Estimation and Validation of Arterial Blood Pressure Using Photoplethysmogram Morphology Features in Conjunction With Pulse Arrival Time in Large Open Databases

Seungman Yang, Jangjay Sohn, *Student Member, IEEE*, Saram Lee, Joonnyong Lee, and Hee Chan Kim, *Member, IEEE*

***Abstract*—Although various predictors and methods for BP estimation have been proposed, differences in study designs have led to difficulties in determining the optimal method. This study presents analyses of BP estimation methods using 2.4 million cardiac cycles of two commonly used non-invasive biosignals, electrocardiogram (ECG) and photoplethysmogram (PPG), from 1376 surgical patients. Feature selection methods were used to determine the best subset of predictors from a total of 42 including PAT, heart rate (HR), and various PPG morphology features, and BP estimation models constructed using linear regression (LR), random forest (RF), artificial neural network (ANN), and recurrent neural network (RNN) were evaluated. 28 features out of 42 were determined as suitable for BP estimation, in particular two PPG morphology features outperformed PAT, which has been conventionally seen as the best non-invasive indicator of BP. By modelling the low frequency component of BP using ANN and the high frequency component using RNN with the selected predictors, mean errors of 0.05 $\pm$ 6.92 mmHg for systolic BP, and -0.05 $\pm$ 3.99 mmHg for diastolic BP were achieved. External validation of the model using another biosignal database consisting of 334 intensive care unit patients led to similar results, satisfying three standards for accuracy of BP monitors. The results indicate that the proposed method can contribute to the realization of ubiquitous non-invasive continuous BP monitoring.**

Manuscript received March 8, 2020; revised May 28, 2020 and July 13, 2020; accepted July 13, 2020. Date of publication July 16, 2020; date of current version April 5, 2021. This work was supported by the Bio and Medical Technology Development Program of the National Research Foundation (NRF) funded by the Korean Government (MSIT) under Grant 2016M3A9F1939646 *(Corresponding authors: Joonnyong Lee; Hee Chan Kim.)*

Seungman Yang and Jangjay Sohn are with the Interdisciplinary Program in Bioengineering, Seoul National University Graduate School, Seoul 08826, South Korea (e-mail: ysmgreen@melab.snu.ac.kr; jjaysohn@melab.snu.ac.kr).

Saram Lee is with the Seoul National University Hospital Biomedical Research Institute, Seoul 03080, South Korea (e-mail: hommelee@snu.ac.kr).

Joonnyong Lee is with the Mellowing Factory Co. Ltd., Seoul 03726, South Korea (e-mail: joonnyonglee@melab.snu.ac.kr).

Hee Chan Kim is with the Department of Biomedical Engineering, Seoul National University College of Medicine, Seoul 03080, South Korea, and also with the Institute of Medical and Biological Engineering, Medical Research Center, Seoul National University, Seoul 08826, South Korea (e-mail: hckim@snu.ac.kr).

Digital Object Identifier 10.1109/JBHI.2020.3009658

***Index Terms*—Continuous blood pressure monitoring, biosignal database, photoplethysmogram morphology, feature selection, pulse arrival time.**

I. Introduction

CARDIOVASCULAR diseases (CVD) are the leading cause of worldwide mortality [1], and since hypertension is a major predictor of CVD, proper monitoring and management of blood pressure (BP) is important [2]–[4]. However, a discrete BP measurement, generally performed in clinical settings with a conventional cuff-type oscillometric device, can lead to misdiagnosis and cannot observe circadian fluctuations of BP [4], [5]. As the importance of ambulatory blood pressure has been stressed in many recent studies on the topic [5], [6], it can be deduced that while continuous monitoring is required in daily life for accurate diagnosis of BP and cardiovascular health, the conventional cuff-based method is not practical due to its inconvenient and cumbersome nature.

As one of the promising techniques for realizing NCBPM, pulse wave velocity (PWV) based approaches have gained the most interest over the recent decades [7]–[14]. This technique based on the concept that the velocity of the arterial pulse traveling in an artery is affected by physiological variations of the arterial vessels, which is closely related to BP [15]. PWV can be approximated using its surrogate, pulse transit time (PTT), which is the time delay for the arterial pulse to propagate between two arterial sites. Although numerous studies have tried to establish a BP estimation method using PTT, most studies have used pulse arrival time (PAT) instead, because of the ease of measurement [8], [16]–[20]. PAT is defined as the time difference between the R-peak of the electrocardiogram (ECG) and a point on distal pulse waveform such as the photoplethysmogram (PPG) measured at the fingertips. Although some studies have shown remarkable results in NCBPM using PAT and/or PTT methods [8], [10], [14], [21], [22], the results are still widely varied in terms of accuracy [19], [23], [24].

In order to find alternative predictors associated with BP to improve the estimation accuracy, many researchers have attempted

to analyze the morphology of the PPG waveform [25]–[31], since it theoretically could reflect both the ejection of blood pulse from the heart and the conditions of the peripheral artery [32]. However, most of these studies have analyzed different characteristics extracted from the morphology of the PPG waveform in less than 100 subjects. While a few studies have presented results from large pools of subjects by using an open biosignal database called the Physionet Medical Information Mart for Intensive Care (MIMIC) [33]–[38], the database had issues with inter-waveform alignments due to unspecified filtering delays or channel delays, which made the data unsuitable for traditional inter-waveform analysis [39]. In addition, there were no studies that clearly confirmed the relations between BP and various PPG morphology features from a large group of heterogeneous subjects.

In a previous research [40], a moderate degree of association between BP and PAT from a large database called the VitalDB [41] was found, but a linear model based BP estimation using PAT as a sole predictor with a simple calibration did not show satisfactory level of performance in systolic blood pressure (SBP). In order to improve the performance of the BP estimation model, in this study, various PPG morphology based features were analyzed from the same surgical biosignal database to determine the valid markers of BP, and BP estimation models using these features with PAT were constructed and evaluated. Furthermore, the models were validated externally in MIMIC database. First, the VitalDB database and the data exclusion process for the selection of usable data are introduced and the detailed definitions of various features are described. Then, the method of constructing BP estimation models is presented. Lastly, the results of analyses are presented and discussed.

II. Methods

A. Data Acquisition & Data Pre-Processing

Data loading, pre-processing and feature extraction were performed automatically using a MATLAB script (MATLAB 2018b; Mathworks, Natick, MA, USA). The data used in this study were retrieved from the VitalDB, an open multi-biosignal database provided by the Seoul National University Hospital Department of Anesthesia (Seoul, Korea) [41]. The VitalDB contains time-synchronized biosignals during various surgeries including arterial blood pressure (ABP), ECG, and PPG from a total of 6388 patients. The experimental setup of the VitalDB is described in detail in two previous studies [40], [41]. The data used in this study were selected and pre-processed according to data selection criteria as shown in Fig. 1. Since the available recordings contained segments in which the biosignals were severely distorted due to saturation and/or unknown causes generated during the surgery, several exclusion criteria were applied to remove these corrupted sections prior to analyses. The example plots of data exclusion process and data pre-processing are shown in Fig. 2.

First, the recordings were checked for ECG, PPG, and ABP signals. Second, recordings less than 30 minutes were removed.

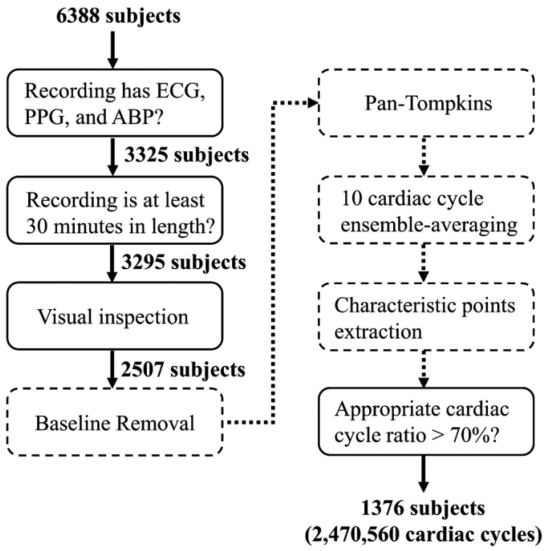

Fig. 1. Outline of the data exclusion process for the selection of usable data from the VitalDB and data pre-processing. Solid lines correspond to the data selection criteria, and dashed lines correspond to the data pre-processing.

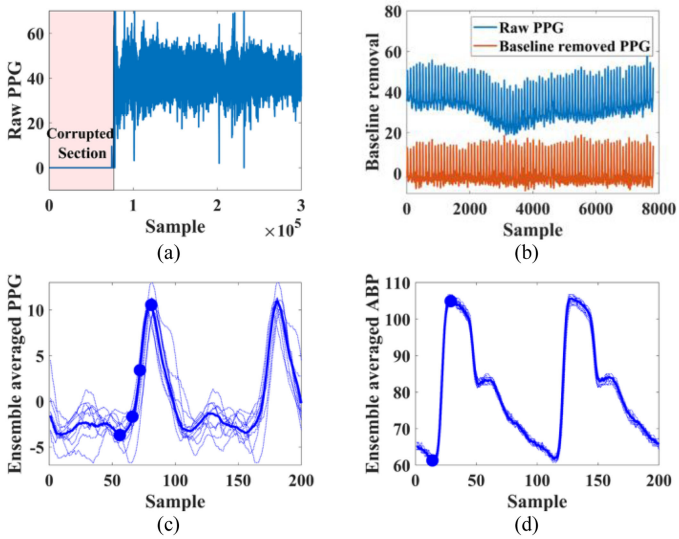

Fig. 2. Example plots of data exclusion process and data pre-processing. (a) Example plot for visual inspection of PPG waveform with the corrupted section highlighted; (b) Example plot for baseline removal of PPG waveform; (c) Example plot of ensemble averaged PPG; (d) Example plot of ensemble averaged ABP. In (c) and (d), dotted lines denote to 10 adjacent cardiac cycles of PPG/ABP, and solid line denotes to the ensemble averaged waveform with the characteristic points of PPG/ABP labeled.

Third, the recordings were visually inspected to manually remove sections where the biosignals were saturated as shown in Fig. 2(a).

