A Novel Continuous Blood Pressure Estimation Approach Based on Data Mining Techniques

Fen Miao, Nan Fu, Yuan-Ting, Zhang, *Fellow, IEEE*, Xiao-Rong Ding, Xi Hong, Qingyun He, and Ye Li

***Abstract*—Continuous blood pressure (BP) estimation using pulse transit time (PTT) is a promising method for unobtrusive BP measurement. However, the accuracy of this approach must be improved for it to be viable for a wide range of applications. This study proposes a novel continuous BP estimation approach that combines data mining techniques with a traditional mechanism-driven model. First, 14 features derived from simultaneous electrocardiogram and photoplethysmogram signals were extracted for beat-to-beat BP estimation. A genetic algorithm-based feature selection method was then used to select BP indicators for each subject. Multivariate linear regression and support vector regression were employed to develop the BP model. The accuracy and robustness of the proposed approach were validated for static, dynamic, and follow-up performance. Experimental results based on 73 subjects showed that the proposed approach exhibited excellent accuracy in static BP estimation, with a correlation coefficient and mean error of 0.852 and -0.001 ± 3.102 mmHg for systolic BP, and 0.790 and -0.004 ± 2.199 mmHg for diastolic BP. Similar performance was observed for dynamic BP estimation. The robustness results indicated that the estimation accuracy was lower by a certain degree one day after model construction but was relatively stable from one day to six months after construction. The proposed approach is superior to the state-of-the-art PTT-based model for an approximately 2-mmHg reduction in the standard derivation at different time intervals, thus providing potentially novel insights for cuffless BP estimation.**

***Index Terms*—Continuous blood pressure (BP), feature selection, multivariate linear regression (MLR), support vector regression (SVR).**

Manuscript received July 28, 2016; revised January 11, 2017 and March 8, 2017; accepted March 31, 2017. Date of publication April 28, 2017; date of current version November 3, 2017. This work was supported in part by the National Natural Science Foundation of China under Grant 61502472, in part by the National 863 Project of China (SS2015AA020109), in part by the Basic Research Program of Shenzhen (JCYJ20150630114942316), in part by the Science and Technology Planning Project of Guangdong Province (2015B010129012), and in part by the Guangdong Province Special Support Program (2014TX01X060). *(Corresponding author: Ye Li.)*

F. Miao, N. Fu, Y.-T. Zhang, X. Hong, Q. He, and Y. Li are with the Key Laboratory for Health Informatics of the Chinese Academy of Sciences, Shenzhen Institutes of Advanced Technology, Shenzhen 518172, China (e-mail: fen.miao@siat.ac.cn; nan.fu1@siat.ac.cn; ytzhangapple@icloud.com; xi.hong@siat.ac.cn; qy.he@siat.ac.cn; ye.li@siat.ac.cn).

X.-R. Ding is with the Department of Electronic Engineering, Chinese University of Hong Kong, Hong Kong (e-mail: xrding@ee.cuhk.edu.hk).

Digital Object Identifier 10.1109/JBHI.2017.2691715

I. Introduction

HYPERTENSION is one of the most critical predictors of cardiovascular disease (CVD), which is the leading cause of death worldwide. Each incremental elevation of 20/10 mmHg in systolic blood pressure/diastolic blood pressure (SBP/DBP) over 115/75 mmHg doubles the risk of CVD [1]. Ambulatory blood pressure (BP) and related BP variability have been shown to be more reliable predictors of CVD than BP measured in a clinical setting [2], [3]. Although traditional 24-hour BP devices can monitor BP at regular intervals through repeated inflation with a cuff, such methods are discontinuous and unsuitable for daily use. Developing an unobtrusive device for high-resolution BP monitoring is thus of great significance for real-time hypertension detection and would thus benefit CVD prevention [4].

Several continuous BP measurement methods have been proposed, but all of them are performed either manually or with a cuff and are thus impractical for constant monitoring [5], [6]. Pulse transit time (PTT) is a potential indicator for BP estimation, referring to the time for a pulse wave to travel between two locations in the cardiovascular system. PTT can be calculated from two pulse signals generated by the cardiovascular system, such as electrocardiogram (ECG) and photoplethysmogram (PPG) signals [7]. The mechanism-driven BP estimation approach of using PTT has been extensively studied over the past 15 years [8]–[13]. In 2001, Chan *et al.* proposed a linear model for estimating BP with PTT [8]. A nonlinear model was proposed by Poon *et al.* [10] in 2005, and it attained an accuracy of 0.6 ± 9.8 mmHg for SBP and 0.9 ± 5.6 mmHg for DBP. In 2015, Ding *et al.* [14] proposed the PPG intensity ratio (PIR) as a crucial DBP indicator. In that study, the combination of PIR with PTT (PTT+PIR) outperformed previous PTT algorithms, achieving an accuracy of -0.37 ± 5.21 mmHg for SBP, -0.08 ± 4.06 mmHg for mean BP, and -0.18 ± 4.13 mmHg for DBP. In a 24-hour correlation study between BP and PTT [13], PTT correlated closely with BP at night time, but the correlation was limited during the daytime because of confounding factors such as vascular tone [15] and the pre-ejection period (PEP) [16]. Therefore, extant mechanism-driven BP estimation approaches that use PTT have limited accuracy. Additionally, frequent calibrations must be performed to ensure the estimation accuracy [17]. In that study, the impact of the length of the calibration interval on the estimation accuracy of the PTT-based BP approach was also studied over 15-min, 2-week, and 1-month

periods. The experimental results showed that longer calibration intervals reduce the accuracy of both the SBP and DBP estimates. Overall, PTT-based BP mechanism models are inferior in accuracy and robustness, especially in long-term tracking performance, because the fixed relationship hypothesis is highly influenced by various factors such as vascular tone, physiological status, and individual variability.

Data mining can be used to automatically and accurately reconstruct relationships between variables and response values from big data. It provides an efficient means for overcoming the weaknesses of the mechanism-driven model [18]. Data mining has been widely applied over the past few decades, including in investigations of factors influencing BP [19]–[21]. Moreover, numerous features extracted from PPG signals as well as the acceleration of PPG signals have been proposed for quantifying vascular tone, and these factors correlate favorably with BP [22], [23]. Multivariate analysis using data mining techniques might improve the accuracy of mechanism-based BP estimation approaches by including more features reflecting vascular tone and physiological status. Only a few researchers have attempted to develop a BP estimation model through multivariate analysis [24]–[26]. A random forest-based BP estimation approach was recently proposed by He [27], who used the Multi parameter Intelligent Monitoring in Intensive Care research database. All of these approaches, however, lack an intrinsic mechanism analysis, and their long-term tracking performance has not been validated over different time intervals.

In this paper, we describe a novel approach for estimating continuous BP, which involves a combination of data mining techniques with traditional mechanism-based models. This study adds several contributions to the field of continuous BP measurement methods. First, in contrast to previous mechanism-driven models based on the fixed hypothesis of the PTT–BP relationship for different subjects, the proposed personalized BP model is based on individual patterns derived from data mining. Second, compared with previously proposed methods, our method extracts more BP indicators from simultaneous ECG and PPG signals for each subject and determines the relative importance for each subject by using a genetic algorithm-based feature selection method. Finally, the robustness of the proposed approach was fully validated at different time intervals after model construction. The accuracy and robustness of the proposed approach were verified by comparing them with the state-of-the-art mechanism-based method presented in [14].

