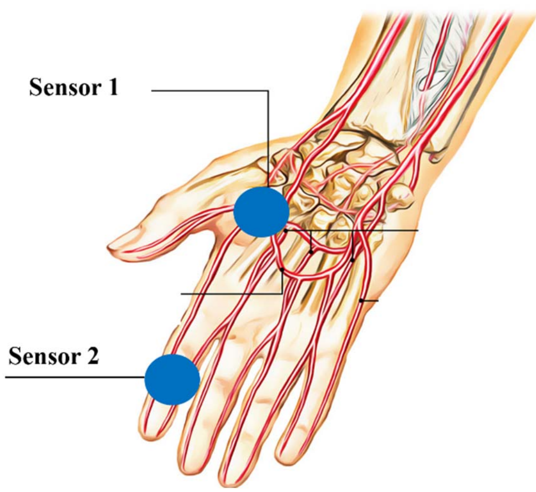

Fig. 1 Sensors fixed position on hand

PPG can often become contaminated by noise such as motion artifacts and electromyogram (EMG) signals. In order to obtain data with obvious characteristics, it is necessary to do pre-processing to increase the signal to noise ratio. This step can remove unnecessary information from the PPG signal and acquire waveform features that meet the needs of feature extraction. Excessive processing will result in the loss of features in the processed waveform. Reasonable pre-processing steps are needed to ensure the stable extraction of final waveform features. The figure below shows the data pre-processing workflow.

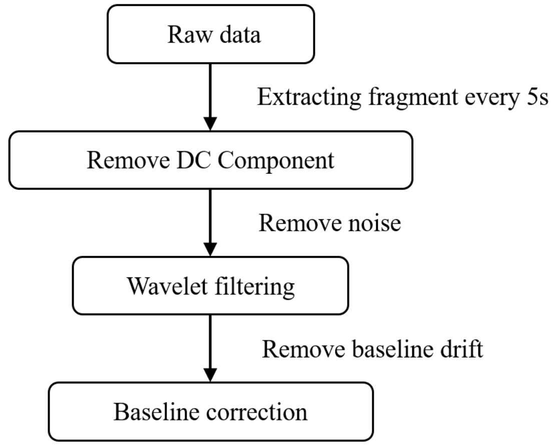

Fig. 2 Data pre-processing workflow

The following pre-processing steps are applied to process the noisy PPG signal.

1) Remove DC component

PPG signal is transmitted by arterial tissue signal (termed AC part) and other tissues (termed DC part). The DC component does not help the result and needs to be removed.

2) Filtering noise

PPG signal can be reconstructed by discrete wavelet transform which can remove the noise signal while retaining useful information. In Yadhuraj S. R's research [16], the reconstruction results of different wavelets such as db4, bior3.3, coif1, sym2, haar for PPG signal are verified. The Signal reconstruction is more accurate in db4 whereas the others are less effective. In this Paper, discrete wavelet transform based on db4 is also used to filter the noise signal.

3) Baseline correction

Due to the slight movement of the subject during the measurement, the amplitude of the signal is inconsistent, and the baseline is a drift state. This will result in the extraction of time-domain features not within a uniform standard.

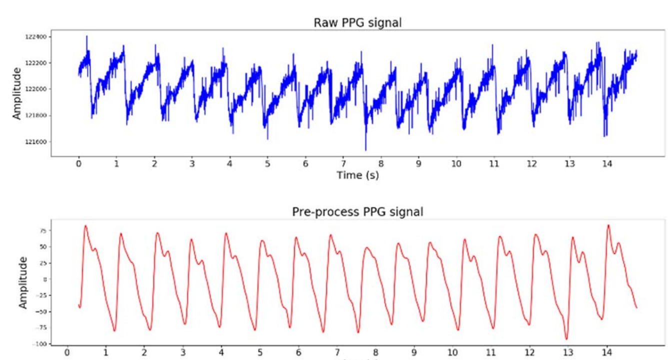

Fig. 3 Raw data and Pre-processed data

Fig. 3 is the comparison result of the collected signal for filtering. The pre-processed PPG signal can be observed stable waveforms and time-domain feature points.