Following data selection, biosignals were pre-processed to detect the characteristic points of ABP and PPG for each cardiac cycle. First, the low frequency baseline fluctuations of ECG and PPG waveforms were removed using non-linear filtering as

shown in Fig. 2(b) [42]. Then, the ECG R-peaks detected by Pan-Tompkins algorithm [43] were used to separate the waveform data into cardiac cycles for ensemble-averaging; 10 adjacent cardiac cycles of ABP and PPG were ensemble-averaged in a smoothing window manner to accentuate the waveform features as shown in Fig. 2(c) and Fig. 2(d). After ensemble-averaging, the characteristic points of ABP and PPG were detected. For ABP, the peak and valley points were detected. In the case of PPG, the valleys, peaks, maximum derivatives, and intersecting-tangent points (or the intersecting point between the tangent lines of the maximum derivative and the diastolic minimum) [44] were detected. SBP and diastolic blood pressure (DBP) were derived from the peak and the valley points of ABP, and four different PAT (PAT_V, PAT_P, PAT_{MD}, and PAT_{IT}) values were derived from the time difference between ECG R-peak and one of the characteristic points of PPG ('V' denotes the valley, 'P' denotes the peak, 'MD' denotes the maximum derivative, and 'IT' denotes the intersecting-tangent points of PPG). Since the waveforms of ABP and PPG could be severely distorted by various artifacts, signal quality of each ensemble-averaged cardiac cycle was evaluated using the following conditions:

- Is the extracted SBP greater than 50 mmHg and less than 250 mmHg?
- Is the extracted DBP greater than 30 mmHg and less than 160 mmHg?
- Is the change in the extracted BP (SBP or DBP) during the previous 5 s interval less than 30 mmHg?
- Is the change in the extracted PATs during the previous 5 s interval less than 30ms?

If the above conditions were satisfied, the cardiac cycle was evaluated as having high signal quality, and was determined to be appropriate for further analyses. If the ratio of the number of appropriate cardiac cycles to the total number of cardiac cycles was less than 70%, the recording was excluded as shown in Fig. 1. Finally, 30 minutes sections in each recording were selected to adjust the number of data points per subject to a similar level for proper validation. As a result, 1376 recordings (2,470,560 cardiac cycles) were selected. The demographic characteristics of the patients and the BP characteristics of the selected data are shown in Table I.

TABLE I
DEMOGRAPHIC AND BP CHARACTERISTICS OF THE DATA (N = 1376)

Characteristics	Subjects
Age (years)	58 ± 14 (range 8–92)
Gender (male)	720 (52%)
Height (cm)	162 ± 9
Weight (kg)	61 ± 11
BMI (kg/m^2) [a]	23 ± 3
Hypertension	484 (35%)
Diabetes	159 (12%)
# of cardiac cycles	1788 ± 289
SBP	
Mean value (mmHg)	115 ± 15
Δ value [b] (mmHg)	37 ± 19
DBP	
Mean value (mmHg)	62 ± 10
Δ value [b] (mmHg)	21 ± 11

[a]Body mass index.
[b]The difference between max and min values of each recording.

B. Feature Extraction

A total of 42 features including four different PAT (PAT_V, PAT_P, PAT_{MD}, and PAT_{IT}) values, HR, and various PPG morphology features used in previous studies were extracted from each cardiac cycle as shown in Fig. 3. The definitions of the features are listed in Table II. The detailed description of PPG morphology features is as followed:

1) Photoplethysmogram Intensity Ratio (PIR): Photoplethysmogram intensity ratio (PIR), defined as the ratio of the intensity of the peak point of PPG to the intensity of the valley point of PPG, is used as BP estimation predictor by Ding *et al.* [45]. According to the Beer-Lambert law, Photoplethysmogram intensity of the peak and valley point of PPG (PI_P and PI_V) can be derived as below.

$$PI_P = PI_0 \cdot e^{-\varepsilon_{DC} \cdot c_{DC} \cdot d_{DC}} \cdot e^{-\alpha \cdot d_D} \quad (1)$$

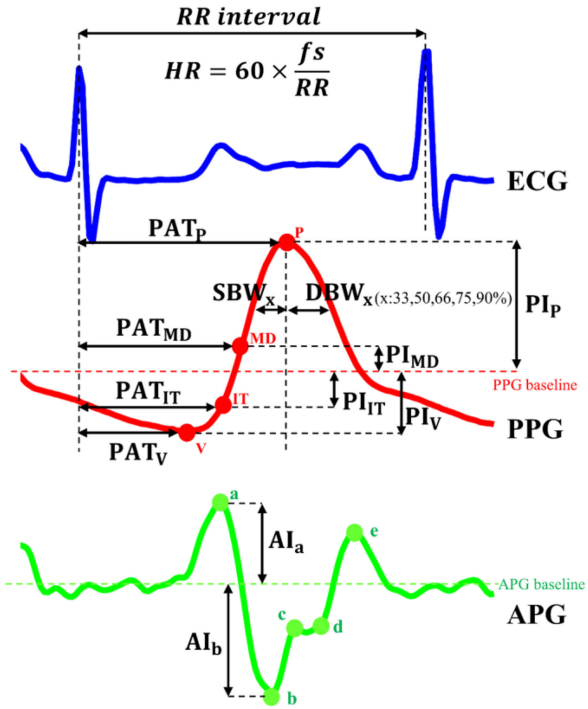

Fig. 3. Features labeled on ECG, PPG, and APG waveforms.

$$PI_V = PI_0 \cdot e^{-\varepsilon_{DC} \cdot c_{DC} \cdot d_{DC}} \cdot e^{-\alpha \cdot d_S} \quad (2)$$

where PI_0 is the incident light intensity, ε_{DC} is the absorbance coefficient, c_{DC} is the concentration, d_{DC} is the optical path of the DC component, α is the characteristic parameter, d_S and d_D are the systolic diameter and diastolic diameter, respectively. Therefore, as taking the ratio of (1) and (2), PIR is exponentially

TABLE II
DEFINITIONS OF THE ANALYZED FEATURES

Index	Feature	Definition
1	PAT_V	Time delay from the R-peak of ECG to the valley point of PPG
2	PAT_P	Time delay from the R-peak of ECG to the peak point of PPG
3	PAT_{MD}	Time delay from the R-peak of ECG to the 'MD' point of PPG
4	PAT_{IT}	Time delay from the R-peak of ECG to the 'IT' point of PPG
5	HR	Time delay between two adjacent R-peaks of ECG
6	PI_V	Intensity of the valley point of PPG
7	PI_P	Intensity of the peak point of PPG
8	PI_{MD}	Intensity of the 'MD' point of PPG
9	PI_{IT}	Intensity of the 'IT' point of PPG
10	PIR_P	Ratio of PPG peak point intensity to PPG valley point intensity
11	PIR_{MD}	Ratio of PPG 'MD' point intensity to PPG valley point intensity
12	PIR_{IT}	Ratio of PPG 'IT' point intensity to PPG valley point intensity
13	CT	Time delay from the valley point of PPG to the peak point of PPG
14	AS	Ascending slope
15	PPGK	PPG characteristic value
16	SBW_{33}	Systolic branch width at x % of the pulse height of PPG (x: 33, 50, 66, 75, 90)
17	SBW_{50}	
18	SBW_{66}	
19	SBW_{75}	
20	SBW_{90}	
21	DBW_{33}	Diastolic branch width at x % of the pulse height of PPG (x: 33, 50, 66, 75, 90)
22	DBW_{50}	
23	DBW_{66}	
24	DBW_{75}	
25	DBW_{90}	
26	BW_{33}	Branch width at x % of the pulse height of PPG (x: 33, 50, 66, 75, 90)
27	BW_{50}	
28	BW_{66}	
29	BW_{75}	
30	BW_{90}	
31	BWR_{33}	Branch width ratio at x % of the pulse height of PPG (x: 33, 50, 66, 75, 90)
32	BWR_{50}	
33	BWR_{66}	
34	BWR_{75}	
35	BWR_{90}	
36	AI_a	Intensity of the a-wave of APG
37	AI_b	Intensity of the b-wave of APG
38	AIR	Ratio of APG b-wave intensity to APG a-wave intensity
39	PC1	1st principal component of the one cardiac cycle PPG
40	PC2	2nd principal component of the one cardiac cycle PPG
41	PC3	3rd principal component of the one cardiac cycle PPG
42	PC4	4th principal component of the one cardiac cycle PPG

linked with arterial diameter change through the following equation:

$$\mathrm{PIR_P} = \frac{\mathrm{PI_P}}{\mathrm{PI_V}} = e^{\alpha \cdot (d_S - d_D)} \tag{3}$$

Since the arterial diameter change affects total peripheral resistance, a direct influencer of BP [15], $\mathrm{PIR_P}$ can theoretically be used for BP estimation. In this paper, $\mathrm{PIR_{MD}}$ and $\mathrm{PIR_{IT}}$ are additionally defined using PPG intensity of the maximum derivative or intersecting-tangent points of PPG ($\mathrm{PI_{MD}}$ or $\mathrm{PI_{IT}}$) instead of $\mathrm{PI_P}$ in (3), respectively. As a result, three different PIR ($\mathrm{PIR_P}$, $\mathrm{PIR_{MD}}$, and $\mathrm{PIR_{IT}}$) values and four different PI ($\mathrm{PI_V}$, $\mathrm{PI_P}$, $\mathrm{PI_{MD}}$, and $\mathrm{PI_{IT}}$) values are extracted.

2) Crest Time (CT) and Ascending Slope (AS): Crest time (CT), defined as the time delay from the valley point to the peak point of the PPG waveform is known to be a useful feature for cardiovascular disease classification [46]. Ascending slope (AS) was previously reported as the rate of wave front sloping for BP estimation [47]. In this study, CT and AS was calculated as below.

$$\mathrm{CT} = \mathrm{PAT_P} - \mathrm{PAT_V} \tag{4}$$

$$\mathrm{AS} = \frac{\mathrm{PI_P} - \mathrm{PI_V}}{\mathrm{PAT_P} - \mathrm{PAT_V}} \tag{5}$$

3) PPGK: PPGK, also called PPG characteristic value or K value, was previously reported to be related to total peripheral resistance and blood viscosity [48]. PPGK is based on the change of PPG area and defined as below.

$$\mathrm{PPGK} = \frac{\mathrm{P_m} - \mathrm{PI_V}}{\mathrm{PI_P} - \mathrm{PI_V}} \tag{6}$$

where $\mathrm{P_m}$ is the average value of one cardiac cycle of PPG waveform.

