2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Continuous Blood Pressure Estimation From Non-Invasive Measurements Using Support Vector Regression

Solmaz Rastegar A, Hamid GholamHosseini A, *Senior Member, IEEE,* Andrew Lowe A, *Member, IEEE*, and Maria Lindén B, *Member, IEEE*

Abstract— **Blood pressure (BP) is one of the most crucial vital signs of the human body that can be assessed as a critical risk factor for severe health conditions such as cardiovascular diseases (CVD) and hypertension. An accurate, continuous, and cuff-less BP monitoring technique could help clinicians improve the prevention, detection, and diagnosis of hypertension and manage related treatment plans. Notably, the complex and dynamic nature of the cardiovascular system necessitates that any BP monitoring system could benefit from an intelligent technology that can extract and analyze compelling BP features. In this study, a support vector regression (SVR) model was developed to estimate systolic blood pressure (SBP) and diastolic blood pressure (DBP) continuously. We selected a set of features commonly used in previous studies to train the proposed SVR model. A total of 120 patients with available ECG, PPG, DBP and SBP data were chosen from the Medical Information Mart for Intensive Care (MIMIC III) dataset to validate the proposed model. The results showed that the average root mean square error (RMSE) of 2.37 mmHg and 4.18 mmHg were achieved for SBP and DBP, respectively.**

I. Introduction

Blood pressure (BP) is a critical physiological parameter of the human body that can be one of the most important risk indicators for hypertension (high BP) and cardiovascular diseases (CVD)[1]. CVD is dysfunctions of the heart and blood vessels and includes hypertension, cardiac arrhythmia, cardiac ischemia, and stroke. CVD is the primary cause of universal death and is the leading cause of damage to arteries in organs such as the heart, brain, eyes, and kidneys [2].

High BP or hypertension is the single most crucial adjustable risk factor for CVD, and arterial blood pressure (ABP) is an efficient way to detect and control CVD [3, 4]. The physical condition, physiological rhythm, environmental conditions, and so many other factors could vary the BP over time. By collecting activities of daily living (ADL) and BP variations, there is a possibility of improving the assessment of a patient's hypertension state.

Fortunately, most of the problems caused by CVD are preventable and treatable. Continuous and cuff-less BP monitoring techniques would help reduce CVD risk through early detection and maintaining control at earlier stages of the disease. Moreover, early detection of hypertension could dramatically decrease the likelihood of disability and mortality and reduce treatment costs.

With recent revolutionary developments in machine learning, researchers have recognized its remarkable potential in the healthcare industry to improve people wellbeing [5, 6]. The support vector machines (SVM), linear regression, regression trees, model trees, the ensemble of trees, and random forest are few machine learning algorithms that are considered feasible approaches for cuff-less and continuous BP monitoring [7, 8]. Moreover, the ability of machine learning to learn the function of the complex system makes it a promising method for BP estimation. The main idea is to use machine learning to extract surrogate cardiovascular features from time-domain or frequency-domain of physiological signals, train the machine learning-based model, and estimate BP through the developed model [9].

This study aims to develop an accurate continuous BP estimation model using feature extraction techniques and a machine learning model.

II. Methodology

A. Database

The data used in this study was obtained from 120 patients selected from the Medical Information Mart for Intensive Care (MIMIC-III) Waveform Database Matched Subset [10], a subset of the MIMIC-III waveform Database. For each subject, 420 samples of ECG, PPG, SBP, and DBP individually were selected, with a total of 50400 samples for each signal (equal to 14 hours of training data for each signal).

B. Features Selection and Feature Extraction

ECG and PPG signals can be employed to calculate BP. Therefore, we investigated how to enhance the accuracy of the BP estimation model using related features from ECG and PPG signals. Based on the best performance reported by recent studies, a total of 11 BP-related features associated with heart rate (HR), pulse transit time (PTT), and characteristics of PPG waveforms were carefully chosen from those studies. The definition of all selected features is given in Table I. Two different PTT features were considered by calculating the time between the R-peak of the ECG signal and the specific point

S. R. is with the Electrical & Electronics Engineering Department, Auckland University of Technology, Auckland 1010, New Zealand (e-mail: solmaz.mansouri@aut.ac.nz).

G. H. Associate Professor is with the Electrical & Electronics Engineering Department, Auckland University of Technology, Auckland, New Zealand (e-mail: hgholamh@aut.ac.nz).

A. L. is with the Institute of Biomedical Technologies, Auckland University of Technology, Auckland 1010, New Zealand (e-mail: andrew.lowe@aut.ac.nz).

M. L. is with Division of Intelligent Future Technologies, Mälardalen University, Box 883, 721 23 Västerås, Sweden, (e-mail: maria.linden@mdh.se).

of PPG evaluated in each cardiac cycle.