B. *Feature Extraction*

There have been many studies to prove that PTT can be used to estimate BP through linear or nonlinear formulas. In Manuja Sharma's research [2], multiple models were used to estimate blood pressure by PTT. However, this method has obvious shortcomings. Due to the difference of each person's physical characteristics that the parameters in the formula need to advanced correction to get the desired prediction result. In addition, another type of PPG-based blood pressure estimation model is machine learning with features extraction by a single sensor signal of PPG. Machine learning methods can use multiple PPG features to build a model to estimate blood pulse in a small error range without correction. In this paper, feature points used by the above two methods are extracted and combined with the pre-processed waveform for selection.

1) Extracting PTT features based on two-channel PPG signal.

PTT is the measurement of the traveling time of blood between two points inside the body. PTT can be obtained by calculating the time difference between the same peak positions of the two sensor data. Many studies use the following linear regression model to calculate blood pressure continuously using PTT as the main parameter.

$$\mathrm{BP} = \frac{w_1}{PTT} + w_2 \tag{1}$$

The unknown parameters w_1 and w_2 are determined by a least-square algorithm during the calibration process. And the determination of parameters is performed separately in each subject.

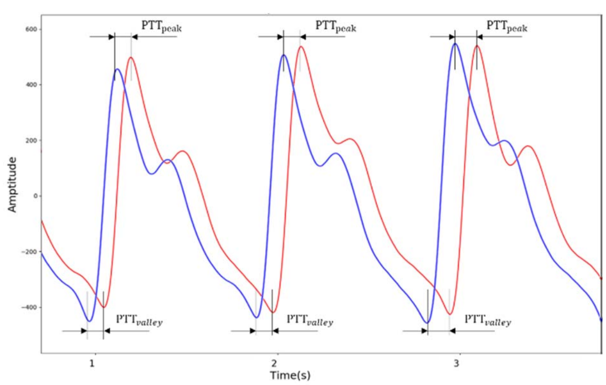

Fig. 4 PPT characteristics between two pulse wave.

Figure 4 is a diagram of PTT parameters extraction based on two PPG signal where PTT_{peak} represents the difference between two groups of PPG signal peak point and PTT_{vally} represents the difference between two groups of PPG signal valley point.

2) Extracting the time-domain characteristics of single-channel PPG signal.

The time-domain characteristics of the PPG signal contain a large number of human physiological characteristics. By mining time-domain features, more correlations between PPG and BP can be analyzed. Machine learning methods can be used to mining the implicit relationships, and estimates of blood pressure through these indicators. Figure 5 shows the time-domain features extracted from a single PPG signal. The meaning of the selected features is illustrated in Table 1.

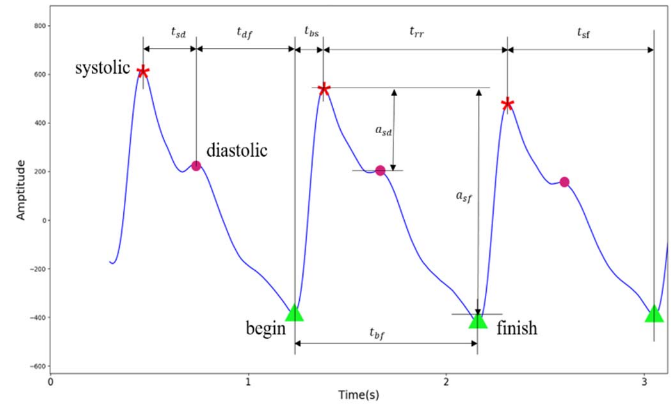

Fig. 5 Time-domain characteristics of the single pulse wave.

TABLE I
DEFINITIONS OF SELECTED FEATURES

Features	**Definitions**
t_{bf}: timeBeginToFinish	The time from the begin of the cycle to the finish
t_{bs}: timeBeginToSystolic	The time from the begin of the cycle to the systolic
t_{sd}: timeSystolicToDiastolic	The time from the of the cycle to the diastolic
t_{df}: timeDiastolicToFinish	The time from the diastolic of the cycle to the finish
t_{sf}: timeSystolicToFinish	The time from the systolic of the cycle to the finish
t_{rr}: timePeakToPeak	The time from the peak of the cycle to the peak of another cycle
a_{sd}: amplitudeSystolicToDiastolic	Amplitude difference between Systolic and diastolic.
a_{sf}: amplitudeSystolicToFinish	Amplitude difference between Systolic and Finish.