4) Branch Width (BW) Based Features: Awad *et al.* previously reported that pulse width at half height of a PPG waveform is related to total peripheral resistance [49]. Following this study, in order to extract as much information as possible, Kurylyak *et al.* calculated the widths at 25%, 33% and 75% of the pulse height and extracted separate values for the systolic part and for the diastolic part, and for the ratio between these values [34]. In this paper, systolic branch width (SBW) is defined as the systolic part of the pulse width, diastolic branch width (DBW) is defined as the diastolic part of the pulse width as shown in Fig. 3. Branch width (BW) is defined as the sum of SBW and DBW, and branch width ratio (BWR) is defined as the ratio of DBW to SBW. BW, SBW, DBW, and BWR are calculated at 33%, 50%, 66%, 75%, and 90% height of the beat respectively.

5) Acceleration Plethysmogram (APG) Features: The second derivative of PPG waveform, also called the acceleration plethysmogram (APG) is an indicator of the acceleration of the pulse through an artery. The waveform of the APG includes five waves, namely a-wave to e-wave as shown in Fig. 3. The ratios of the height of the each wave to that of the a-wave are commonly used as characteristic waveform features of APG [50]. APG intensity ratio (AIR), defined as the ratio of the APG intensity of the b-wave ($\mathrm{AI_b}$) to the APG intensity of the a-wave ($\mathrm{AI_a}$), is reported to reflect arterial stiffness and distensibility of the peripheral artery [51]. The c-wave to e-wave of APG are usually used to detect the dicrotic notch and diastolic peak point of PPG waveform [50]. However, a dicrotic notch is usually only seen in subjects with healthy compliant arteries [52]. According to a previous study, the higher harmonic frequency components of PPG diminish with age, consistent with the loss of the dicrotic notch features in older subjects [53]. Therefore, in this study, features related to the dicrotic notch such as c-wave to e-wave in APG waveform, augmentation index [51], and large artery stiffness index [54] were excluded from further analyses in order to develop a generalized BP estimation model for heterogeneous subjects. As a result, three features (AIR, $\mathrm{AI_a}$, and $\mathrm{AI_b}$) are extracted from the APG waveform.

6) Principal Component Analysis (PCA) Features: All of the above mentioned features are based on the assumptions that these characteristic points of the PPG are physiologically meaningful. On the other hand, principal component analysis (PCA) features are a representation of the whole PPG waveform shape and timing. This approach is presented as a "whole-based

feature" in a previous study [55]. In this paper, PCA features were extracted as followed. First, the PPG waveform in each cardiac cycle was re-sampled to 100 points. Second, each cardiac cycle of PPG was normalized by dividing by the maximum value of the waveform (the offset was previously removed during pre-processing steps). Finally, PCA is used to aggregate and analyze each pre-processed cardiac cycles of PPG for each subject. In other words, one cardiac cycle of PPG is treated as a 100-dimensional data point. Top 4 principal components (PC1, PC2, PC3, and PC4) were chosen among results of the PCA for further analyses.

C. Feature Selection

In general, the aims of feature selection process are: to eliminate irrelevant and redundant features to reduce the dimensionality of the data; and to gain a better understanding of the features and their relationship to the response variables [56]. In this study, several feature selection methods, which were more suitable for latter purpose, were applied to the extracted features in order to gain an interpretation of the features and their relationship to BP. Feature redundancy was not considered since the fundamental purpose of feature selection in this study was not to reduce the dimensionality of the features. Furthermore, noise reduction and better estimation performance might be obtained by adding variables that were presumed to be redundant [56].

Before performing feature selection process, the mean value of BP was subtracted from the BP values and the mean value of each feature was subtracted from the feature values in a given recording for the purpose of calibration, and then these values obtained from 1376 subjects were analyzed. As a result, a total of 2,470,560 pairs of BP and features were analyzed for feature interpretation.

All feature selection processes were performed using Python (version 3.6.3) and Scikit-learn library (version 0.19.1). Feature selection methods were categorized into filter methods, wrapper methods, and embedded methods [56]. A detailed description of each method is as follows:

1) Filter Methods: Filter methods are based on univariate statistical techniques to evaluate the strength of the relationship of each feature and the response variable [57]. In this study, Pearson correlation coefficient and mutual information (MI) were analyzed. While Pearson correlation captures linear dependency, MI, which measures mutual dependence between variables, is able to present the non-linear relationship between variables [58].

2) Wrapper Methods: Wrapper methods incorporate learning algorithms in the process of feature selection [57]. Features are ranked based on the regression performance. One of the representative techniques of a wrapper method is the recursive feature elimination (RFE) [59]. In this study, RFE with linear support vector machine (SVM) regression was performed. Varying the hyper-parameters of SVM such as epsilon or regularization parameter did not lead to significant difference in the RFE results. Therefore, epsilon was set to 0, and regularization parameter was set to 1.

3) Embedded Methods: Embedded methods perform feature selection in the process of model training and are usually model-specific [57]. Among the embedded methods, ridge regression [60], randomized lasso regression [61], and random forest's impurity based ranking [62] were performed in this study. Parameter setting for the random forest method is discussed in Section E.

Finally, a total of six feature selection methods were applied in this study in order to reliably assess the statistical significance of the relevance of the features to BP. The feature importance scores calculated by using each of the methods were normalized to a scale between 0 and 1. Then, the mean value of feature importance scores for each features were analyzed to gain an interpretation of the features and to determine the best BP estimation predictors amongst all the features.

For the purpose of validating the feature selection without considering feature redundancy, the result was compared with two well-known feature selection algorithms introduced in a prior research [63]. The first is the MaxRel algorithm that maximizes the average of the mutual information values between the response variable and each feature of the subset, and the second is the minimal-redundancy-maximal-relevance (mRMR) algorithm which combines MaxRel algorithm and min-redundancy criterion that minimizes the dependencies between features of the subset.

D. Frequency Component Separation of BP and Features

Some previous studies have suggested that BP estimation features such as PAT could track only certain frequency components of BP variation. Chen *et al.* reported that lower frequency component of PAT was not well correlated to SBP [8]. Therefore, they combined the high frequency component of PAT and used intermittent calibration to adjust for low frequency fluctuations of BP. Ding *et al.* also suggested that PAT could track BP in high frequency range well, but was inadequate to follow the low frequency variations in BP [29]. They proposed PIR as an indicator that could track BP in the low frequency range. These studies showed that the linear relationship between BP and certain predictors can be more apparent when BP and features are separated based on frequency. Therefore, estimation performance may be improved by separating BP and feature variations into low frequency components (LFC) and high frequency components (HFC), and modeling each separately.

In order to verify this hypothesis, extracted BP and features were separated into LFC and HFC. First, since the BP and features were extracted on beat-by-beat basis, they were re-sampled to a sampling rate of 1 Hz using cubic spline interpolation (i.e. resampled in the time domain to be 1 cardiac cycle per second). Then, LFC of BP and features were obtained by filtering using a 1st order low-pass Butterworth digital filter. HFC of BP and features were obtained by subtracting the LFC of BP and features from the raw BP and features. The cutoff frequency of the low-pass filter was set at 0.004 Hz, referring to a previous study showing a linear relationship between SBP and PAT between 0.00053 Hz to 0.004 Hz [8]. Fig. 4 shows the example plots of separating SBP and PAT_{IT} into their LFC and HFC. The optimal cutoff frequency was determined by comparing the performance

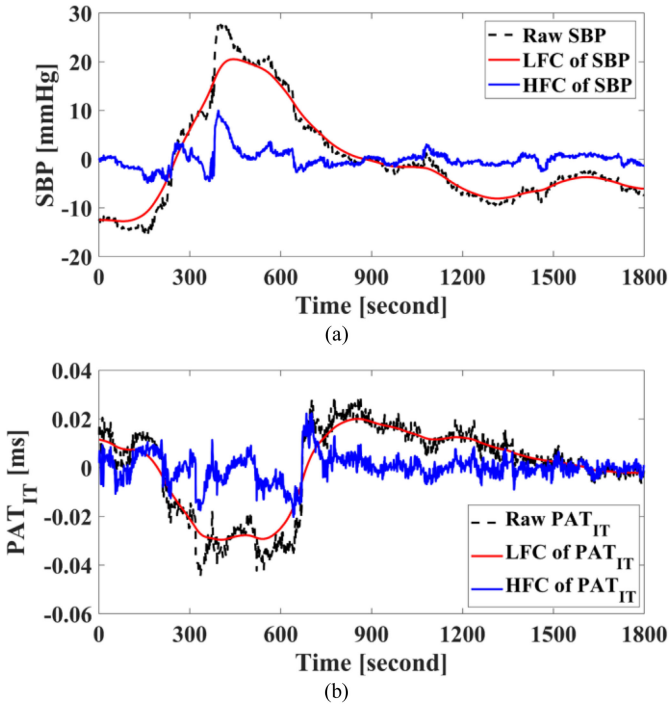

Fig. 4. Example plots of separating SBP and PAT_{IT} into LFC and HFC. (a) Example plot of raw SBP (dotted black line), LFC of SBP (solid red line), and HFC of SBP (solid blue line) from one representative subject; (b) Example plot of raw PAT_{IT} (dotted black line), LFC of PAT_{IT} (solid red line), and HFC of PAT_{IT} (solid blue line) from one representative subject. Raw SBP and PAT_{IT} are the calibrated values.