II. Methodology

A. Basic Principles and Feature Extraction

As presented in [28] by Hughes *et al.*, vascular elasticity, E, can be expressed as

$$E = E_0 e^{\alpha P} \tag{1}$$

where P is the mean BP, E_0 denotes the vascular elasticity when the pressure is 0, and α is a correction factor. Equation (1) can be transferred using the following logarithm:

$$P = \frac{1}{\alpha} \ln \frac{E}{E_0} \tag{2}$$

TABLE I
Definitions of the Selected Features

Features	Definitions
f_1 : Heart rate	Time span between two adjacent ECG R wave
f_2 : PTT_ppgBottom	Time span between the ECG R wave and the bottom of the simultaneously collected PPG
f_3 : PTT_ppgPeak	Time span between ECG R wave and the peak of the simultaneously collected PPG
f_4 : PTT_MaxDeri	Time span between ECG R wave and the maximum first derivative of the simultaneously collected PPG
f_5 : Systolic time	Ascending time from PPG foot to PPG peak
f_6 : ppgFirstDeriHeight	Intensity of the first derivate of the PPG waveform
f_7 : ppgFirstDeriWidth	Time width of the first derivate of the PPG waveform
f_8 : ppgSecondDeriHeight	Total intensity of the second derivate of the PPG waveform
f_9 : ppgSecondDeriPeakHeight	Peak intensity of the second derivate of the PPG waveform
f_{10} : ppgSecondDeriFootHeight	Foot intensity of the second derivate PPG waveform
f_{11} : ppgSecondDeriWidth	Time width of the second derivate of the PPG waveform
f_{12} : PIR	Ratio of PPG peak intensity to PPG bottom intensity
f_{13} : Diastolic time	Descending time from PPG peak to PPG foot
f_{14} : ppg_k	PPG characteristic value

From (2), we can see that P depends on E/E_0, which indicates the degree of arteriosclerosis. According to a previous study [22], arteriosclerosis can be evaluated using factors extracted from the second derivative of the PPG signal (2nd PPG). Therefore, features extracted from the 2nd PPG might be indicators for BP measurement.

Alternatively, the degree of arteriosclerosis can be inferred from the Moens–Korteweg (M–K) equation, which shows that the pulse pressure, PP (i.e., the difference between SBP and DBP), is proportional to $1/PTT^2$ [10], [13], [15]; that is,

$$PP = PP_o . \left(\frac{PTT_0}{PTT} \right)^2 \tag{3}$$

where PP_0 and PTT_0 are the initial calibrated PP and PTT, respectively.

The two-element Windkessel model presented in [29] shows that DBP is proportional to the reciprocal of PIR:

$$DBP \propto \frac{1}{PIR} \tag{4}$$

With the initial calibrated PIR_0 and DBP_0, the following highly accurate BP estimation model was proposed in [14]:

$$DBP = DBP_0 \frac{PIR_0}{PIR} \tag{5}$$

$$SBP = DBP_0 \frac{PIR_0}{PIR} + PP_0 \left(\frac{PTT_0}{PTT} \right)^2 \tag{6}$$

Based on the aforementioned studies, 14 features that potentially influence BP were considered in the present study and are depicted in Fig. 1. The definitions of the selected features are

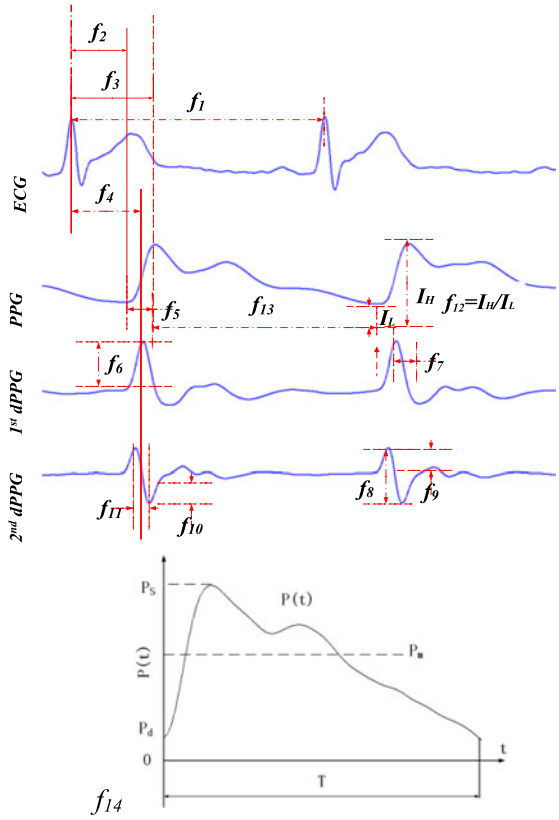

Fig. 1. Extracted features.

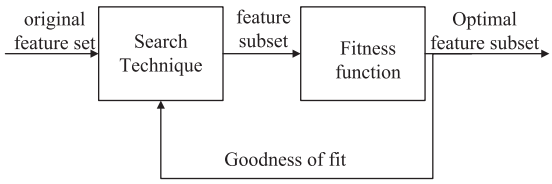

Fig. 2. Block diagram of the feature selection process.

listed in Table I, with ppg_K defined as

$$ppg\_K = \frac{p_m - p_d}{p_s - p_d} \tag{7}$$

where $p_m = \frac{1}{T}\int p_t dt$.

B. Feature Selection

Because of the possibility of introducing irrelevant and redundant features, a feature selection process should be employed before model construction to determine the optimal feature subset and avoid over-fitting. Genetic algorithms can identify near-optimal solutions for such optimization problems [30]. Genetic algorithms use iterations to evolve a candidate solution with a satisfactory fitness level. Fig. 2 presents a block diagram of the feature selection process using a genetic algorithm. This process can be considered a combination of a search technique for generating a new feature subset and a fitness function for scoring different subsets and selecting the optimal one. First, a subset of features is randomly generated from the extracted feature set presented in Table I. In each generation, the fitness

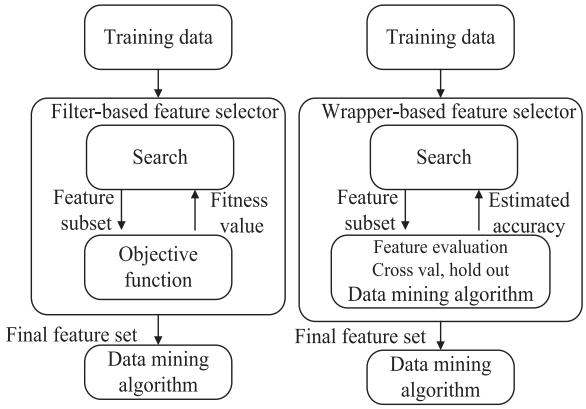

Fig. 3. Filter and wrapper feature selectors [32].

of each feature in the population is evaluated on the basis of the value of fitness function in the optimization problem. The features with a closer fit are selected and modified to form a new generation. This new generation of candidate solutions is then used in the subsequent generation. The algorithm terminates when either a maximum number of generations has been reached or a satisfactory fitness level has been achieved.

The adopted search technique and proposed fitness function are detailed as follows.

1) Search technique: the search technique is a "survival of the fittest" process. In our study, we adopted the roulette wheel selection method, which is the simplest and most common method used in genetic algorithms [31]. The probability of selection is proportional to the fitness level. If f_i is the fitness value of an individual,i,in the population, its probability of being selected is

$$p_i = \frac{f_i}{\sum_{j=1}^{N} f_j} \tag{8}$$

where N is the population size.

2) Fitness function: an appropriate fitness function is an essential step for feature selection using genetic algorithms. According to different fitness functions, a feature selector can be divided into filter- and wrapper-based selectors [32], as depicted in Fig. 3. Filter-based selectors evaluate the fitness of a feature subset by using the intrinsic properties of the data by ignoring the classifier, whereas wrapper-based selectors evaluate the fitness of a feature subset on the basis of accuracy estimates from subsequent data-mining processes. Wrapper-based selectors are considerably slower than filter-based selectors because they must repeatedly call the data-mining algorithm and re-run while switching to a different algorithm; therefore, a filter-based selector was considered in the present study.

Therefore, a correlation coefficient-based fitness function, which is independent of a specific type of classifier, was proposed for the regression model. Let the response values be $y_1, y_2, \cdots, y_N$ and the corresponding n-dimensional feature sets be $X_1, X_2, \cdots, X_N$ respectively. For each pair of $\{y_i, y_j\}$,

the distance for the response values can be computed as follows:

$$D_y = d_{y_i,y_j} = y_i - y_j \tag{9}$$

Concurrently, the distance of the corresponding feature sets can be denoted as the Euclidean distance:

If $D_y \geq 0$,

$$D_X = d_{X_i,X_j} = \sqrt{\frac{(X_{i,1}-X_{j,1})^2+(X_{i,2}-X_{j,2})^2+\cdots(X_{i,n}-X_{j,n})^2}{n}} \tag{10}$$

If $D_y < 0$,

$$D_X = d_{X_i,X_j} = -\sqrt{\frac{(X_{i,1}-X_{j,1})^2+(X_{i,2}-X_{j,2})^2+\cdots(X_{i,n}-X_{j,n})^2}{n}} \tag{11}$$

The correlation coefficient between D_X, D_y can be expressed as follows:

$$R = \frac{S_{D_X D_y}}{\sqrt{S_{D_X} S_{D_y}}} \tag{12}$$

where

$$S_{D_X D_y} = \frac{\sum_i (D_{X_i} - \bar{D}_X)(D_{y_i} - \bar{D}_y)}{n-1}$$

$$S_{D_X} = \frac{\sum_i (D_{X_i} - \bar{D}_X)^2}{n-1}, S_{Dy} = \frac{\sum_i (D_{y_i} - \bar{D}_y)^2}{n-1} \tag{13}$$

As the goodness of fit improves, the correlation coefficient becomes larger. Therefore, the optimization problem in our study was defined as the maximization of the fitness function R.