3) Feature selection

In the process of data acquisition, due to the slight movement of the subject during the measurement process and some equipment acquisition errors, even after the pre-processing of the signal, there still be a large number of feature points that cannot be accurately obtained. As a result, in the process of real-time calculation, all the feature points required by the machine learning model cannot be calculated. If all the features are discarded, the blood pressure in the current time segment cannot be calculated or there will be a lot of errors in the calculation. Faced with the characteristics of low-quality PPG signal, useful features need to be retained rather than abandon them completely. Only the standard feature points will be selected for use. Figure 6 shows the PPG signal with different waveform quality part. The main criterion for distinguishing the quality of waveforms is whether systolic point and diastolic point can be found. For part A, these two key feature points are hard to find. All feature points of Part A need to be completely abandoned. In part B, the only systolic point can be found. Therefore, the feature points associated with the diastolic need to be discarded while the feature points associated with the systolic need to be retained. All key points can be extracted accurately from Part C which is a normal quality waveform.

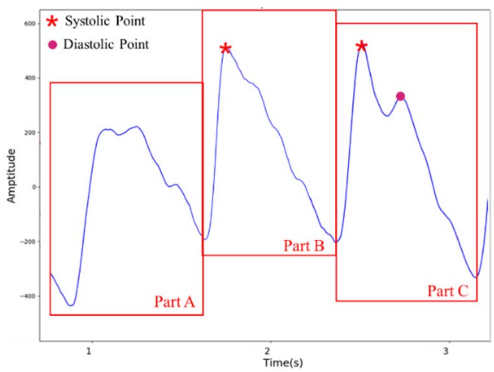

Fig. 6 PPG signal with different waveform quality part.

Useful waveforms are mainly Part A and Part B cases. Algorithm1 explains the feature selection process in processing the PPG signal. First, need to determine whether the systolic point meets the standard. In this step, all non-standard waveforms will be discarded directly. The second step is to detect the position of the diastolic point. Features that meet the criteria will be extracted.

Algorithm1 Features selection

i = 1, Num = # PulseWaveData points;
idxOfSystolic = findWavePeak(PulseWaveData);
idxOfDiastolic = findWavePeak(PulseWaveData);
idxOfVally= findWaveVally(PulseWaveData);
for each i ∈ {1, Num} **do**
    tmpBS = idxOfSystolic (i) – idxOfVally(i);
    tmpSF = idxOfVally(i+1) – idxOfSystolic (i);
    if δ_1 < tmpBS < δ_2 **and** δ_3 < tmpSF < δ_4 **then**
        t_{bs} = tmpBS, t_{sf} = tmpSF, $t_{bf} = t_{bs} + t_{sf}$;
        t_{rr} = idxOfSystolic (i+1) –idxOfSystolic (i);
        tmpSD = idxOfDiastolic(i) - idxOfSystolic (i);
        if δ_5 < tmpSD < δ_6 **then**
            t_{sd} = tmpSD;
            t_{df} = idxOfVally(i+1) - idxOfDiastolic(i);
            a_{sd} = PulseWaveData(idxOfDiastolic(i)) - PulseWaveData(idxOfSystolic(i))
            a_{sf} = PulseWaveData(idxOfVally(i+1)) - PulseWaveData(idxOfSystolic(i))
        else
            t_{sd} = None, t_{df} = None,
            a_{sd} = None, a_{sf} = None
        end if
    end if
end for

C. Model Training

1. XGBoost model training process

In the field of continuous blood pressure assessment, many studies have adopted the machine learning method. The blood pressure value is estimated through a large number of pulse wave feature points by trained model. A large number of prediction models are used, including multiple regression model, support vector machines (SVM) model, decision tree model, and neural network model. XGBoost [14] is an improved tree model based on Gradient Boosting Decision Tree. It is an integrated algorithm formed by combining the base function and weight with Boosting idea. The generalization ability of XGBoost is strong, and it can also handle with training and prediction with a large number of value missing cases. Compared with the traditional GBDT algorithm, XGBoost can perform parallel computing of multi-core CPUs, so its calculation speed has been significantly improved. In addition, when optimizing the target error function, XGBoost performs second-order derivative expansion. Compared with the traditional first-order expansion, XGBoost prediction accuracy has been greatly improved

The process of training the XGBoost model is as follows:

1) The first step is to initialize the model, give weight to all samples with the same value.
2) Iterative training of samples, and calculate the error weight of each sample at the m^{th} iteration.

$$err_m = \frac{\sum w_i I(y_i \neq G_m x)}{\sum w_i} \qquad (2)$$

3) Calculation of a_m

$$a_m = \log((1 - err_m)/err_m) \qquad (3)$$

4) Recalculate the weight of each sample at m^{th} iteration.

$$w_i \times e^{a_m \times I(y_i \neq G_m x)} \qquad (4)$$

5) After a training session, weights will be redistributed, and samples with larger errors will be assigned a higher weight. otherwise, they will be weighted down.