TABLE I. DEFINITIONS OF SELECTED FEATURES.

Features	Definition
f_1: HR	The time distance between two R_R intervals
f_2: PTT peak	The time distance between ECG R peak and the peak of simultaneously PPG
f_3: PTT foot	The time distance between ECG R peak and the bottom of simultaneously PPG
f_4: PIR	The ratio of PPG peak to PPG bottom
f_5: T1	The time delay between the systolic and diastolic peak
f_6: DT	Diastolic time
f_7: ST	Systolic time
f_8: T_s	Time from cycle start to systolic PPG peak
f_9: T_d	Time from systolic PPG peak to cycle end
f_{10}: AI	Augmentation Index
f_{11}: LAF	Large artery stiffness

As illustrated in Figure 1, the PTT peak was defined as the distance between the R-peak of the ECG signal and the PPG peak, while the PTT foot was defined as the distance between the R-peak of the ECG signal and the PPG foot point. The maxima and minima detection routines were implemented for each cardiac cycle to locate the PPG systolic peaks and feet, as shown in Fig. 1.

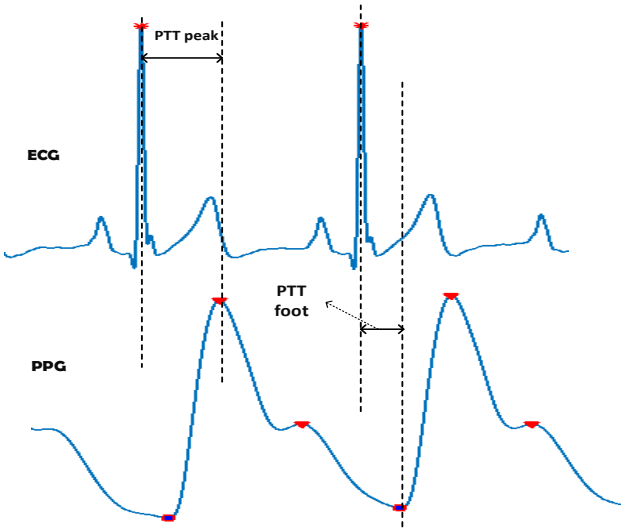

Figure 1. The PTT peak is the distance between the R-peak of the ECG and the PPG peak, and the PTT foot is the distance between the R-peak of the ECG and the PPG foot.

The effective features of the PPG signal for estimating SBP and DBP that were used in this study include; systolic time (ST), diastolic time (DT), large artery stiffness index (LAF), PPG peak intensity ratio (PIR), augmentation index (AI), and the time delay between the systolic and diastolic peak (T1) tracted. The extracted features are defined below:

- Systolic Time (ST), as shown in Fig. 2, ST is described as a relative time from PPG foot to PPG peak, presenting the cardiac output changes [11].
- Diastolic Time (DT), as shown in Fig 2., DT is defined as the relative time for PPG peak to PPG foot which changes with peripheral resistance [11].
- Large Artery Stiffness index (LAF) is the time interval between the systolic peak and diastolic peak, which measures arterial stiffness [12].
- PIR is the ratio of PPG peak intensity to PPG bottom intensity, reflecting arterial vasomotion [13]. The arterial diameter change Δd is theoretically reflected by PIR during one cardiac cycle from systole to diastole point. The PIR is defined as:

$$PIR = e^{\alpha \Delta d} \qquad (1)$$

where α is a constant related to the optical absorption coefficients in the light path.

- Augmentation Index (AI), as illustrated in Fig 2., is defined as the ratio of the diastolic peak and systolic peak, which measure the arteries wave reflection as following [12].

$$AI = \frac{x}{y} \qquad (2)$$

- The time difference between the systolic and diastolic peaks (T1), as shown in Fig 2. The absolute maximum of each PPG wave is the systolic peak, and the diastolic peak is the relative one.

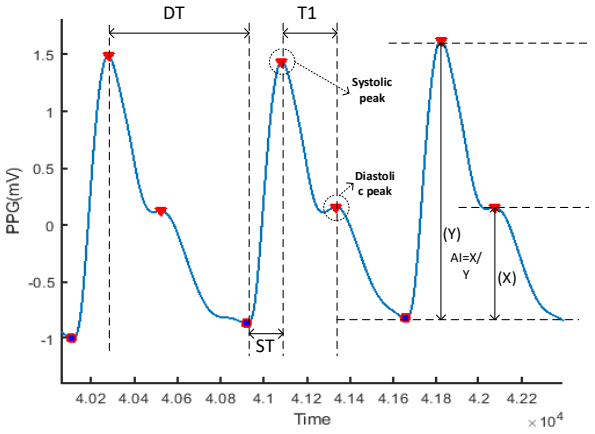

Figure II2. The schematic diagram of PPG Characteristic features.