C. Model Construction and Validation

Two multivariate analysis methods, namely multivariate linear regression (MLR) and support vector regression (SVR),were adopted to construct the BP model on the basis of the features selected by the genetic algorithm. MLR is the most commonly used method because of its low computational complexity and easily interpretational power. However, because the extracted features often have nonlinear effects on the BP value, MLR is prone to low accuracy and high instability. Therefore, a nonlinear regression model, SVR [33], an efficient method for reconstructing nonlinear relationships between variables and estimation results, was employed in our study. We validated the performance of the proposed models by comparing them with the PTT+PIR-based model presented in [14].

1) MLR: As the most widely used regression method, MLR models the relationship between a response variable and explanatory variables by fitting a linear formula from observed data. In our study, BP was the response variable and $f_1, f_2, \cdots f_p$ were the independent variables derived from feature selection process. The relationship between BP and the features can be expressed as follows:

$$BP = \beta_0 + \beta_1 f_1 + \beta_2 f_2 + \cdots \beta_p f_p \tag{14}$$

Fig. 4. Experimental scenario.

where $\beta_1, \beta_2, \cdots \beta_p$ denotes the regression coefficient, which can be fitted from a training dataset and then used for estimation.

2) SVR: Support vector machines are specific class of algorithm that can be applied to both classification problems and regression by using kernels to reflect nonlinear relationships. In SVR, the input feature set, f, is first mapped to an m-dimensional feature space through nonlinear mapping, wherein a linear model is then constructed. The linear model in the feature space for BP estimation in our study is expressed as follows:

$$BP = \sum_{j=1}^{m} \omega_j g_j(\mathrm{f}) + b \tag{15}$$

where $g_j(\mathrm{f}), j = 1, \ldots, m$ denotes a set of nonlinear transformations, and b is the bias term.

D. Experiment

To validate the proposed MLR-based BP estimation model, three experiments were conducted, namely a static experiment, a dynamic experiment, and a follow-up experiment. We conducted the static BP estimation experiment on 73 healthy adults (40 men and 33 women) with a mean age of 26.41 ± 4.26 years (range, 21–43 years). Fig. 4 shows the experimental scenario. A continuous blood pressure monitor (Finpres® NOVA, SMART Medical, UK) was used to measure the reference BP with the finger cuff on the left middle finger and the brachial cuff on the left upper arm. ECG and PPG signals were acquired simultaneously using a multi-parameter monitor (Biopac system). I-lead ECG signals were collected with ECG electrodes placed on the left and right arms and right leg. PPG signals were collected from the left index finger. All tests lasted 10 min and were performed with the subjects in a seated position. The sample rate was set to 1000 Hz. In the dynamic experiment, 35 available subjects (20 men and 15 women) with a mean age of 27.89 ± 4.98 years (range, 22–42 years) were required to perform rope skipping for 5 min to produce dynamic BP changes. The same experimental procedure was adopted for each subject. The overall mean SBP for all subjects was 114.88 ± 16.04 mmHg (static experiment) and 132.98 ± 19.81 mmHg (dynamic experiment). The corresponding means were 65.98 ± 11.32 and 74.78 ± 13.09 mmHg for DBP and 260.39 ± 24.37 and 235.25 ± 24.29 ms for PTT. In the static and dynamic experiments, the original dataset was separated into two subsets, each 5 min in length. The subset from the first 5 min was used to develop the BP model for each

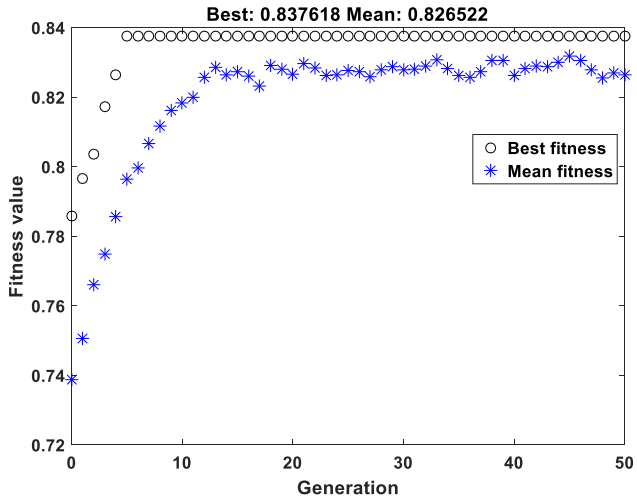

Fig. 5. Improvement in feature set fitness over time.

subject, whereas that from the final 5 min was used for performance validation. In the follow-up experiment, 10 available subjects were followed up at 1 day, 3 days, and then 6 months after the first experiment to investigate the robustness of the proposed approach, with the same experimental procedure used each time. The dataset from the follow-up experiment was used to validate the robustness of the model constructed from the static experiment in tracking long-term BP.

The study was approved by the Institutional Ethics Committee of the Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences. Informed consent was obtained from all subjects before the experiment.

III. Experimental Results

A. Feature Selection Results

The aforementioned feature selection method was employed to select the most critical features for each subject. Fig. 5 shows an example of the evolution of a subject's mean and highest correlation coefficient as a function of generations. The overall increase in the mean correlation coefficient demonstrates that using the genetic algorithm eliminates models with low predictive power while supporting those with high predictive power. The predictive power stabilized after 10 generations, as indicated by the optimal correlation coefficient curve.

Additionally, we found that the selected features differed between subjects as a result of individual variability, as presented in Table II. To evaluate the importance of each variable in the population, a weight-based strategy was proposed. The importance level for feature f_i can be expressed as

$$FI_i = \sum_{n=1}^{N} \omega_{i,n} \Big/ N \tag{16}$$

where N is the total number of subjects in the experiment and $\omega_{i,n}$ is the weight of f_i for the nth subject. If f_i is not selected as a critical feature, $\omega_{i,n} = 0$; otherwise, $\omega_{i,n}$ is the relative weight for improving the fitness value in the genetic algorithm, which

TABLE II
Selected Features for 10 Randomly Selected Subjects

	1	2	3	4	5	6	7	8	9	10
f1	√	√								√
f2	√	√			√					
f3								√	√	
f4	√	√	√	√	√		√		√	√
f5				√						
f6	√	√		√						
f7			√		√					
f8										
f9										
f10		√		√			√			
f11		√		√	√				√	√
f12			√		√	√	√	√		√
f13										
f14	√			√		√			√	√

can be determined as follows:

$$\omega_{i,n} = R_{i,n} \Big/ \sum_{j=1}^{K} R_{j,n} \tag{17}$$

where $R_{i,n}$ is the fitness value of the selected feature f_i for the nth subject, and K is the total number of selected features. Table III presents the feature importance of each variable in the static and dynamic experiments. The table shows that the distribution of feature importance was similar between the static and dynamic experiments. Specifically, the dominant features for SBP were found to be PTT_MaxDeri, ppg_k, PIR, ppgSecondDeriWidth, and ppgSecondDeriFootHeight, whereas the dominant features for DBP were PIR, ppg_k, ppgSecondDeriWidth and ppgSecondDeriFootHeight (see Discussion for further details).