2. Training model

According to the feature extraction in the previous section, the characteristics of the two-channel pulse wave data can be obtained, which are related to the PTT-related features acquired based on the two-channel PPG signal, and the time-domain features obtained by the two single-channel PPG data respectively. All of these features are commonly used as input features of the XGBoost model. After calculating the model, in real-time calculations, the model can be used to calculate real-time blood pressure prediction results, and the test data can also be used as an update model training. The data is used to update the XGBoost model so that the next measurement is predicted using the updated model results. The following figure shows the workflow chart of the entire prediction method. The specific steps are:

1) The subject wears the device and begins measuring the two-channel PPG signal. At the same time, the subject measures the blood pressure value through a blood pressure measuring device
2) The device performs pre-processing on the collected data, extracts features, and retains the features which meet the standard.
3) Predict results using the XGBoost model and compare against measured BP values.
4) Test data enters the training database to increase training samples
5) Retraining the XGBoost model

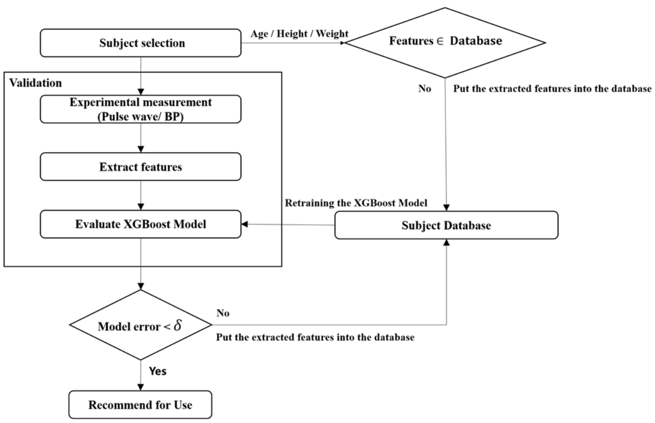

Fig. 7 The process by which the model is used and retrained.

III. RESULT AND ANALYSIS

A. *Subject selection and measurement*

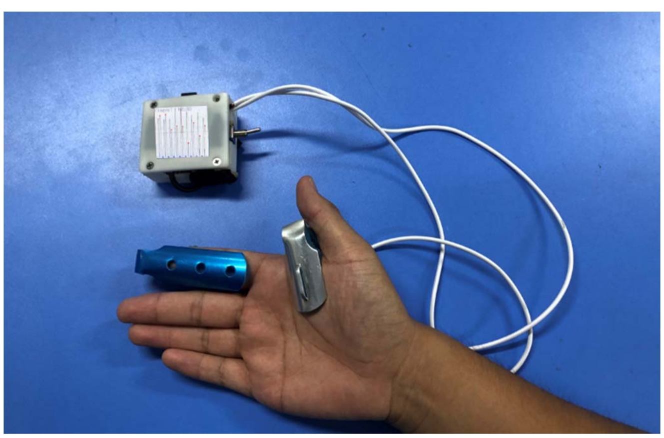

Fig. 8 The subject wears two-channel PPG measuring device

In this study, all two-channel PPG data was obtained using the self-made device. The device is shown in Fig. 8. The selection of subjects is shown in Table 2. A total of 20 subjects participated in the experiment, with 15 males and 5 females. All of them between the age of 18 and 35. In these test groups, the group containing normal blood pressure also included pre-hypertension and hypertensive subjects. In addition, the age, height, weight, date of birth, medication history and current physical state of these subjects were recorded.