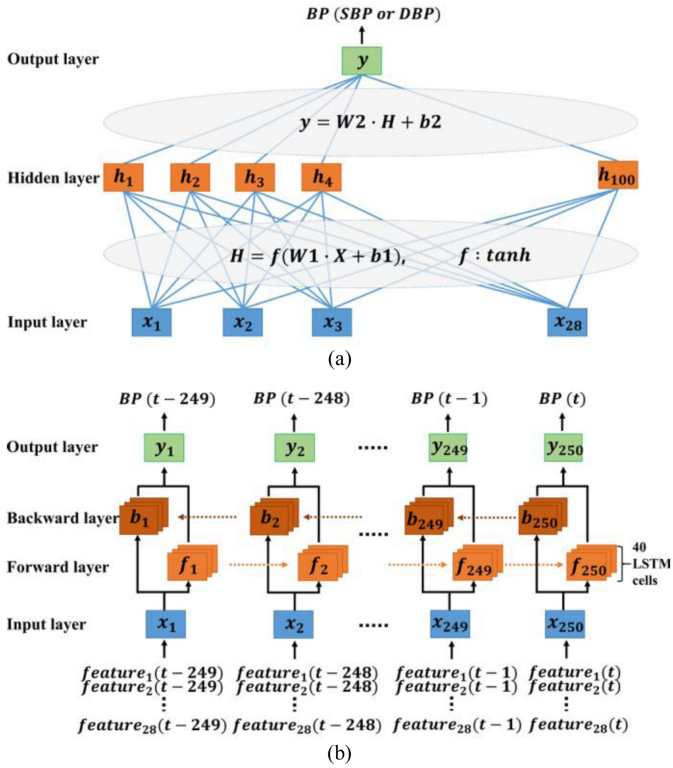

Fig. 5. The architectures of the ANN and RNN used in this study. (a) The architecture of the ANN; (b) The architecture of the RNN.

of the BP estimation model constructed by linear regression algorithm for each cutoff frequency.

E. BP Estimation Models

BP estimation models based on the selected features using several modeling algorithms were constructed and evaluated. The LFC and HFC of BP were modelled separately. That is, the LFC of BP was estimated by the LFC of the features, and the HFC of BP was estimated by the HFC of the features. In order to calibrate the model to each subject, the mean value of BP was subtracted from the BP values and the mean value of each feature was subtracted from the feature values in a given recording prior to modeling. Then, the mean BP was added back to the BP estimate model output for each subject.

The entire dataset consisting of 1376 subjects was randomly divided into two sets in 7:3 ratio; 963 recordings were used to train BP estimation model and 413 recordings were used for validation accounting for a total of 741,690 cardiac cycles. The two sample t-test between the training and the validation dataset showed no statistically significant differences in age, gender, height, weight, BMI, SBP values, and DBP values between the two groups. The modeling algorithms used in this study included linear regression, random forest, artificial neural network, and recurrent neural network. The detailed description of these algorithms is as followed:

1) Linear Regression (LR): As the baseline machine learning model, linear regression (LR) model was constructed by using BP as the dependent variable and the selected features as the explanatory variables. LR model was trained using a MATLAB script.

2) Random Forest (RF): Due to the size of the training dataset in this study (more than 1 million cardiac cycles), a method called "Extremely Randomized Trees", which is computationally less expensive than the typical random forest method was used [64]. Parameters such as the number of trees, the maximum number of leaf nodes, and the minimum number of samples required to be at a leaf node were tuned using the grid search method. The RF model in this study consisted of 500 trees with the maximum 10,000 leaf nodes, and the minimum 100 samples required to be at a leaf node. The training of RF was done using Python and Scikit-learn library.

3) Artificial Neural Network (ANN): Parameters of artificial neural network (ANN) such as the number of hidden layers, the number of hidden nodes, and the activation functions were chosen after considering multiple different combinations. The ANN model in this study consisted of an input layer, one hidden layer with 100 nodes using tanh function as the activation function, and one node for SBP or DBP prediction in the output layer as shown in Fig. 5(a). Mean Squared Error (MSE) was used as the cost function. To optimize the network, an ADAM optimizer with initial learning rate of 0.001 was used. The training of ANN was done using Python and Tensorflow (version 1.12.0).

4) Recurrent Neural Network (RNN): Recurrent neural network (RNN) has a recurring connection to itself at consecutive

TABLE III
FEATURE IMPORTANCE SCORES OF TOP 28 RANKED FEATURES

Rank	Feature	Feature importance score (feature selection analyses with SBP)						
		Pearson	MI	RF	RFE	Ridge	R-lasso	Mean
1	PPGK	**1.00**[a]	**0.93**[a]	**1.00**[a]	0.81	**0.83**[a]	**0.89**[a]	**0.91**[a]
2	PIR_P	**0.98**[a]	**0.95**[a]	**0.90**[a]	0.84	**0.74**[a]	**0.97**[a]	**0.90**[a]
3	HR	0.43	**1.00**[a]	**0.90**[a]	**1.00**[a]	**1.00**[a]	**1.00**[a]	**0.89**[a]
4	PAT_{IT}	**0.65**[a]	0.74	**0.98**[a]	**1.00**[a]	0.59	**0.91**[a]	**0.81**[a]
5	BW_{66}	**0.64**[a]	**0.88**[a]	0.33	**1.00**[a]	0.59	0.61	**0.68**[a]
6	BW_{75}	**0.63**[a]	**0.89**[a]	0.45	0.65	0.22	**0.89**[a]	0.62
7	AI_b	0.26	0.55	0.10	0.76	0.44	0.49	0.43
8	PI_P	0.22	0.66	0.11	0.78	0.55	0.25	0.43
9	SBW_{66}	0.23	0.30	0.06	0.97	**0.81**[a]	0.09	0.41
10	BWR_{50}	0.09	0.34	0.03	**1.00**[a]	**1.00**[a]	0.00	0.41
11	PAT_V	0.32	0.51	0.18	0.89	0.06	0.45	0.40
12	SBW_{50}	0.20	0.32	0.06	**1.00**[a]	0.74	0.00	0.39
13	DBW_{50}	0.34	0.60	0.07	0.86	0.47	0.01	0.39
14	PAT_{MD}	0.54	0.52	**0.57**[a]	0.11	0.06	0.50	0.38
15	BW_{90}	0.44	0.78	0.21	0.38	0.03	0.36	0.37
16	DBW_{90}	0.36	0.51	0.10	0.92	0.23	0.03	0.36
17	PC2	0.21	0.31	0.11	0.95	0.41	0.03	0.34
18	BW_{50}	0.48	0.82	0.13	0.46	0.03	0.09	0.34
19	AS	0.14	0.41	0.09	0.68	0.40	0.09	0.30
20	DBW_{33}	0.17	0.62	0.04	0.49	0.31	0.00	0.27
21	SBW_{75}	0.22	0.28	0.04	0.73	0.29	0.04	0.27
22	PAT_P	0.14	0.52	0.16	0.51	0.13	0.12	0.26
23	DBW_{75}	0.49	0.66	0.21	0.05	0.00	0.09	0.25
24	BWR_{66}	0.19	0.39	0.02	0.70	0.20	0.00	0.25
25	DBW_{66}	0.48	0.67	0.13	0.14	0.06	0.03	0.25
26	PI_V	0.03	0.55	0.06	0.62	0.26	0.00	0.25
27	CT	0.09	0.40	0.07	0.59	0.27	0.00	0.24
28	PC1	0.01	0.28	0.18	0.57	0.11	0.00	0.19

[a]Top 5 features in each feature selection method.

time points, thus the features from a previous time step can influence the prediction of BP at a future time step [65]. In this study, a RNN model was constructed in many-to-many fashion consisting of bidirectional long short-term memory (LSTM) cells [66], [67]. Similar to ANN, the architecture of RNN developed in this study was chosen after considering multiple different architectures. The chosen network consisted of 2 hidden layers with 40 LSTM cells in each layer, and 250 time steps as shown in Fig. 5(b). To optimize the network, an ADAM optimizer with initial learning rate of 0.01 was used. The training of RNN was done using Python and Tensorflow.

The performance of the constructed BP models was evaluated against three international standards. First, mean error (ME) and standard deviation of the error (SDE) were calculated to validate the models against the Association for the Advancement of Medical Instrumentation (AAMI) standards, which requires ME and SDE values lower than 5 and 8 mmHg [68]. Second, the cumulative error percentages within 5, 10, and 15 mmHg were calculated to evaluate the models against the British Hypertension Society (BHS) BP monitor standards, which grades devices based on their cumulative error percentages under three different thresholds [69]. Finally, the mean absolute difference (MAD) was calculated to evaluate the models in accordance with the latest standard for wearable and cuffless BP monitoring devices published by the IEEE Engineering in Medicine and Biology Society (IEEE Standard 1708) [70]. All performance measures were calculated between the estimated BP (i.e. the sum of the outputs of the LFC estimation model and the HFC estimation model) and the reference BP.

III. RESULTS

A. Feature Selection Results

The results of feature selection processes analyzing a total of 2,470,560 pairs of SBP and the extracted 42 features are shown in Table III. Although the top features ranked by different feature selection methods vary as shown in Table III, PPGK, PIR_P, HR, PAT_{IT}, BW_{66}, and BW_{75} rank high across almost all methods. Among them, PPGK has the highest feature importance score with a mean value of 0.91. The absolute Pearson correlation coefficient value of PPGK was 0.49 for SBP, and 0.48 for DBP. Among the four PAT values, PAT_{IT} showed the highest mean feature importance score with the highest Pearson correlation, consistent with the findings of the previous study [40]. Among the PPG morphology features, PPGK and PIR_P along with HR showed higher mean feature importance score when compared to PAT_{IT}. Similar results were found on the same feature selection analyses with DBP.

When comparing SDE and MAD values calculated from the linear regression models constructed by sequentially adding features in order of highest mean feature importance score, there were no significant differences between the model using only 28 features and one using all features. Therefore, PPGK, PIR_P, HR, PAT_{IT}, BW_{66}, BW_{75}, AI_b, PI_P, SBW_{66}, BWR_{50}, PAT_V,

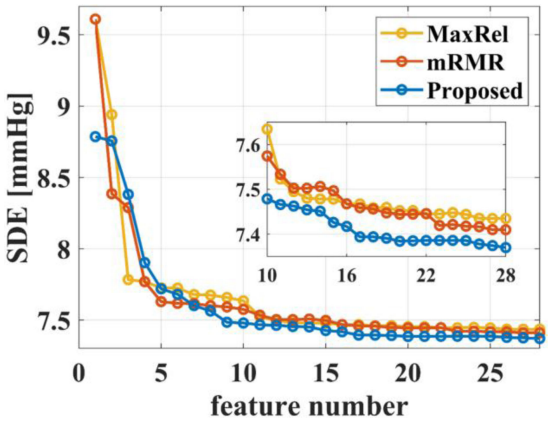

Fig. 6. Comparison of estimation performance of proposed, MaxRel, and mRMR algorithms.