B. Accuracy Performance of the Proposed Models

We evaluated the overall accuracy performance of the proposed BP models according to Pearson's correlation coefficient (CC), the mean difference (MD), and the difference in the standard derivation (SD) with the Finapres BP measurement as the reference. The CC is a measure of consistency from a mathematical perspective, MD is a measure of bias in the BP estimation, and SD is a measure of error variability. Formulas for these indices are as follows:

$$CC = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{18}$$

$$MD = \frac{\sum_{i=1}^{n}(y_i - x_i)}{n} \tag{19}$$

$$SD = \sqrt{\frac{\sum_{i=1}^{n}(y_i - x_i - MD)^2}{n - 1}} \tag{20}$$

TABLE III
STATISTICAL RESULTS FOR THE FEATURE IMPORTANCE OF EACH VARIABLE. (A) STATIC EXPERIMENT FOR 73 SUBJECTS. (B) DYNAMIC EXPERIMENT FOR 35 SUBJECTS

(a)

Features	SBP	DBP
f1: Heart rate	0.01	0.05
f2: PTT_ppgBottom	0.07	0.05
f3: PTT_ppgPeak	0.04	0.01
f4: PTT_MaxDeri	0.19	0.08
f5: Systolic time	0.03	0.02
f6: ppgFirstDeriHeight	0.07	0.06
f7: ppgFirstDeriWidth	0.05	0.08
f8: ppgSecondDeriHeight	0.05	0.02
f9: ppgSecondDeriPeakHeight	0.05	0.07
f10: ppgSecondDeriFootHeight	0.11	0.11
f11: ppgSecondDeriWidth	0.10	0.12
f12: PIR	0.13	0.16
f13: Diastolic time	0.01	0.06
f14: ppg_k	0.14	0.11

(b)

Features	SBP	DBP
f1: Heart rate	0.08	0.09
f2: PTT_ppgBottom	0.07	0.05
f3: PTT_ppgPeak	0.03	0.02
f4: PTT_MaxDeri	0.20	0.07
f5: Systolic time	0.04	0.01
f6: ppgFirstDeriHeight	0.02	0.05
f7: ppgFirstDeriWidth	0.02	0.08
f8: ppgSecondDeriHeight	0.01	0.03
f9: ppgSecondDeriPeakHeight	0.01	0.03
f10: ppgSecondDeriFootHeight	0.10	0.10
f11: ppgSecondDeriWidth	0.10	0.11
f12: PIR	0.12	0.16
f13: Diastolic time	0.05	0.05
f14: ppg_k	0.15	0.15

TABLE IV
STATISTICAL ANALYSIS FOR PERFORMANCE WITH MLR (A) STATIC PERFORMANCE FOR 73 SUBJECTS. (B) DYNAMIC PERFORMANCE FOR 35 SUBJECTS. (C) LONG-TERM (6 MONTHS) PERFORMANCE FOR 10 SUBJECTS

(a)

	Index	Mean	Standard derivation	Max	Min
SBP (mmHg)	CC	0.824	0.084	0.939	0.538
	MD (mmHg)	0.0016	0.0199	0.0896	-0.004
	SD (mmHg)	3.449	1.069	8.101	1.695
DBP (mmHg)	CC	0.754	0.120	0.919	0.368
	MD (mmHg)	0.0017	0.017	0.099	0.0004
	SD (mmHg)	2.468	0.779	6.238	1.369

(b)

	Index	Mean	Standard derivation	Max	Min
SBP (mmHg)	CC	0.941	0.038	0.969	0.794
	MD (mmHg)	-0.046	0.136	0.063	0.001
	SD (mmHg)	4.705	0.843	5.949	3.300
DBP (mmHg)	CC	0.923	0.093	0.970	0.557
	MD (mmHg)	-0.071	0.056	0.279	-0.021
	SD (mmHg)	2.839	0.228	3.459	2.406

(c)

	Index	Mean	Standard derivation	Max	Min
SBP (mmHg)	CC	0.619	0.154	0.857	0.373
	MD (mmHg)	-1.267	4.678	6.813	0.569
	SD (mmHg)	5.98	1.753	8.852	2.680
DBP (mmHg)	CC	0.549	0.175	0.814	0.178
	MD (mmHg)	-1.38	5.926	9.83	0.806
	SD (mmHg)	5.49	0.934	8.184	2.751

where $\{x_1, x_2, \cdots x_n\}$ are the BP values from the Finapres device and $\{y_1, y_2, \cdots y_n\}$ are the estimated BP values.

1) Multivariate Linear Regression-Based Model: Table IV gives the statistical analysis results of the aforementioned performance indices with MLR. In the table, the static performance in 73 subjects is shown in Panel (a), the dynamic performance in 35 subjects is given in Panel (b), and the long-term tracking performance at 6 months in 10 subjects is given in Panel (c). For the static experiment, the mean CC, MD, and SD for SBP and DBP are 0.824 and 0.754; 0.0016 and 0.0017 mmHg; and 3.449and 2.468 mmHg, respectively. For the dynamic experiment, the estimation results also exhibited favorable performance, with a mean CC, MD, and SD for SBP and DBP of 0.941 and 0.923; −0.046 and −0.071 mmHg; and 4.705 and 2.839 mmHg, respectively. For the long-term tracking experiment, the performance was lower to a certain degree, with a mean CC, MD, and SD for SBP and DBP of 0.619 and 0.549; −1.267 and −1.38 mmHg, and 5.98 and 5.49 mmHg, respectively.

Fig. 6 gives a typical example of the correlation and Bland–Altman plot for the proposed SBP and DBP estimations relative to the Finapres BP measurements. The correlation plot shows that the CCs for the static estimation between the SBP and DBP estimates and the Finapres measurements were 0.91 and 0.77, respectively, indicating a high correlation between the estimated BP and the reference. Similar results were obtained for the dynamic estimation, with CCs of 0.92 and 0.75, respectively. The long-term estimation underperformed the static and dynamic estimations by approximately 20% in CC. The Bland–Altman plot indicates that most of the estimated points for SBP and DBP are within the limits of agreement (bias $\pm$ 1.96 $\times$ SD), demonstrating that the estimated BP with the proposed method approximate those measured by the Finapres system in the static, dynamic, and long-term estimations.

2) SVR-Based Model: Table V(a)–(c) respectively shows the statistical performance of the static, dynamic, and long-term BP tracking approach with SVR. Panel (a) indicates that, in the static estimation, the mean CC, MD, and SD for SBP and DBP were 0.852 and 0.790;−0.001 and −0.004 mmHg; and 3.102 and 2.199 mmHg, respectively. The estimation results from the dynamic experiment also showed favorable performance, with a mean CC, MD, and SD of 0.946 and 0.929; −0.085 and −0.077 mmHg; and 4.493 and 2.735 mmHg, respectively. Similar to the proposed approach with MLR, the long-term tracking performance was lower to a certain degree, with a mean CC, MD, and SD for SBP and DBP of 0.617 and 0.554; −1.148 and −1.194 mmHg; and 5.79 and 5.29 mmHg, respectively.

Fig. 7 presents a typical example of the correlation and Bland–Altman plot for the proposed SBP and DBP estimations compared with the Finapres BP with SVR. The CCs between the SBP and DBP estimates and the Finapres measurements were 0.92 and 0.82, indicating very high correlations between the BP estimates and the reference for the static estimation. Similar results were obtained for the dynamic estimation, which attained CCs of 0.92 and 0.78 for SBP and DBP, respectively. The Bland–Altman plot also indicates that the estimated BP