III. SVR MODEL

Support vector machine (SVM) encompasses a range of different models divided into linear not-separable vector machines, linear separable SVMs, and non-linear SVMs. Using the Kernel function, the non-linear SVM converts linearly non-separable problems into linearly separable problems with high-dimensional space. Thus, the SVM model could be used for both classifications and regression networks.

The version of SVM that use for the regression is called support vector machine regression (SVR) [14]. The ability of this model to deal with the non-linear relationship of both input and output data makes it suitable to address the non-linear relation between the features extracted from ECG and PPG signals and the actual BP [15].

A. Kernel Function Selection

In some cases, the data is not linearly separable in the original, so the kernel function is used to map the projection of

input space to a higher dimension where a linear separation is feasible. Therefore, support vector machine models can use the kernel function $K(x, x')$ to establish a non-linear support machine. The polynomial kernel functions, Gaussian Radial Basis Function (RBF), and sigmoid kernel function are the most common kernel functions [5].

In this paper, the RBF kernel was implemented in the SVR model. The original feature space $x = (PTT, HR, \ldots)$ was mapped onto the new feature space $x' = (x_1, x_2, x_3, \ldots x_n)$. Therefore, the new set of the BP indicator data was expressed by a linear regression formula in the feature space to determine a non-linear mapping model between ECG and PPG signals and BP.

B. Description of Proposed SVR Model

The non-linear mapping $\varphi(x)$ mapped the input sample x into a high-dimensional feature space and then estimated the regression function through a linear model built in this feature space. The main idea was to use non-linear mapping to map the input space onto a high-dimensional feature space. The non-linear model is shown as follow:

$$f(x, \omega) = \omega \cdot \varphi(x) + b \qquad (3)$$

where $x = (PTT, HR, \ldots), \omega$ is the weight vector, and b is the threshold. The ω and b can be obtained as follow:

$$\min \frac{1}{2} \parallel \omega \parallel^2 + c \sum_{i=1}^{1} (\xi_i + \xi_i^*), \qquad (4)$$

is subject to:

$$y_i - (\omega^T x_i + b) < \varepsilon + \xi_i$$
$$(\omega^T x_i + b) - y_i < \varepsilon + \xi_i^*$$
$$\xi_i + \xi_i^* \geq 0$$

where C is a penalty factor, ε is the loss function, and ξ_i and ξ_i^* are different relaxation factors. The solution of (3) is as follows:

$$f(x) = \sum_i^l (-\alpha_i + \alpha_i^*) K(x_i, x) + b, \qquad (5)$$

where α_i and α_i^* are Lagrange multipliers, l is the number of SVs, and $K(x_i, x)$ is a kernel function.

Unlike the linear kernel, the RBF kernel transforms the database into a non-linear high dimensional space, making it possible to overcome the non-linear relationship between features and BP. Also, compare with the polynomial kernel, it has less model complexity due to fewer tuning parameters. The RBF is defined as:

$$K(x_i, x) = \exp(-\gamma \parallel x - x_i \parallel^2) \qquad (6)$$

where γ is the kernel parameter. The kernel parameter, gamma (γ), can adjust the influence of a training sample. The larger value will decrease under the influence of the training sample.

C. SVR BP Model

The HR, PTT, and PPG features were extracted and combined using the dataset. These features were used to train and test the SVR model. To evaluate the SVR method, the database was split into training and testing; 70% of the subjects were selected for training and 30% of the rest for the testing dataset.

To establish an SVR model to predict the SBP and DBP using the optimal parameters C and γ, the parameters were set to (100, 10) for C , (100, 10) for γ , and [0.01,1] for ε. These values were chosen based on the promising results achieved by Zhang et al. [16], which were achieved by the optimization function created on 10-fold cross-validation. The proposed SVR model was implemented in the MATLAB environment.

The details of the SBP and DBP prediction process using the SVR model is shown in Fig 3. After extraction and combination of features, the data was divided into training and test set. Then the SVR model was established based on the optimal parameter of RBF kernel (C, γ, ε), and then the training dataset was used to train the SVR model. Finally, the testing dataset was fed to a trained network in the next step to estimate the SBP and DBP.

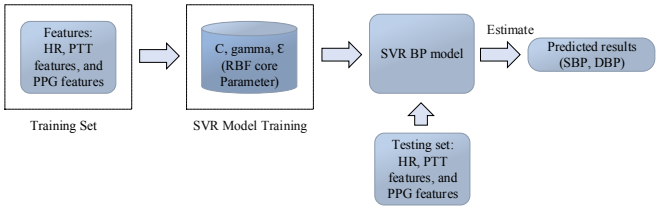

Figure 3. Schematic diagram of SBP and DBP estimation using SVR model.