TABLE IV
PERFORMANCE OF BP LINEAR MODELS WITH DIFFERENT CUTOFF FREQUENCIES OF FILTER SEPARATING LFC/HFC ACROSS 413 VALIDATION RECORDINGS

Cutoff frequency	SBP			DBP		
	ME	SDE	MAD	ME	SDE	MAD
0[a]	<0.001	7.43	5.49	<0.001	4.17	3.00
0.001	-0.018	7.37	5.43	-0.004	4.14	2.98
0.002	-0.011	7.27	5.35	-0.005	4.10	2.95
0.004	-0.004	7.19	5.30	-0.002	4.08	2.93
0.006	**-0.002**	**7.17**	**5.29**	**-0.001**	**4.07**	**2.92**
0.008	-0.001	7.18	5.30	-0.001	4.07	2.93
0.016	-0.001	7.22	5.34	<0.001	4.09	2.95

[a]In the case of modeling without LFC/HFC separation.

SBW_{50}, DBW_{50}, PAT_{MD}, BW_{90}, DBW_{90}, PC2, BW_{50}, AS, DBW_{33}, SBW_{75}, PAT_{P}, DBW_{75}, BWR_{66}, DBW_{66}, PI_{V}, CT, and PC1 were selected as the BP estimation predictors (totaling 28 features) for further analyses.

In order to validate the feature selection results, the proposed feature selection algorithm in this study without considering feature redundancy was compared with the MaxRel algorithm and the mRMR algorithm. The three methods (proposed, MaxRel, and mRMR) were used to select 28 sequential feature sets and computed the respective SDE values using linear regression. As shown in Fig. 6, for most $k \in \{1, 28\}$, smaller SDE values on proposed feature sets were obtained.

B. Optimization of the BP Estimation Models

The performance of the BP linear models constructed by changing the cutoff frequency of the filter separating LFC/HFC of BP and features is shown in Table IV. As shown in Table IV, modeling LFC and HFC of BP separately showed slightly better performance as compared to modeling without LFC/HFC separation. The optimal cutoff frequency was found to be 0.006 Hz.

The performance of the SBP models with different modeling algorithms is shown in Table V. All modeling algorithms were constructed by separating LFC/HFC of BP and features with a cutoff frequency of 0.006 Hz. The performance measures were calculated for the LFC estimation model and the HFC estimation model separately. Since the ME values between

TABLE V
PERFORMANCE OF SBP ESTIMATION MODELS WITH DIFFERENT MODELING ALGORITHMS ACROSS 413 VALIDATION RECORDINGS

Modeling algorithm	SBP					
	LFC			HFC		
	ME	SDE	MAD	ME	SDE	MAD
LR	-0.002	6.24	4.63	<0.001	2.20	1.42
RF	-0.135	6.32	4.59	0.003	2.15	1.37
ANN	**-0.025**	**6.07**	**4.48**	0.008	2.16	1.39
RNN	0.310	6.91	4.99	**0.077**	**2.14**	**1.36**

TABLE VI
PERFORMANCE OF BP ESTIMATION MODELS WITH DIFFERENT MODELING SETTINGS ACROSS 413 VALIDATION RECORDINGS

Model index	Modeling methods	SBP			DBP		
		ME	SDE	MAD	ME	SDE	MAD
1	LR (LFC+HFC)[a] – PAT[b]	<0.001	8.83	6.53	<0.001	4.98	3.59
2	LR (LFC+HFC)[a]	<0.001	7.43	5.49	<0.001	4.17	3.00
3	LR(LFC) + LR(HFC)	-0.002	7.17	5.29	-0.001	4.07	2.93
4	ANN(LFC) + RNN(HFC)	0.052	6.92	5.07	-0.054	3.99	2.86

[a]In the case of modeling without LFC/HFC separation.
[b]In the case of modeling using only 4 PAT (PATV, PATP, PATMD, and PATIT) values as estimation predictors.

different algorithms did not differ significantly, the performance comparisons between algorithms were conducted based on their SDE and MAD values. As shown in Table V, ANN showed the best performance in LFC modeling, and RNN showed the best performance in HFC modeling. Similar results were found on the same analyses with DBP.

The performances of the BP estimation models with varying settings are shown in Table VI. Comparing model 1 and model 2, it is confirmed that the estimation performance is improved by adding several PPG morphology features to the estimation parameters in addition to the conventional PAT values. Moreover, it can be seen that the application of the LFC/HFC separation method (model 2 versus model 3) as well as using non-linear machine learning based regression method (model 3 versus model 4) further improve the estimation performance.

C. Performance of the Best-Case BP Estimation Models

The best performing was model 4 in Table VI, which is obtained by modeling LFC of BP as ANN and HFC of BP as RNN using all 28 selected features as predictors. The detailed analysis of the best-case BP estimation model is shown in Table VII and Fig. 7. As shown in Table VII, both SBP and DBP estimation models satisfy the AAMI standards and are rated grade A against the BHS standards. Based on the IEEE standards, both SBP and DBP models pass the standards as grade B for SBP, and as grade A for DBP.

For the purpose of external validation, the constructed BP estimation model was evaluated using an external dataset from the MIMIC III waveform database [71]. Data from part 0 of the MIMIC III waveform database matched subset was used. Data loading, pre-processing and feature extraction processes were performed in the same manner as described above. As a result, a total of 750,898 pairs of BP and features from 334 subjects are extracted. The estimation performance of the model across the

TABLE VII
PERFORMANCE OF THE BEST-CASE BP ESTIMATION MODEL ACROSS 413 VALIDATION RECORDINGS

Performance Measure	SBP estimation model		DBP estimation model	
	Value	Evaluation	Value	Evaluation
ME (mmHg)	0.052	**PASS**[a]	-0.054	**PASS**[a]
SDE (mmHg)	6.92	**PASS**[a]	3.99	**PASS**[a]
Cumulative Error < 5 mmHg (%)	62	**A**[b]	84	**A**[b]
Cumulative Error < 10 mmHg (%)	87	**A**[b]	97	**A**[b]
Cumulative Error < 15 mmHg (%)	96	**A**[b]	99	**A**[b]
MAD (mmHg)	5.07	**B**[c]	2.86	**A**[c]
Correlation to reference (mean ±STD)	0.63 ± 0.36		0.60 ± 0.36	

[a]Evaluation against the AAMI standards.
[b]Evaluation against the BHS BP monitor standards.
[c]Evaluation against the IEEE standards.

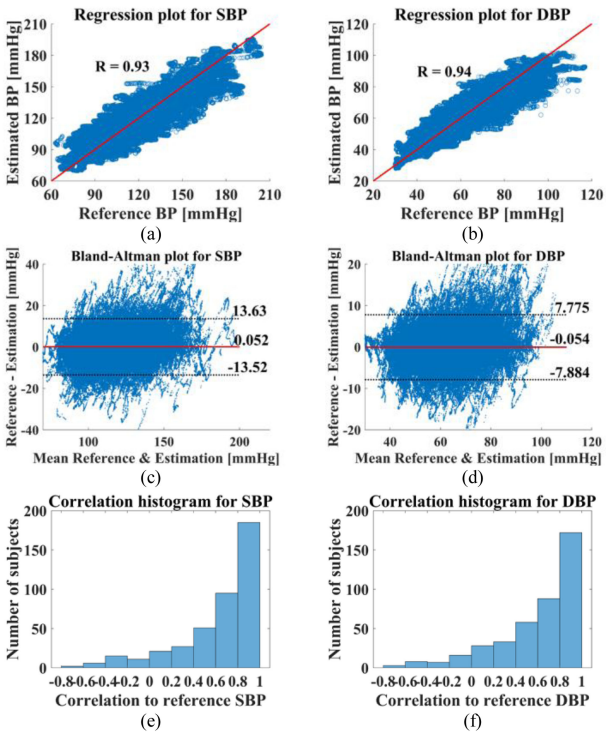

Fig. 7. Regression plots, Bland-Altman plots, and correlation histograms of the best-case SBP and DBP estimation models across 413 validation recordings. (a) Regression plot for SBP estimation; (b) Regression plot for DBP estimation; (c) Bland-Altman plot for SBP estimation; (d) Bland-Altman plot for DBP estimation. (e) Correlation histogram for SBP estimation; (f) Correlation histogram for DBP estimation.

external validation dataset is shown in Table VIII. As shown in Table VIII, the model satisfies AAMI, BHS, and IEEE standards for both SBP and DBP estimated in the external database. Example plots of the reference BP and the BP estimates from four representative subjects in the external database are shown in Fig. 8.

IV. DISCUSSION

In general, previous studies of BP estimation using PPG morphology features have analyzed different features and have used small homogeneous subject pools to generate models of

TABLE VIII
PERFORMANCE OF THE BEST-CASE BP ESTIMATION MODEL ACROSS 334 EXTERNAL VALIDATION RECORDINGS

Performance Measure	SBP estimation model		DBP estimation model	
	Value	Evaluation	Value	Evaluation
ME (mmHg)	-0.006	**PASS**[a]	-0.004	**PASS**[a]
SDE (mmHg)	7.04	**PASS**[a]	4.77	**PASS**[a]
Cumulative Error < 5 mmHg (%)	61	**A**[b]	86	**A**[b]
Cumulative Error < 10 mmHg (%)	87	**A**[b]	96	**A**[b]
Cumulative Error < 15 mmHg (%)	95	**A**[b]	98	**A**[b]
MAD (mmHg)	5.13	**B**[c]	2.81	**A**[c]
Correlation to reference (mean ±STD)	0.52 ± 0.40		0.53 ± 0.38	

[a]Evaluation against the AAMI standards.
[b]Evaluation against the BHS BP monitor standards.
[c]Evaluation against the IEEE standards.