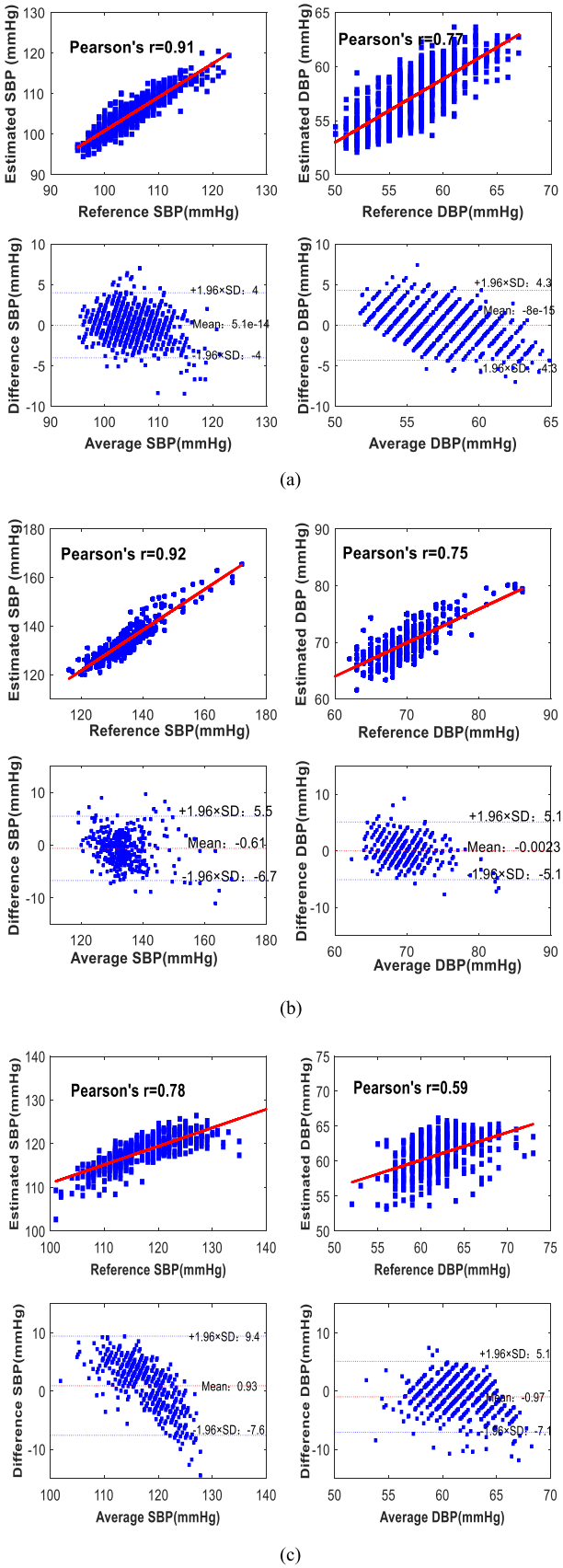

Fig. 6. Correlation and Bland–Altman plots of SBP and DBP with the reference Finapres BP using MLR. (a) Static estimation. (b) Dynamic estimation. (c) Long-term (6 months) estimation.

TABLE V
STATISTICAL ANALYSIS OF PERFORMANCE WITH SVR. (A) SHORT-TERM PERFORMANCE FOR 73 SUBJECTS. (B) DYNAMIC PERFORMANCE FOR 35 SUBJECTS. (C) LONG-TERM (6 MONTHS) PERFORMANCE FOR 10 SUBJECTS

(a)

	Index	Mean	Standard derivation	Max	Min
SBP (mmHg)	CC	0.852	0.127	0.952	0.574
	MD (mmHg)	-0.001	0.027	0.090	-0.001
	SD (mmHg)	3.102	0.992	7.861	1.493
DBP (mmHg)	CC	0.790	0.135	0.95	0.420
	MD (mmHg)	-0.004	0.017	0.043	-0.0001
	SD (mmHg)	2.199	0.770	1.315	5.897

(b)

	Index	Mean	Standard derivation	Max	Min
SBP (mmHg)	CC	0.946	0.034	0.972	0.819
	MD (mmHg)	-0.085	0.092	-0.455	-0.005
	SD (mmHg)	4.493	0.780	5.618	3.156
DBP (mmHg)	CC	0.929	0.082	0.973	0.599
	MD (mmHg)	-0.077	0.069	-0.349	-0.018
	SD (mmHg)	2.735	0.233	3.314	2.267

(c)

	Index	Mean	Standard derivation	Max	Min
SBP(mmHg)	CC	0.617	0.1523	0.848	0.3321
	MD (mmHg)	-1.148	4.566	7.962	0.3785
	SD (mmHg)	5.79	1.831	8.718	2.767
DBP (mmHg)	CC	0.554	0.231	0.817	0.192
	MD (mmHg)	-1.194	5.926	7.757	0.806
	SD (mmHg)	5.29	0.929	8.202	2.712

approximated the reference value for the static, dynamic, and long-term estimates.

C. Robustness Performance of the Proposed Models

We evaluated the robustness performance of the proposed approach by comparing the estimation accuracy of different physical states and at different time intervals; that is, the static and dynamic estimates and the estimations 1 day, 3 days, and 6 months after model construction. We used a two-sample t test to verify whether the estimation error means differed at different calibration intervals, with $p < 0.05$ indicating statistical significance. With the proposed MLR-based models, no significant difference was observed in the estimation error between the static and dynamic experiments. However, the overall estimation error for SBP and DBP for the 10 subjects increased significantly 1 day after model construction: from -0.001 ± 3.102 to 0.85 ± 5.78 mmHg for SBP and from -0.004 ± 2.199 to -1.24 ± 4.63 mmHg for DBP, respectively. From 1 day to 3 days and then to 6 months, the estimation error was relatively stable for both SBP and DBP. Fig. 8 shows the trend in the SBP and DBP estimation error with the proposed MLR- and SVR-based models compared with the PTT + PIR-based approach presented in [14] (based on our dataset). The figure indicates that the estimation errors increased significantly 1 day after model construction, but there was no significant change from 1 day to 6 months. For the MLR and SVR methods, the experimental results demonstrated that the SVR method slightly outperformed the MLR method, indicating nonlinear effects between the selected variables and BP value. However, considering the low-power, low-complexity requirements in continuous BP

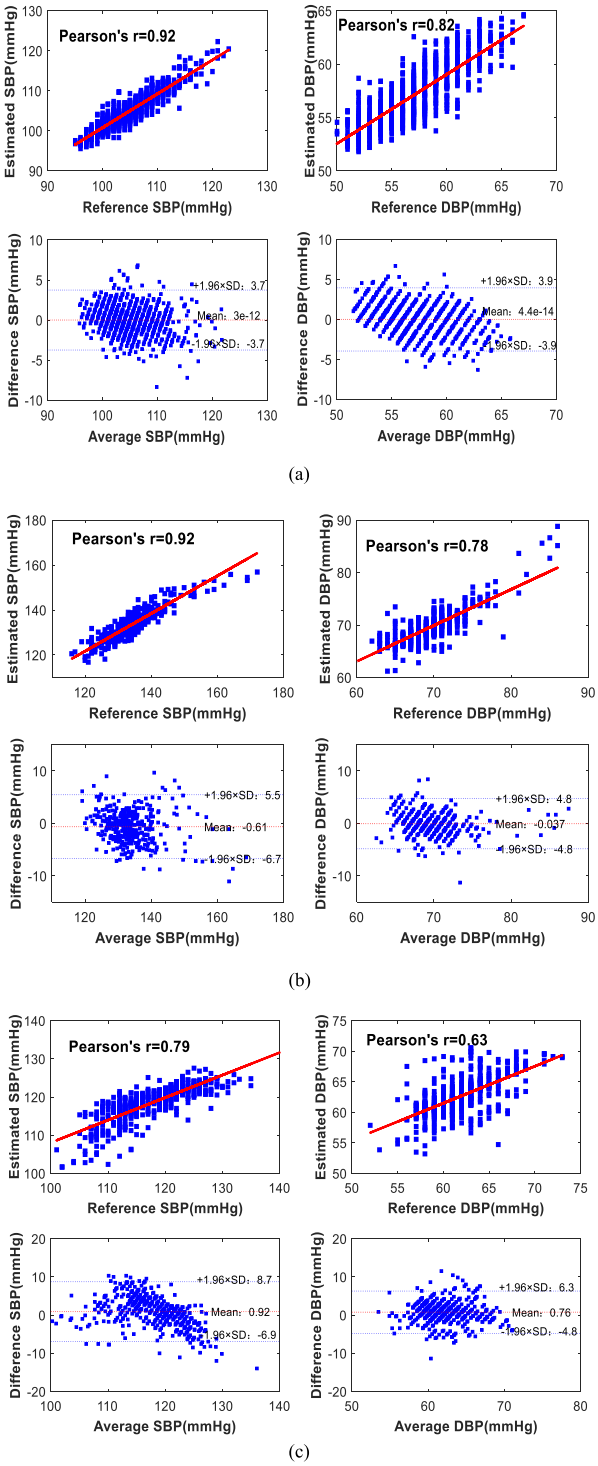

Fig. 7. Correlation and Bland–Altman plots of SBP and DBP with the reference Finapres BP using SVR. (a) Static estimation. (b) Dynamic estimation. (c) Long-term (6 months) estimation.

measurements and the similar estimation accuracy patterns between the two models, we recommend using MLR as the optimal approach.