TABLE II
SUBJECT STATISTICS

Number of subjects: 20				
Blood Pressure ranges:				
Blood pressure classification	**Systolic blood pressure(mmHg)**		**Diastolic blood pressure(mmHg)**	**Number**
Normal	<120	and	<80	10
Prehypertension	120-139	or	80-89	7
Hypertension	>140	or	>100	3
Gender:				
15 males and 5 females				
Age:				
Between 18 and 35 years old				

B. *Model verification*

In order to analyze the data, Mean-Absolute-Difference (MAD) and Relative-Mean-Square-Deviation (RMSD) will be used as a criterion for judging the accuracy of model predictions. These two indicators are the most commonly used in most studies. Their formulas are as follows:

$$\text{MAD} = (\textstyle\sum_{i=1}^{n} |p_i - y_i|)/\text{n} \quad (5)$$

$$\text{RMSD} = \sqrt{\frac{1}{n}\textstyle\sum_{i=1}^{n}(p_i - y_i)^2} \quad (6)$$

Where the p_i means the predicted value of i^{th} sample and the y_i means the measurement value of that sample. The accuracy of result in the form of (MAD ± RMSD) on SBP and DBP. Compare with the accuracy of 8.54 ± 10.9 mmHg for SBP and 4.34 ± 5.8 mmHg for DBP in the method of SVR [7], Our result attained an accuracy of 1.56 ± 3.39 mmHg for SBP and 3.11 ± 3.78 mmHg for DBP. Fig 8 shows the measured and predicted values of SBP and DBP.

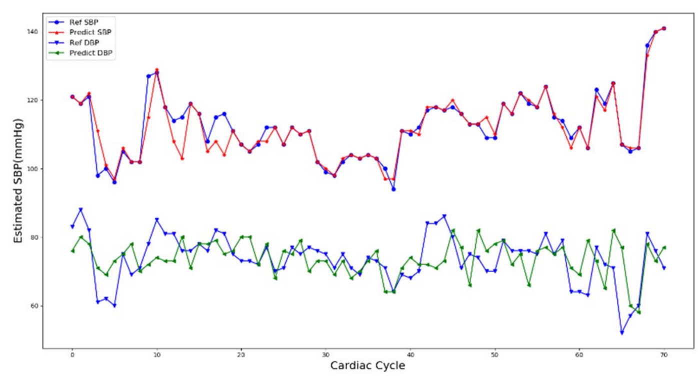

Fig. 9 Measured and predicted values of SBP and DBP

Bland-Altman plot (BA-plot) can visualize the distance between the predicted value and measurement value. For comparison, the same PPG feature was used to train the SVR model for verification. Fig. 9 shows the BA-plot of SBP and DBP. From the BA -plot analysis, it was obtained that for both SBP and DBP 12 out of 80 pairs of data points were located beyond the agreement limits.

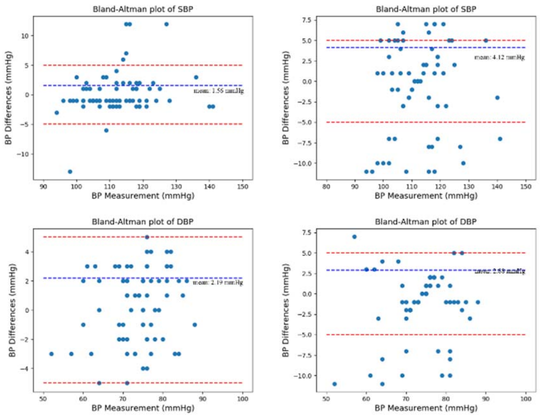

Fig. 9 Bland-Altman plot the SBP and DBP errors of the proposed model (left) and SVR model (right)

C. conclusion and discussion

The traditional blood pressure measuring device is uncomfortable to wear for the patient. And it is difficult to obtain a continuous segment in a common measuring instrument. The way of passing the pulse wave has been demonstrated by a large number of studies to obtain the blood pressure value of the patient. Many previous studies focused on how to establish linear regression model or machine learning model to predict blood pressure value by PTT extracted from PPG/ECG and some other time-domain parameters, but these methods can only extract accurate waveform parameters from well-formed waveform data which is difficult in the actual continuous pulse wave measurement process, the reason is that the slight movement of the body is unavoidable. This study solves the above problem by using feature selection algorithm combined with XGBoost model.

In this paper, we collect data by two sets of pulse wave sensors fixed at the hand. The sensors are fixed at the finger and the palm of the hand, then the two PTT features and the time-domain features of the single-channel data are extracted. In the face of unstable pulse wave data characteristics, the feature points that do not reach the threshold standard are discarded, and only the features that meet the requirements are put into the model as input parameters. We select XGBoost, a machine learning model, to predict blood pressure. Because this model can handle the lack of input features. In the model verification, it also proves that the use of XGBoost as a test model has a smaller MAD value in the blood pressure prediction process, and it attained an accuracy of 1.56 ± 3.39 mmHg for SBP and 3.11 ± 3.78 mmHg for DBP. Compared with the traditional linear model method, not only does it need to be corrected to deal with each subject, it's MAD value is smaller than that of the linear model (0 ± 7 mmHg). In this case, the relative error calculated by XGBoost is also smaller than the SVR model and bp neural network model used in other studies.