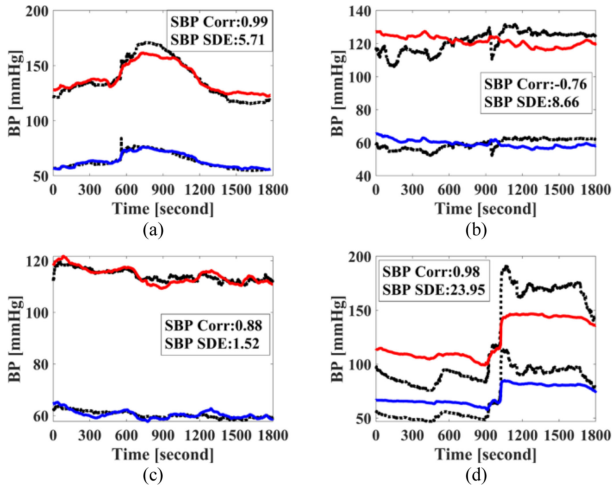

Fig. 8. Example plots of estimated BP (solid red lines for SBP, and solid blue lines for DBP) and reference BP (dotted black lines) in 4 external validation recordings. The correlation and SDE values for the SBP estimation are labeled. (a) The recording with the highest SBP correlation; (b) The recording with the lowest SBP correlation; (c) The recording with the lowest SBP SDE; (d) The recording with the highest SBP SDE.

BP. This is the first study to use a large heterogeneous dataset to validate the potential of a diverse PPG morphology features as predictors of BP comprehensively. The key findings of the study were that 1) 28 features including several PPG morphology features and PAT were determined as suitable BP predictors, in particular two PPG morphology features (PPGK and PIR_P) outperformed PAT, and 2) BP estimation models constructed by modeling LFC using ANN and HFC using RNN with selected 28 predictors showed the best estimation performance, satisfying the three relevant international standards for BP monitors.

A. PPG Morphology Features as Predictors of BP in Ubiquitous NCBPM

Although PAT or PTT based approach for continuous BP monitoring has already shown remarkable progress through numerous studies [8], [10], [14], [19], [21]–[24], it has not yet been widely accepted for ubiquitous NCBPM due to its limited accuracy. One of the reasons for this may be that PAT alone

cannot fully reflect the complex BP dynamics and that conventionally measured PAT in ubiquitous BP monitoring involves PAT measured at peripheral arteries rather than central arteries. PAT is associated with BP through the physiological characteristics of arterial vessels such as arterial elastance, but BP is affected by physiological factors other than arterial elastance [15]. In order to address this issue, many studies have attempted to find new predictors for BP estimation through extracting physiologically meaningful features from the morphology of PPG waveforms [25]–[31], [33]–[38]. However, to the authors' knowledge, there has not been a study that confirmed to the validity of these features as predictors of BP in a large diverse dataset.

In this study, a large biosignal database from a diverse group of 1376 surgical patients were analyzed to assess the efficacy of various PPG morphology features and PAT values. As a result, PPGK, PIR_P, HR, PAT_{IT}, and BW_{66} were found to be the top 5 important features according to several feature selection methods. PPGK and PIR_P, which may be physiologically related to total peripheral resistance as described in Methods B, ranked top 2 in the mean feature importance scores. Similar results were found in a prior study analyzing several PPG morphology features using a genetic algorithm based feature selection method, although the number of subjects in the study was less than 100 [28]. Since BP can be as the product of cardiac output and total peripheral resistance by applying Ohm's law in the hydraulic version, the variation of arterial BP relies upon the changes of the total peripheral resistance to a degree [15]. Among the branch width based features which are also reported to be associated with total peripheral resistance, features extracted at the mid-level height (i.e. 50%, 66%, and 75%) of the PPG waveform ranked high in general. In addition, BW values generally showed high mean feature importance scores when compared to SBW, DBW, and BWR. As a result, 28 features including the conventional PAT values and HR were determined as suitable BP estimation predictors. In particular, 2 features (i.e. PPGK and PIR_P) along with HR were even found to be superior to traditional PAT (PAT_{IT}, specifically) as BP estimation predictors.

The performance of the BP estimation model constructed by using only PPG morphology features without PAT and HR as estimation predictors and applying the best-case modeling method mentioned above is shown in Table IX. Although the model is inferior to the model that includes PAT features and HR, it was able to output seemingly decent performance that satisfies all relevant standards. This implies the potential for simple ubiquitous NCBPM using PPG measured at the finger without bulky electrode attachments for ECG. Since the ECG waveform was used in pre-processing of the PPG waveforms for the extraction of the PPG morphology features in this study, a further study is warranted to develop a method solely based on PPG for BP estimation that completely excludes ECG.

TABLE IX
PERFORMANCE OF THE BP ESTIMATION MODEL USING PPG MORPHOLOGY FEATURES WITHOUT PAT AND HR ACROSS 413 VALIDATION RECORDINGS

Performance Measure	SBP estimation model		DBP estimation model	
	Value	Evaluation	Value	Evaluation
ME (mmHg)	0.016	**PASS**[a]	-0.043	**PASS**[a]
SDE (mmHg)	7.66	**PASS**[a]	4.22	**PASS**[a]
Cumulative Error < 5 mmHg (%)	58	**B**[b]	81	**A**[b]
Cumulative Error < 10 mmHg (%)	84	**B**[b]	97	**A**[b]
Cumulative Error < 15 mmHg (%)	94	**B**[b]	99	**A**[b]
MAD (mmHg)	5.57	**B**[c]	3.03	**A**[c]
Correlation to reference (mean ±STD)	0.57 ± 0.40		0.55 ± 0.39	

[a]Evaluation against the AAMI standards.
[b]Evaluation against the BHS BP monitor standards.
[c]Evaluation against the IEEE standards.

TABLE X
PEARSON CORRELATION COEFFICIENTS AND MI VALUES BETWEEN SBP AND TOE 10 RANKED FEATURES ACROSS 1376 RECORDINGS

Rank	Feature	LFC+HFC[a]		LFC[b]		HFC[c]	
		Pearson[d]	MI	Pearson[d]	MI	Pearson[d]	MI
1	PPGK	0.49	0.35	0.52	1.07	0.28	0.10
2	PIR_P	0.49	0.35	0.52	1.06	0.28	0.10
3	HR	0.34	0.37	0.33	1.06	0.32	0.12
4	PAT_{IT}	0.42	0.30	0.44	1.01	0.21	0.05
5	BW_{66}	0.42	0.35	0.45	1.09	0.18	0.08
6	BW_{75}	0.41	0.35	0.45	1.09	0.18	0.08
7	AI_b	0.29	0.24	0.31	0.99	0.12	0.05
8	PI_P	0.27	0.27	0.29	0.99	0.12	0.06
9	SBW_{66}	0.26	0.17	0.30	0.86	0.08	0.03
10	BWR_{50}	0.18	0.19	0.20	0.91	0.07	0.03

[a]Correlation between the raw SBP and the raw features.
[b]Correlation between the LFC of SBP and the LFC of features.
[c]Correlation between the HFC of SBP and the HFC of features.
[d]Absolute values of Pearson correlation coefficients.

B. Frequency Component Separation of BP and Features

In order to prove the hypothesis that the relationship between BP and features can be clarified in a specific frequency band based on previous studies, frequency components of BP and features were separated. As a result of separately modeling LFC and HFC of BP with the optimal cutoff frequency at 0.006 Hz using linear regression, the performance of the estimation model was improved as compared to modeling both components of BP simultaneously. Similar results were found using non-linear regression methods such as ANN and RNN. The Pearson correlation coefficients and MI values between SBP and top 10 ranked features across 1376 subjects calculated for each of the LFC and HFC are shown in Table X. The correlation values and mutual dependence values of the features are generally lower in HFC, but almost all features are significantly better related to SBP in low frequency domain. Performing power spectral analysis, the power over low frequency bandwidth (0.0005–0.006 Hz) and the power over high frequency bandwidth (over 0.006 Hz) are calculated as the area under the power spectral density of SBP variation of each subject. As a result, the mean power across 1376 subjects are 101.16 mmHg^2 for LFC, and 4.96 mmHg^2 for HFC. Since the LFC of BP contains most of the components of BP variation, the enhanced relationships between BP and features in the LFC might improve the overall estimation performance.

TABLE XI
THE COMPARISON OF THE RESULTS OF THIS WORK WITH OTHER WORKS

Work	Subject	Estimation algorithm	Performance measure of SBP (mmHg)			Performance measure of DBP (mmHg)		
			SDE	MAD	RMSE[a]	SDE	MAD	RMSE[a]
[26]	15	LR	7.32	-	-	4.39	-	-
[34]	15000 (beats)	ANN	-	3.80	-	-	2.21	-
[35]	4254 (records)	SVM, ANN	16.17	12.38	-	8.45	6.34	-
[29]	27	LR	5.21	4.09	-	4.06	3.18	-
[28]	73	LR, SVM	3.10	-	-	2.20	-	-
[25]	84	RNN	-	-	3.73	-	-	2.43
[30]	32	LR, SVM, RF	6.50	-	-	5.20	-	-
[38]	772 (records)	LR, SVM	5.52	3.27	-	1.97	1.16	-
[33]	441	SVM, RF	8.90	3.97	8.90	4.17	2.43	4.18
[31]	103	LR, RF, ANN, RNN	-	-	7.86	-	-	6.49
This work	1376	LR, RF, ANN, RNN	6.92	5.07	6.92	3.99	2.86	3.99

[a]Root mean square error.

C. Comparison of Modeling Algorithms

Several modeling algorithms including LR, RF, ANN, and RNN were evaluated. When comparing the estimation performances between these algorithms, the overall accuracy improved when using non-linear regression methods as compared to LR in terms of SDE and MAD, potentially indicating nonlinear relationships between the features and BP values. However, due to the moderate degree of linear association between the features and BP values, the differences of the error values between LR and nonlinear modeling algorithms were minimal.