To further verify the effectiveness of the proposed BP model with MLR, we also compared the performance with the PTT+PIR-based BP estimation approach presented in [14]

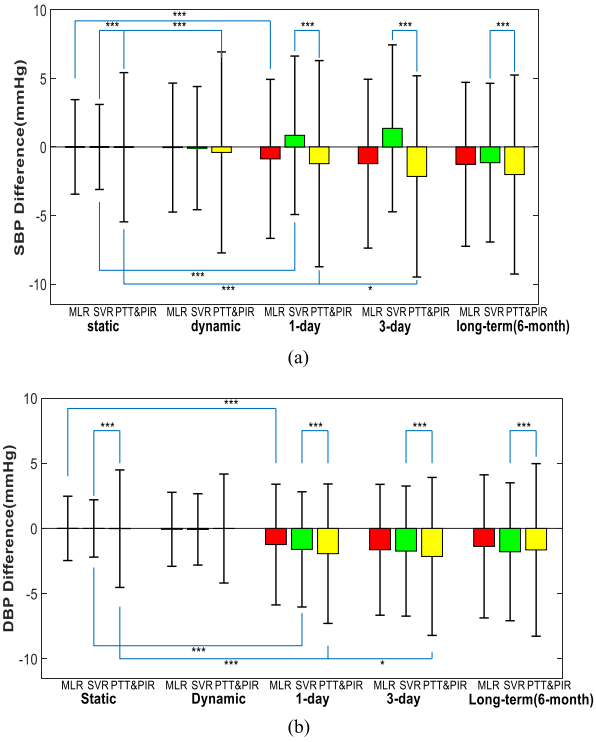

Fig. 8. Estimation error for SBP and DBP at different time intervals (∗ and ∗∗∗ indicate statistical significance at the 0.05 and 0.001 levels, respectively). (a) SBP estimation. (b) DBP estimation.

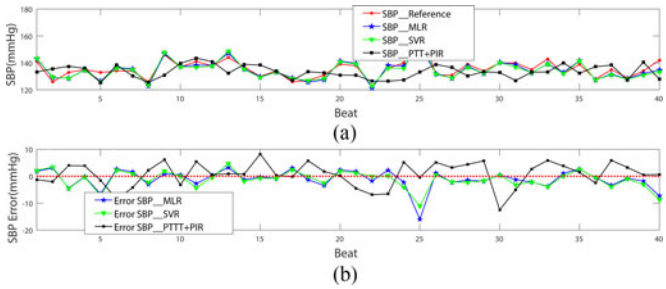

Fig. 9. Typical example of beat-to-beat SBP estimated with the proposed methods and the PTT+PIR method with the Finapres SBP as a reference. (a) Beat-to-beat SBP (b) SBP error.

(see Section II). Figs. 9 and 10 present typical examples of the comparison between the reference BP and the BP estimates derived from the different approaches. For both SBP and DBP, the BP estimates obtained from the proposed models tracked the Finapres BP values more accurately than did the PTT+PIR-based approach. The SBP estimation errors were -0.54 ± 3.39, -0.61 ± 3.11, and -0.23 ± 8.57 mmHg for the MLR-, SVR-, and PTT+PIR-based models, respectively. The corresponding DBP estimation errors were -0.002 ± 2.60, -0.03 ± 2.44, and -0.61 ± 5.51 mmHg. This evidence further verifies that the BP estimates obtained using the proposed approaches attained a closer correlation with the Finapres BP than did the PTT+PIR-based method.

Fig. 8 shows the trend in the estimation error with the PTT+PIR approach. Compared with this approach, the proposed MLR- and SVR-based methods obtained significantly

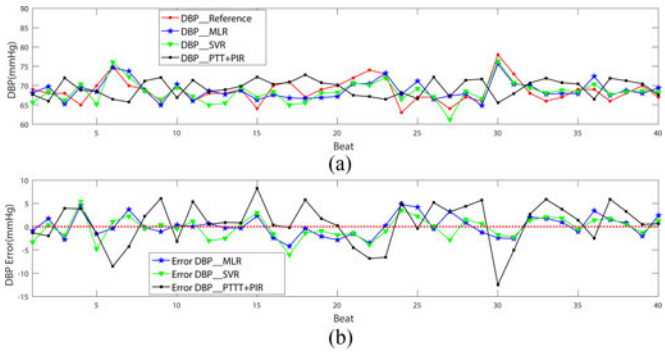

Fig. 10. Typical example of beat-to-beat DBP estimated with the proposed methods and the PTT+PIR method with the Finapres DBP as a reference. (a) Beat-to-beat DBP (b) DBP error.

higher estimation accuracy for both SBP and DBP at different time intervals. Specifically, the proposed approach outperformed the PTT+PIR-based model with an approximately 2-mmHg reduction in SD for the different time intervals. However, the proposed approach exhibited slightly higher robustness of performance compared with the PTT+PIR-based approach. For the proposed approach, the estimation error was stable from 1 day after the initial calibration, whereas for the PTT+PIR-based model, the estimation error continued to deteriorate from 1 day to 3 days.

IV. Discussion

In this study, we proposed a continuous BP estimation approach that combines data mining techniques with a mechanism model. This differs from traditional PTT-based models, in that personalized BP features can be identified with high predictive power. The experimental results demonstrate that the proposed MLR model with selected individual features improved the accuracy of mechanism-based BP approaches using PTT.

A. Features Selected for BP Estimation

Physiologically, several factors are responsible for BP changes, including peripheral resistance, vessel elasticity, cardiac output, and blood volume [34]. One of the main factors affecting BP is peripheral resistance, because more pressure is needed to maintain blood flow when the resistance increases. Apart from peripheral resistance, vessel elasticity also affects BP. A healthy elastic artery expands to absorb the peak systolic pressure and maintain blood flow during diastole. Anything that decreases the cardiac output also decreases the BP because of less pressure on the vessel walls. Blood volume affects BP in that more fluid presses against the arterial wall because of a greater volume of fluid resulting in greater pressure. In our study, the selected features for BP estimation are related to one or two of these factors.

Our study results confirm the findings of prior studies regarding PTT as a critical indicator of BP. It can be inferred from the M–K equation that PTT is related to the elastic modulus of the vessel wall and thus BP. However, the definitions of PTT differ between studies, especially regarding the definition of the foot location in PPG waveforms. The efficacy of these definitions has been compared in [35] and [36], but the overall results have been inconclusive. The most widely used definitions of the foot of the PPG pulse are the diastolic minimum time and the maximum derivatives time. Our study compared three definitions of PTT for estimating BP, namely PTT_ppgBottom, PTT_ppgPeak and PTT_MaxDeri (see Table I for definitions). Experimental results show that the most available indicators differ between subjects as a result of individual variability. Thus, the traditional SBP model based on a fixed definition of PTT exhibited low accuracy when applied to a large population. However, the statistical results for the feature importance in the static and dynamic experiments (Table III) indicate that PTT_MaxDeriwas the most useful indicator for SBP estimation among the three definitions, with a feature importance of 0.19.The feature importance was 0.07 for PTT_ppgBottom and 0.04 for PTT_ppgPeak.

PIR was recently proposed by Ding *et al.* [14] as a novel indicator for DBP estimation; thus, the accuracy of PTT-based models can be improved considerably by using the combination of PIR+PTT. That study showed that PIR can reflect changes in arterial diameter in one cardiac cycle, which is one of the main sources of peripheral resistance. Our study further supports evidence that PIR is the dominant feature for DBP estimation, as indicated by the feature importance value of 0.16. Furthermore, its impact on SBP estimation was also remarkable, with a feature importance value of 0.13.

Previous investigations have demonstrated that vascular tone and PEP exert a large effect on the PTT–BP relationship [13], [14], [37]–[40]. PTT with PEP included was used for BP estimation in our study. Many related studies have demonstrated that this combination has a positive effect on BP estimation. For example, a study conducted in 1981 [37] demonstrated that the correlation between PTT with PEP included and BP was larger than that obtained without PEP included. Furthermore, an investigation by Philips Research Laboratories (Europe) [38] suggested that PEP has a prominent role in SBP estimation. Additionally, a recent study by our group [39] showed that BP estimation was more accurate with PTT including PEP. A plausible explanation is that PEP, as an intra-cardiac component of PTT, can reflect the sympathetic cardiac influence, which would contribute to a change in BP. This is one of the most crucial reasons for why we included PEP in the calculation of PTT. A more recent clinical study [40] that used intra-arterial BP as a reference demonstrated that PTT (calculated from ECG and PPG, including PEP) correlated strongly with the gold standard of BP measurement, especially for SBP. Impairment of arterial vascular tone in hypertensive patients can be identified by the PPG waveform, as described previously in [41]–[44]. Takazawa *et al.* [41] reported that vascular function can be characterized by the 2nd PPG. A reflective index obtained from noninvasive PPG was also demonstrated to determine and quantify vascular tone in [45]. Therefore, a feature selection method was proposed in [27] for estimating SBP by using features extracted from the 1st and 2nd PPG. As demonstrated in previous studies, we found that multiple features extracted from the 2nd PPG, such as ppgSecondDeriWidth and ppgSecondDeriFootHeight, are crucial BP indicators with high feature importance values

of 0.10 and 0.11 for SBP estimation and 0.12 and 0.11 for DBP estimation, respectively. Additionally, the pulse characteristic value, K, was first studied to estimate cardiac output by using a three-element model [46], [47]. Furthermore, in [48], the researchers demonstrated that vascular elasticity can be quantified using K. Our study results confirm the importance of ppg_K in BP estimation with feature importance values of 0.14 for SBP and 0.11 for DBP. Therefore, the proposed model is significantly more accurate in combination with features extracted from the PPG signal.