However, the method we proposed also has some limitations. First, when using XGBoost as a model to predict blood pressure, there are a lot of parameters to be adjusted. When retraining the model, it takes a long time to find the optimal parameters. It is beneficial to the rapid iterative update of the model in the real-time blood pressure measurement process, it is necessary to find a certain method to optimize the parameters search process of the XGBoost model.

ACKNOWLEDGMENT

We are thankful to the students of Shanghai University for participating in the collection and preparation of the experimental data. Min Li is the corresponding author.

REFERENCES

[1] C. C. Y.Poon and Y. T. Zhang, "Cuff-less and Noninvasive Measurements of Arterial Blood Pressure by Pulse Transit Time," 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference, Shanghai, 2005, pp. 5877-5880.

[2] Sharma, Manuja, et al. "Cuff-less and continuous blood pressure monitoring: a methodological review." Technologies 5.2 (2017): 21.

[3] Ghosh, Shrimanti, et al. "Continuous blood pressure prediction from pulse transit time using ECG and PPG signal." 2016 IEEE Healthcare Innovation Point-Of-Care Technologies Conference (HI-POCT). IEEE, 2016.

[4] Xing-Yun, Ding, Chang Qing, and Song Chao. "Study on the extract method of time domain characteristic parameters of pulse wave." 2016 IEEE International Conference on Signal and Image Processing (ICSIP). IEEE, 2016.

[5] Lin, Chen, et al. "Pulse waveform as an indicator of baseline offset in pulse transit time based blood pressure estimation." 2017 IEEE Healthcare Innovations and Point of Care Technologies (HI-POCT). IEEE, 2017.

[6] Pernice, Riccardo, et al. "Comparison of short-term heart rate variability indexes evaluated through electrocardiographic and continuous blood pressure monitoring." Medical & biological engineering & computing 57.6 (2019): 1247-1263.

[7] Liu, Mengyang, Lai-Man Po, and Hong Fu. "Cuffless blood pressure estimation based on photoplethysmography signal and its second derivative." International Journal of Computer Theory and Engineering 9.3 (2017): 202.

[8] Kurylyak, Yuriy, Francesco Lamonaca, and Domenico Grimaldi. "A Neural Network-based method for continuous blood pressure estimation from a PPG signal." 2013 IEEE International instrumentation and measurement technology conference (I2MTC). IEEE, 2013.

[9] Wang, Ludi, et al. "A novel neural network model for blood pressure estimation using photoplethysmography without electrocardiogram." Journal of healthcare engineering 2018 (2018).

[10] Lo, Frank P-W., et al. "Continuous systolic and diastolic blood pressure estimation utilizing long short-term memory network." 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2017.

[11] Wang, Yuemeng, et al. "Continuous blood pressure prediction using pulse features and Elman neural networks." 2017 IEEE 17th International Conference on Communication Technology (ICCT). IEEE, 2017.

[12] Xu, Zhihong, et al. "Continuous blood pressure estimation based on multiple parameters from eletrocardiogram and photoplethysmogram by Back-propagation neural network." Computers in Industry 89 (2017): 50-59.

[13] Rundo, Francesco, et al. "Advanced bio-inspired system for noninvasive cuff-less blood pressure estimation from physiological signal analysis." Computation 6.3 (2018): 46.

[14] Pan, Junjun, and Yue Zhang. "Improved blood pressure estimation using photoplethysmography based on ensemble method." 2017 14th International Symposium on Pervasive Systems, Algorithms and Networks & 2017 11th International Conference on Frontier of Computer Science and Technology & 2017 Third International Symposium of Creative Computing (ISPAN-FCST-ISCC). IEEE, 2017.

[15] Chen, Tianqi, and Carlos Guestrin. "Xgboost: A scalable tree boosting system." Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. ACM, 2016.

[16] Yadhuraj, S. R, H. Harsha, and K. V. Padmaja. "REMOVAL OF NOISE IN PPG SIGNAL USING WAVELETS." International Journal of Computer Science & Mobile Computing 2.6(2013).