For LFC modeling, ANN showed the best performance in terms of SDE and MAD. Comparing the mean squared error values between models constructed by changing the number of hidden layers and the number of hidden nodes in the architecture of ANN, the optimal architecture was 100 hidden nodes with 1 hidden layer. In the case of the hidden layer activation function, models using tanh function outperformed models using relu or sigmoid functions.

In the case of RNN, it achieved the best performance in HFC modeling. However, it performed worse than ANN and LR in LFC modeling. There may be a few reasons for this lack of prediction performance, but one may be due to the problem of long-term dependencies [72]. Since LFC only contains slowly varying components below 0.006 Hz, the size of the time step of RNN modeling LFC should be large enough to properly capture the sequential change of features over time in the input data. A further study in which the period between each input time step may provide a suitable solution to this issue and such methods can be used to further improve estimation performance in the future.

Due to the range of complexities of the modeling algorithms used in this study, the duration for training varied widely. However, in terms of post optimization application of the algorithms, all of the methods were able to output a BP estimate within 600 microseconds using the current graphics processing unit (Geforce RTX 2080Ti; NVIDIA, Santa Clara, CA, USA), which is sufficient for pseudo real-time estimation of BP.

D. Performance of the Best-Case BP Estimation Model

The best-case BP estimation model was obtained by modeling the LFC of BP values using ANN and the HFC of BP values using RNN with a total of 28 features as estimation predictors. The model satisfied AAMI, BHS, and IEEE standards for both SBP and DBP estimates. In particular, in the case of SBP estimation, even though the high absolute magnitude of SBP led to large errors, the model was rated grade A against the BHS standards and nearly grade A against the IEEE standards. The comparisons of the estimation results with prior studies which have estimated BP based on PPG morphology features are presented in Table XI. However, direct comparisons of estimation error values are difficult since the factors affecting the absolute magnitude of the errors, such as the degree of BP variations of the data used, the number of subjects, and the calibration method, are all different.

The distribution of the correlation values between the reference BP and the estimated BP across 413 validation datasets are shown in Fig. 7(e) and Fig. 7(f). For SBP, about 70% of the subjects showed the correlation values higher than 0.6, and 45% of the subjects showed higher than 0.8. Similar trends were seen on the same analyses with DBP. Unlike most of the previous studies used only error-based validation criteria, these results present the degree of association in trends between the reference BP and the estimated values. However, as some cases still show low or opposite correlation, a further study is warranted to further improve the estimation performance in terms of correlation.

According to the external validation results, the model satisfies AAMI, BHS, and IEEE standards for both SBP and DBP estimates in the external database, indicating that the developed BP estimation model can be applied to data with different measurement modalities and different demographics. As shown in Table VIII and Fig. 8(b), in terms of correlation to reference BP, the performance was degraded, which might be caused due to the lower BP variations in the test dataset (the standard deviation values of SBP were 10.3 mmHg for validation dataset and 8.6 mmHg for test dataset), or incorrect PAT values due to the inter-waveform alignment issue. As shown in Fig. 8(d), a sample with very low or very high BP values tended not to be accurately estimated. It might be caused due to the limited number of subjects with very high or very low BP values in the training dataset, or it may have been caused by sharp changes in BP during this particular surgery, which is not usually observed. However, the overall trend of BP variations was well tracked as indicated by the high correlation value in Fig. 8(d).

E. Limitations

As mentioned in previous research [40], there is a consistent delay between the biosignals in VitalDB due to the experimental setup. The mean PAT_{IT} across 2,470,560 cardiac cycles of 1376 subjects was 659 ms, which is abnormally large compared to previous reports of PAT values derived from ECG to finger PPG. However, the standard deviation of PAT_{IT} across 1376 subjects was about 30 ms, which indicates the time delay between ECG and PPG was nearly consistent. The authors have verified with the creators of the database the consistency of this delay throughout all recordings used. Through the calibration procedure, it is ensured that this delay does not affect the analyses of this study.

Another limitation is that the VitalDB database used in this study only contains biosignals obtained from patients undergoing surgery. Most of the data was measured from motionless subjects who were in supine position. However, previous research has shown that postural changes affect the morphology of PPG [73]. In terms of ubiquitous NCBPM, the underlying mechanism of BP variation during surgery may be very different as compared to the ones dominating the circadian fluctuations of BP, and the range of BP during surgery is dependent on each case and may vary wildly as compared to the normal range of BP observed in daily life. Therefore, a further study is warranted to validate the estimation model against data measured in less controlled environment.

V. Conclusion

In this study, various features obtained from the morphology of PPG waveforms were analyzed across a large heterogeneous group of 1376 subjects, and BP estimation models using these features in conjunction with PAT were evaluated. The results of the analyses indicate that 1) several PPG morphology features outperform PAT as a predictor for BP, 2) modeling low and high frequency components of BP separately can improve estimation performance, and 3) modeling using non-linear regression methods based on ANN or RNN can improve estimation performance as compared to linear regression. The best-case BP estimation model constructed by applying these three points yielded an improved performance that satisfied AAMI, BHS, and IEEE standards for both SBP and DBP in validation and in external validation datasets.

VII. Conflicts of Interest

Joonnyong Lee declares conflict of interest as the CEO of Mellowing Factory Co., Ltd., which is developing a deep learning based non-invasive blood pressure estimation system. Hee Chan Kim is an investor of Mellowing Factory Co., Ltd.

References

[1] World Health Organization (WHO), "Cardiovascular Diseases," WHO: Geneva, Switzerland, 2015.

[2] World Health Organization (WHO), "A global brief on hypertension: Silent Killer, global public health crisis: World health day 2013," WHO: Geneva, Switzerland, 2013.

[3] P. K. Whelton, "Epidemiology of hypertension," *The Lancet*, vol. 344, pp. 101–106, 1994.

[4] V. L. Burt *et al.*, "Prevalence of hypertension in the US adult population," *Results 3rd Nat. Health Nutrition Examination Survey, 1988-1991*, vol. 25, pp. 305–313, 1995.

[5] T. G. Pickering, G. A. Harshfield, R. B. Devereux, and J. H. Laragh, "What is the role of ambulatory blood pressure monitoring in the management of hypertensive patients?" *Hypertension*, vol. 7, pp. 171–177, 1985.

[6] P. Verdecchia *et al.*, "Ambulatory blood pressure. An independent predictor of prognosis in essential hypertension," *Hypertension*, vol. 24, pp. 793–801, 1994.

[7] V. Chandrasekaran, R. Dantu, S. Jonnada, S. Thiyagaraja, and K. P. Subbu, "Cuffless differential blood pressure estimation using smart phones," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 4, pp. 1080–1089, Apr. 2013.

[8] W. Chen, T. Kobayashi, S. Ichikawa, Y. Takeuchi, and T. Togawa, "Continuous estimation of systolic blood pressure using the pulse arrival time and intermittent calibration," *Med. Biol. Eng. Comput.*, vol. 38, pp. 569–574, 2000.

[9] L. A. Geddes, M. H. Voelz, C. F. Babbs, J. D. Bourland, and W. A. Tacker, "Pulse transit time as an indicator of arterial blood pressure," *Psychophysiology*, vol. 18, pp. 71–74, 1981.

[10] H. Gesche, D. Grosskurth, G. Küchler, and A. Patzak, "Continuous blood pressure measurement by using the pulse transit time: comparison to a cuff-based method," *Eur. J. Appl. Physiol.*, vol. 112, pp. 309–315, Jan. 01, 2012.

[11] F. Parry, G. Dumont, C. Ries, C. Mott, and M. Ansermino, "Continuous noninvasive blood pressure measurement by pulse transit time," in *Proc. 26th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2004, pp. 738–741.

[12] C. C. Y. Poon and Y. T. Zhang, "Cuff-less and noninvasive measurements of arterial blood pressure by pulse transit time," in *Proc. IEEE Eng. Med. Biol. 27th Annu. Conf.*, 2005, pp. 5877–5880.

[13] J. Solà *et al.*, "Noninvasive and nonocclusive blood pressure estimation via a chest sensor," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 12, pp. 3505–3513, Dec. 2013.

[14] M. Y.-M. Wong, C. C.-Y. Poon, and Y.-T. Zhang, "An evaluation of the cuffless blood pressure estimation based on pulse transit time technique: a half year study on normotensive subjects," *Cardiovascular Eng.*, vol. 9, pp. 32–38, Mar. 01, 2009.

[15] X. Ding and Y.-T. Zhang, "Pulse transit time technique for cuffless unobtrusive blood pressure measurement: From theory to algorithm," *Biomed. Eng. Lett.*, vol. 9, pp. 37–52, Feb. 01, 2019.

[16] F. S. Cattivelli and H. Garudadri, "Noninvasive cuffless estimation of blood pressure from pulse arrival time and heart rate with adaptive calibration," in *Proc. 6th Int. Workshop Wearable Implantable Body Sensor Netw.*, 2009, pp. 114–119.

[17] L. Geddes, M. Voelz, S. James, and D. Reiner, "Pulse arrival time as a method of obtaining systolic and diastolic blood pressure indirectly," *Med. Biol. Eng. Comput.*, vol. 19, pp. 671–672, 1981.

[18] J. Muehlsteff, X. Aubert, and G. Morren, "Continuous cuff-less blood pressure monitoring based on the pulse arrival time approach: The impact of posture," in *Proc. 30th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2008, pp. 1691–1694.

[19] Z. Tang *et al.*, "A chair-based unobtrusive cuffless blood pressure monitoring system based on pulse arrival time," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 5, pp. 1194–1205, Sep. 2017.

[20] X. Ding, B. P. Yan, Y.-T. Zhang, J. Liu, N. Zhao, and H. K. Tsang, "Pulse transit time based continuous cuffless blood pressure estimation: A new extension and a comprehensive evaluation," *Sci. Rep.*, vol. 7, 2017, Art. no. 11554.

[21] S. Puke, T. Suzuki, K. Nakayama, H. Tanaka, and S. Minami, "Blood pressure estimation from pulse wave velocity measured on the chest," in *Proc. 35th Annu. Int. Conf. IEEE Eng. Medicine Biol. Soc.*, 2013, pp. 6107–6110.