B. Influence of Time Intervals on the Accuracy of the Proposed Approach

Although various studies have demonstrated the reliability of PTT-based methods for cuffless BP measurement [13], [49], [50], most of them have been validated on for very short time periods. The accuracy would deteriorate when applied to long-term estimates after the initial calibration [51]. In [18], a significant increase in the estimation error (from -1.44 ± 10.69 to -1.87 ± 14.42 mmHg for SBP and from -1.21 ± 5.07 to -0.11 ± 8.51 mmHg for DBP) was observed at 1 month after the initial calibration. Therefore, recalibration is the most challenging problem for improving estimation accuracy and must be resolved for this approach to be realized in widespread applications [14]. Frequent calibrations are necessary to ensure the accuracy of the PTT-based approach for BP estimation.

Except for static and dynamic validations within a short time of the initial calibration, our study also evaluated the proposed model in long-term BP tracking at different time intervals to analyze the robustness of the proposed models. Experimental results show that the proposed MLR-based model exhibited excellent performance in static and dynamic BP estimation, which was attributed to the consideration of multiple factors that influence BP changes. However, similar to previous studies on long-term BP validation, the accuracy of the proposed model was lower to a certain degree when used in long-term tracking. Specifically, the estimation error increased from -0.001 ± 3.102 to 0.85 ± 5.78 mmHg for SBP and from -0.004 ± 2.199 to -1.24 ± 4.63 mmHg for DBP at 1 day after the initial calibration. However, with longer time intervals (i.e., from 1 day to 3 days and then to 6 months), the estimation error became relatively stable. This finding supports the results reported in [17] that for healthy subjects, the estimation accuracy decreased significantly 2 weeks after the initial calibration, but no significant change was observed from 2 weeks to 1 month. There are three reasons for this. First, the dataset for training the BP model was collected with subjects in a seated position with lower BP variability; consequently, the robustness of the data-driven model is low, especially for BP estimations derived from the test dataset, which varied considerably from the training set. Second, a previous study stated that the gold standard of the Finapres System is prone to substantial variability [52], which would introduce instability into the model. Finally, the contribution of the dominant features identified from the short-term dataset diminished over time when applied in long-term tracking.

V. Conclusions and Future Work

In this study, we proposed a novel BP estimation approach by combining data mining techniques with a mechanism-driven model. Through the combination of multiple factors that can reflect changes in BP, a feature selection method based on a genetic algorithm was employed to obtain the most promising indicators for each subject. MLR and SVR were employed to develop the BP model. We validated the effectiveness of the proposed approach according to the estimation accuracy and robustness of the model. The experimental results show that compared with the PTT+PIR-based model, the proposed approach achieves significantly higher estimation accuracy for different states and time intervals and slightly improved robustness performance. Overall, the proposed model provides potentially novel insights for unobtrusive BP estimation.

In the future, we will endeavor to provide a continuous BP model that conforms to the IEEE 1708 standard, which was proposed specifically for wearable and cuffless BP devices [53]. First, some effective solutions to overcome the low accuracy of long-term estimations should be considered. For example, the training dataset should comprise data that spans different time lengths and includes different physical states; this would enable the data mining algorithm to identify domain features in long-term tracking. Second, an invasive method for continuous BP measurement should be referred to as the gold standard. Finally, the accuracy should be verified in a large, heterogeneous population—especially for hypertensive patients and long-term validation—to evaluate the applicability of the proposed approach.

Acknowledgment

The authors thank Dr. Zhiqiang Zhang from the University of Leeds for the suggestions on improving this manuscript.

References

[1] A. Vchobanian, G. Lbakris, and H. Rblack, "The seventh report of the joint national committee on prevention, detection, evaluation, and treatment of high blood pressure: The JNC 7 Report," *JAMA*, vol. 289, no. 19, pp. 2560–2572, 2003.

[2] N. Zakopoulos, G. Tsivgoulis, and G. Barlas, "Time rate of blood pressure variation is associated with increased common carotid artery intima-media thickness," *Hypertension*, vol. 45, no. 4, pp. 505–512, 2005.

[3] P. Palatini, G. Reboldi, and L. Jbeilin, "Added predictive value of nighttime blood pressure variability for cardiovascular events and mortality the ambulatory blood pressure international study, " *Hypertension*, vol. 64, no. 3, pp. 487–493, 2014.

[4] Y. T. Zhang, Y. L. Zheng, W. H. Lin, H. Y. Zhang and X. L. Zhou, "Challenges and opportunities in cardiovascular health informatics," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 3, pp. 633–642, Mar. 2013.

[5] J. Rmartina, B. Ewesterhof, and J. V. Goudoever, "Noninvasive continuous arterial blood pressure monitoring with Nexfin," *Anesthesiology*, vol. 116, no. 5, pp. 1092–1103, 2012.

[6] K. Hwesseling, "Finger arterial pressure measurement with Finapres," *Zeitschrift Fur Kardiologie*, vol. 85, Suppl 3, pp. 38–44, 1996.

[7] X. R. Ding *et al.*, "Continuous blood pressure measurement from invasive to unobtrusive: Celebration of 200th birth anniversary of Carl Ludwig," *IEEE J. Biomed. Health Informat.*, vol. 20, no. 6, pp. 1455–1465, Nov. 2016.