[22] C. Kim, A. M. Carek, O. T. Inan, R. Mukkamala, and J. Hahn, "Ballistocardiogram-based approach to cuffless blood pressure monitoring: Proof of concept and potential challenges," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 11, pp. 2384–2391, Nov. 2018.

[23] J. Proença, J. Muehlsteff, X. Aubert, and P. Carvalho, "Is pulse transit time a good indicator of blood pressure changes during short physical exercise in a young population?" in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol.*, 2010, pp. 598–601.

[24] L. Peter, N. Noury, and M. Cerny, "A review of methods for non-invasive and continuous blood pressure monitoring: Pulse transit time method is promising?" *IRBM*, vol. 35, pp. 271–282, 2014.

[25] P. Su, X.-R. Ding, Y.-T. Zhang, J. Liu, F. Miao, and N. Zhao, "Long-term blood pressure prediction with deep recurrent neural networks," in *Proc. Biomed. Health Inform. IEEE EMBS Int. Conf.*, 2018, pp. 323–328.

[26] X. Teng and Y. Zhang, "Continuous and noninvasive estimation of arterial blood pressure using a photoplethysmographic approach," in *Proc. Eng. Med. Biol. Soc. Proc. 25th Annu. Int. Conf. IEEE*, 2003, pp. 3153–3156.

[27] W.-H. Lin, H. Wang, O. W. Samuel, G. Liu, Z. Huang, and G. Li, "New photoplethysmogram indicators for improving cuffless and continuous blood pressure estimation accuracy," *Physiological Meas.*, vol. 39, 2018, Art. no. 025005.

[28] F. Miao *et al.*, "A Novel Continuous Blood Pressure Estimation Approach Based on Data Mining Techniques," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 6, pp. 1730–1740, Nov. 2017.

[29] X. Ding, Y. Zhang, J. Liu, W. 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.

[30] S. G. Khalid, J. Zhang, F. Chen, and D. Zheng, "Blood pressure estimation using photoplethysmography only: Comparison between different machine learning approaches," *J. Healthcare Eng.*, vol. 2018, pp. 1–13, 2018.

[31] M. Radha *et al.*, "Estimating blood pressure trends and the nocturnal dip from photoplethysmography," *Physiological Meas.*, vol. 40, 2019, Art. no. 025006.

[32] Q. Liu, B. P. Yan, C.-M. Yu, Y.-T. Zhang, and C. C. 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. 2013.

[33] S. S. Mousavi, M. Firouzmand, M. Charmi, M. Hemmati, M. Moghadam, and Y. Ghorbani, "Blood pressure estimation from appropriate and inappropriate PPG signals using A whole-based method," *Biomed. Signal Process. Control*, vol. 47, pp. 196–206, 2019.

[34] Y. Kurylyak, F. Lamonaca, and D. Grimaldi, "A Neural network-based method for continuous blood pressure estimation from a PPG signal," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf.*, 2013, pp. 280–283.

[35] M. Kachuee, M. M. Kiani, H. Mohammadzade, and M. Shabany, "Cuffless high-accuracy calibration-free blood pressure estimation using pulse transit time," in *Proc. IEEE Int. Symp. Circuits Syst.*, 2015, pp. 1006–1009.

[36] R. He, Z. Huang, L. Ji, J. Wu, H. Li, and Z. Zhang, "Beat-to-beat ambulatory blood pressure estimation based on random forest," in *Proc. IEEE 13th Int. Conf. Wearable Implantable Body Sensor Netw.*, 2016, pp. 194–198.

[37] A. Gaurav, M. Maheedhar, V. N. Tiwari, and R. Narayanan, "Cuff-less PPG based continuous blood pressure monitoring — A smartphone based approach," in *Proc. 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2016, pp. 607–610.

[38] S. Chen, Z. Ji, H. Wu, and Y. Xu, "A non-invasive continuous blood pressure estimation approach based on machine learning," *Sensors*, vol. 19, 2019, Art. no. 2585.

[39] G. D. Clifford, D. J. Scott, and M. Villarroel, "User guide and documentation for the MIMIC II database," *MIMIC-II Database Version*, vol. 2, pp. 1–110, 2009.

[40] J. Lee, S. Yang, S. Lee, and H. C. Kim, "Analysis of pulse arrival time as an indicator of blood pressure in a large surgical biosignal database: recommendations for developing ubiquitous blood pressure monitoring methods," *J. Clin. Med.*, vol. 8, 2019, Art. no. 1773.

[41] H.-C. Lee and C.-W. Jung, "Vital Recorder-a free research tool for automatic recording of high-resolution time-synchronised physiological data from multiple anaesthesia devices," *Sci. Rep.*, vol. 8, pp. 1527–1527, 2018.

[42] J. M. Łęski and N. Henzel, "ECG baseline wander and powerline interference reduction using nonlinear filter bank," *Signal Process.*, vol. 85, pp. 781–793, 2005.

[43] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Trans. Biomed. Eng.*, vol. BME-32, no. 3, pp. 230–236, Mar. 1985.

[44] N. Gaddum, J. Alastruey, P. Beerbaum, P. Chowienczyk, and T. Schaeffter, "A technical assessment of pulse wave velocity algorithms applied to non-invasive arterial waveforms," *Ann. Biomed. Eng.*, vol. 41, pp. 2617–2629, 2013.

[45] X. Ding and Y. Zhang, "Photoplethysmogram intensity ratio: A potential indicator for improving the accuracy of PTT-based cuffless blood pressure estimation," in *Proc. 37th Annu. Int. Conf. IEEE Eng. Medicine Biol. Soc.*, 2015, pp. 398–401.

[46] S. R. Alty, N. Angarita-Jaimes, S. C. Millasseau, and P. J. Chowienczyk, "Predicting arterial stiffness from the digital volume pulse waveform," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 12, pp. 2268–2275, Dec. 2007.

[47] Y. S. Putyatina, "Measurement of arterial blood pressure by processing pulse wave data," in *Proc. 3rd Annu. Siberian Russian Workshop Electron Dev. Mater.*, 2002, pp. 77–78.

[48] L. Zhichang, Z. Song, Y. Wenming, and Y. Zibin, "A research on characteristic information of pulse wave," *J. Beijing Polytech. Univ.*, vol. 22, pp. 71–79, 1996.

[49] A. A. Awad *et al.*, "The relationship between the photoplethysmographic waveform and systemic vascular resistance," *J. Clin. Monit. Comput.*, vol. 21, pp. 365–372, Dec., 2007.

[50] M. Elgendi, "On the analysis of fingertip photoplethysmogram signals," *Current Cardiol. Rev.*, vol. 8, pp. 14–25, 2012.

[51] K. Takazawa *et al.*, "Assessment of vasoactive agents and vascular aging by the second derivative of photoplethysmogram waveform," *Hypertension*, vol. 32, pp. 365–370, 1998.

[52] A. John, "Photoplethysmography and its application in clinical physiological measurement," *Physiological Meas.*, vol. 28, 2007, Art. no. R1.

[53] M. Sherebrin and R. Sherebrin, "Frequency analysis of the peripheral pulse wave detected in the finger with a photoplethysmograph," *IEEE Trans. Biomed. Eng.*, vol. 37, no. 3, pp. 313–317, Mar. 1990.

[54] S. Millasseau, R. Kelly, J. Ritter, and P. Chowienczyk, "Determination of age-related increases in large artery stiffness by digital pulse contour analysis," *Clin. Sci.*, vol. 103, pp. 371–377, 2002.

[55] M. Kachuee, M. M. Kiani, H. Mohammadzade, and M. Shabany, "Cuffless blood pressure estimation algorithms for continuous health-care monitoring," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 4, pp. 859–869, Apr. 2017.

[56] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, 2003.

[57] U. M. Khaire and R. Dhanalakshmi, "Stability of feature selection algorithm: A review," *J. King Saud Univ. - Comput. Inf. Sci.*, 2019, to be published, doi: 10.1016/j.jksuci.2019.06.012.

[58] B. C. Ross, "Mutual information between discrete and continuous data sets," *PLOS One*, vol. 9, no. 2, 2014, Art. no. e87357.

[59] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Mach. Learn.*, vol. 46, pp. 389–422, 2002.

[60] D. W. Marquardt and R. D. Snee, "Ridge regression in practice," *Amer. Statistician*, vol. 29, pp. 3–20, 1975.

[61] N. Meinshausen and P. Bühlmann, "Stability selection," *J. Roy. Stat. Soc.: Ser. B (Stat. Methodol.)*, vol. 72, pp. 417–473, 2010.

[62] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. Newton, MA, USA: O'Reilly Media, 2019.

[63] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005.

[64] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Mach. Learn.*, vol. 63, pp. 3–42, 2006.

[65] R. J. Williams and D. Zipser, "A learning algorithm for continually running fully recurrent neural networks," *Neural Comput.*, vol. 1, pp. 270–280, 1989.

[66] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, pp. 1735–1780, 1997.

[67] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Trans. Signal Process.*, vol. 45, no. 11, pp. 2673–2681, Nov. 1997.

[68] Association for the Advancement of Medical Instrumentation, "American national standards for electronic or automated sphygmomanometers," in *Proc. ANSI/AAMI SP*, 1987.

[69] E. O'Brien *et al.*, "The British hypertension society protocol for the evaluation of blood pressure measuring devices," *J. Hypertension*, vol. 11, pp. S43–S62, 1993.

[70] I. S. Association, "IEEE standard for wearable cuffless blood pressure measuring devices," IEEE Std., pp. 1708–2014, 2014.

[71] A. E. W. Johnson *et al.*, "MIMIC-III, a freely accessible critical care database," *Sci. Data*, vol. 3, 2016, Art. no. 160035.

[72] S. Li, W. Li, C. Cook, C. Zhu, and Y. Gao, "Independently recurrent neural network (indrnn): Building a longer and deeper RNN," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2018, pp. 5457–5466.

[73] S. P. Linder, S. M. Wendelken, E. Wei, and S. P. McGrath, "Using the morphology of photoplethysmogram peaks to detect changes in posture," *J. Clin. Monit. Comput.*, vol. 20, pp. 151–158, 2006.