[8] K. W. Chan, K. Hung, and Y. T. Zhang, "Noninvasive and cuffless measurements of blood pressure for telemedicine," in *Proc. 23rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2001, vol. 4, pp. 3592–3593.
[9] W. B. Gu, C. C. Y. Poon, M. Y. Sy, H. K. Leung, Y. P. Liang, and Y. T. Zhang, "A h-shirt-based body sensor network for cuffless calibration and estimation of arterial blood pressure," in *Proc. 2009 6th Int. Workshop Wearable Implantable Body Sensor Netw.*, Berkeley, CA, USA, 2009, pp. 151–155.
[10] C. C. Y. Poon and Y. T. Zhang, "Cuff-less and noninvasive measurements of arterial blood pressure by pulse transit time," in *Proc. 2005 IEEE Eng. Med. Biol. 27th Annu. Conf.*, Shanghai, China, 2005, pp. 5877–5880.
[11] Y. Chen, C. Wen, G. Tao and M. Bi, "A new methodology of continuous and noninvasive blood pressure measurement by pulse wave velocity," in *Proc. 2010 11th Int. Conf. Control Autom. Robot. Vis.*, Singapore, 2010, pp. 1018–1023.
[12] Q. Liu, B. P. Yan, C. M. Yu, Y. T. Zhang and C. C. Y. Poon, "Attenuation of systolic blood pressure and pulse transit time hysteresis during exercise and recovery in cardiovascular patients," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 2, pp. 346–352, Feb. 2014.
[13] Y. L. Zheng, B. P. Yan, Y. T. Zhang, and C. C. Y. Poon, "An armband wearable device for overnight and cuff-less blood pressure measurement," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 7, pp. 2179–2186, Jul. 2014.
[14] X. R. Ding, Y. T. Zhang, J. Liu, W. X. Dai and H. K. Tsang, "Continuous cuffless blood pressure estimation using pulse transit time and photoplethysmogram intensity ratio," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 5, pp. 964–972, May 2016.
[15] R. A. Payne *et al.*, "Pulse transit time measured from the ECG: an unreliable marker of beat-to-beat blood pressure," *J. Appl. Physiol.*, vol. 100, no. 1, pp. 136–141, 2000.
[16] M. Y. Wong *et al.*, "The effects of pre-ejection period on post-exercise systolic blood pressure estimation using the pulse arrival time technique," *Eur. J. Appl. Physiol.*, vol. 111, no. 1, pp. 135–144, 2011.
[17] X. R. Ding, Y. T. Zhang, and Y. H. K. Tsang, "Impact of heart disease and calibration interval on accuracy of pulse transit time–based blood pressure estimation," *Physio. Meas.*, vol. 37, no. 2, pp. 227–237, 2016.
[18] C. Christopher, "Encyclopædia britannica: Definition of data mining." 2010. [Online]. Available: https://www.britannica.com/technology/data-mining
[19] S. H. Liao, P. H. Chu, and P. Y. Hsiao, "Data mining techniques and applications – A decade review from 2000 to 2011," *Expert Syst. Appl.*, vol. 39, no. 12, pp. 11303–11311, 2012.
[20] P. R. Hachesu *et al.*, "Use of data mining techniques to determine and predict length of stay of cardiac patients," *Healthcare Informat. Res.*, vol. 19, no. 2, pp. 121–129, 2013.
[21] P. Wasiewicz, "Data mining analysis of factors influencing children's blood pressure in a nation-wide health survey," *Proc. SPIE - Int. Soc. Opt. Eng.*, vol. 7502, pp. 75022R-1–75022R-8, 2009.
[22] K. Takazawa, N. Tanaka, and M. Fujita, "Assessment of vasoactive agents and vascular aging by the second derivative of photoplethysmogram waveform," *Hypertension*, vol. 32, no. 2, pp. 365–370, 1998.
[23] X. F. Teng and Y. T. Zhang, "Continuous and noninvasive estimation of arterial blood pressure using a photoplethysmographic approach," in *Proc. 2003 25th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2003, vol. 4, pp. 3153–3156.
[24] S. Suzuki and K. Oguri, "Cuffless and non-invasive systolic blood pressure estimation for aged class by using a photoplethysmograph," in *Proc. 2008 30th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Vancouver, BC, Canada, 2008, pp. 1327–1330.
[25] J. Y. Kim, B. H. Cho, S. M. Im, M. J. Jeon, I. Y. Kim, and S. I. Kim, "Comparative study on artificial neural network with multiple regressions for continuous estimation of blood pressure," in *Proc. 2005 IEEE Eng. Med. Biol. 27th Annu. Conf.*, Shanghai, China, 2005, pp. 6942–6945.
[26] A. Suzuki and K. Ryu, "Feature selection method for estimating systolic blood pressure using the Taguchi method," *IEEE Trans. Ind. Inf.*, vol. 10, no. 2, pp. 1077–1085, May 2014.
[27] R. He, Z. P. Huang, L. Y. Ji, J. K. Wu, H. Li and Z. Q. Zhang, "Beat-to-beat ambulatory blood pressure estimation based on random forest," in *Proc. 2016 IEEE 13th Int. Conf. Wearable Implantable Body Sensor Netw.*, San Francisco, CA, USA, 2016, pp. 194–198.
[28] D. J. Hughes *et al.*, "Measurements of young's modulus of elasticity of the canine aorta with ultrasound," *Ultrason. Imag.*, vol. 1, no. 1, pp. 356–367, 1979.
[29] N. Westerhof *et al.*, "The arterial windkessel," *Med. Biol. Eng. Comput.*, vol. 47, pp. 131–141, 2009.
[30] J. Yang and V. Honavar, "Feature subset selection using a genetic algorithm," *IEEE Intell. Syst. Appl.*, vol. 13, no. 2, pp. 44–49, Mar./Apr. 1998.
[31] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning.* Reading, MA, USA: Addison-Wesley, 1989.
[32] G. Isabelle and A. Elisseeff, "An introduction to variable and feature selection." *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, 2003.
[33] H. Drucker *et al.*, "Support vector regression machines," *Adv. Neural Inf. Process. Syst.*, vol. 28, no. 7, pp. 779–784, 1996.
[34] "Factors that affect blood pressure." 2000. [Online]. Available: http://www.edises.it/file/minicd/germ002/misc/assignmentfiles/cardiovascular/Fact_Aff_Blood_Pressure.pdf
[35] D. Špulák, R. Čmejla and V. Fabián, "Parameters for mean blood pressure estimation based on electrocardiography and photoplethysmography," in *Proc. Int. Conf. Appl. Electron.*, 2011, pp. 1–4.
[36] I. J. Wood and J. Finkelstein, "Using individualized pulse transit time calibration to monitor blood pressure during exercise," *Studies Health Technol. Informat.*, vol. 190, no. 190, pp. 39–41, 2013.
[37] D. B. Newlin, "Relationships of pulse transmission times to pre-ejection period and blood-pressure," *Psychophysiology*, vol. 18, no. 3, pp. 316–321, 1981.
[38] J. Muehlsteff, X. L. Aubert, and M. Schuett, "Cuffless estimation of systolic blood pressure for short effort bicycle tests: The prominent role of the pre-ejection period," in *Proc. 28th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, vol. 1–15, 2006, pp. 3460–3464.
[39] M. Y. M. Wong, E. Pickwell-Macpherson, Y. T. Zhang,and J. C. Y. Cheng, "The effects of pre-ejection period on post-exercise systolic blood pressure estimation using the pulse arrival time technique," *Eur. J. Appl. Physiol.*, vol. 111, no. 1, pp. 135–144, 2011.
[40] Q. Liu, B. P. Yan, C. M. Yu, Y. T. Zhang, and C. C. Y. Poon, "Attenuation of systolic blood pressure and pulse transit time hysteresis during exercise and recovery in cardiovascular patients," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 2, pp. 346–352, Feb. 2014.
[41] K. Takazawa *et al.*, "Assessment of vasoactive agents and vascular aging by the second derivative of photoplethysmogram waveform," *Hypertension*, vol. 32, no. 2, pp. 365–370, 1998.
[42] P. J. Chowienczyk *et al.*, "Photoplethysmographic assessment of pulse wave reflection: Blunted response to endothelium-dependent beta2-adrenergic vasodilation in type II diabetes mellitus," *J. Amer. CollCardiol*, vol. 34, pp. 2007–2014, 1999.
[43] S. C. Millasseau *et al.*, "Noninvasive assessment of the digital volume pulse. Comparison with the peripheral pressure pulse," *Hypertension*, vol. 36, no. 6, 2000, Art. no. 9520956.
[44] S. C. Millasseau *et al.*, "Determination of age-related increases in large artery stiffness by digital pulse contour analysis," *Clin. Sci.*, vol. 103, no. 4, pp. 371–377, 2002.
[45] S. C. Millasseau *et al.*, "The vascular impact of aging and vasoactive drugs: Comparison of two digital volume pulse measurements," *Amer. J. Hypertension*, vol. 16, no. 6, pp. 467–471, 2003.
[46] K. H. Wesseling, W. B. De, J. Web, and N. T. Smith, "A simple device for the continuous measurement of cardiac output," *Adv. Cardiovascular Phys.*, vol. 5, pp. 16–52, 1983.
[47] K. H. Wesseling *et al.*, "Computation of aortic flow from pressure in humans using a nonlinear, three-element model," *J. Appl. Physiol.*, vol. 74, no. 5, pp. 2566–2573, 1993.
[48] P. Uangpairoj and M. Shibata, "Evaluation of vascular wall elasticity of human digital arteries using alternating current-signal photoplethysmography," *Vascular Health Risk Manage.*, vol. 9, no. 9, pp. 283–295, 2013.
[49] H. Gesche *et al.*, "Continuous blood pressure measurement by using the pulse transit time: Comparison to a cuff-based method," *Eur. J. Appl. Physiol.*, vol. 112, no. 1, pp. 309–315, 2011.
[50] M. Forouzanfar, S. Ahmad, I. Batkin, H. R. Dajani, V. Z. Groza and M. Bolic, "Coefficient-free blood pressure estimation based on pulse transit time–cuff pressure dependence," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 7, pp. 1814–1824, Jul. 2013.
[51] B. M. Mccarthy *et al.*, "An examination of calibration intervals required for accurately tracking blood pressure using pulse transit time algorithms, " *J. Human Hypertension*, vol. 27, no. 12, pp. 744–750, 2013.
[52] B. Silke and D. Mcauley. "Accuracy and precision of blood pressure determination with the Finapres: An overview using re-sampling statistics," *J. Human Hypertension*, vol. 12, no. 6, pp. 403–409, 1998.
[53] *IEEE Standard for Wearable Cuffless Blood Pressure Measuring Devices*, IEEE Standard 1708-2014, pp. 1–38, Aug. 26, 2014.

Authors', photographs and biographies not available at the time of publication.