IV. Experimental Results and Comparison

The prediction results of the SVR model using hand-engineered features were evaluated in terms of accuracy and RMSE. The accuracy was defined as:

$$\text{Accuracy} = 100 * \frac{\text{Number of correct estimation}}{\text{Total number of testing data}} \qquad (7)$$

where the number of correct estimations was calculated as the number within the acceptable error margin ≤ 5mmHg set by AAMI. The experimental results show that the SVR model reaches an SBP estimation accuracy of 93.52 % and DSB estimation accuracy of 90.12 %.

Moreover, the RMSE obtained from (8) was used as an indicator for evaluating the performance of the estimation model.

$$\text{RMSE} = \sqrt{\frac{\sum(x_i - y_i)^2}{n}} \qquad (8)$$

where x_i is the estimated BP, y_i is the actual value, and n is the total number of values in the testing dataset. The RMSE of 2.37 mmHg and 4.18 mmHg for SBP and DBP were calculated, respectively.

Table II summarizes a comparison between the SVR-based model employed in this study and previously established studies that used the machine learning model to estimate BP. To establish a reasonable comparison, only the studies utilizing RSME metrics were chosen for comparison. Nidigattu et al. [17] employed extensive signal processing and feature engineering. They employed three types of machine learning methods to estimate the BP and calculated the RMSE for SBP. They used data related to 140 individual healthy

subjects of just over three minutes as their database. The database is divided into training and test set based on the number of subjects. To determine the specific scale factors of the pulse wave analysis (PWA) based BP estimation model, Yoon et al. [18] employed pulse arrival time (PAT) as a BP related parameter. A linear regression model was designed to measure SBP and DBP. The experimental results achieved lower error but used only 23 subjects to evaluate the results.

TABLE II. COMPARISON OF THE PROPOSED METHOD WITH WELL-ESTABLISHED RELATED WORK

Model	Number of subjects	RMSE for SBP(mmHg)	RSME for DBP(mmHg)
Three types of machine learning [17]	140 healthy subjects	5.86, 9.50, 10.30	5.30, 8.30, 8.25
PWA and PAT [18]	MIMIC, 23 subjects	10.6	Yes
The proposed method	**120**	**2.37**	**4.18**

The SVR BP model proposed in the study was estimated with higher accuracy and least error while avoiding overlapping the subjects of the training set with that of the test set by dividing them based on the number of subjects. This implies that the proposed model notably outperforms the other models.

V. CONCLUSION

The main contribution of this study is the investigation of the association between BP and human physiological index data by using the SVR algorithm. The SVR model has reconstructed the non-linear relationship between the extracted features and BP. To evaluate the accuracy and efficiency of the proposed method, the physiological signals (ECG and PPG) that collected by medical devices in the hospital was used in this study. , Within a relative error range of ±5 mmHg (according to the American ANSI/AAMI SP10-1992 standard specified error range), the proposed SVR model achieved an SBP estimation accuracy of 93.52 % and DBP estimation accuracy of 90.12 %, respectively. The SVR model proposed in the study was estimated with higher accuracy and less error while avoiding overlapping the subjects of the training set with that of the test set by dividing them based on the number of subjects.

Future research will consider expanding the clinical applicability of the model by adding more physiological, clinical, and demographic data, including age, sex, respiratory rate, diagnosed disease, body temperature, age, and weight. In addition, the model with more indicators could improve and enrich current health care provisions by discovering the relationship between BP related diseases and other diseases.

REFERENCES

[1] M. J. Banet, Z. Zhou, M. S. Dhillon, R. J. Kopotic, A. S. Terry, and I. H. VISSER, "Vital sign monitor for measuring blood pressure using optical, electrical and pressure waveforms," ed: Google Patents, 2013.

[2] G. Cohuet and H. Struijker-Boudier, "Mechanisms of target organ damage caused by hypertension: therapeutic potential," *Pharmacology & Therapeutics,* vol. 111, no. 1, pp. 81-98, 2006.

[3] S. Ahmad, M. Bolic, H. Dajani, V. Groza, I. Batkin, and S. Rajan, "Measurement of heart rate variability using an oscillometric blood pressure monitor," *IEEE Transactions on Instrumentation and Measurement,* vol. 59, no. 10, pp. 2575-2590, 2010.

[4] S. Ahmad *et al.*, "A Prototype of An Integrated Blood Pressure and Electrocardiogram Device for Multi-Parameter Physiologic Monitoring," in *IEEE Instrumentation and Measurement Technology Conference (I2MTC)*, 2010: IEEE, pp. 1244-1249.

[5] M. Kachuee, M. M. Kiani, H. Mohammadzade, and M. Shabany, "Cuff-Less High-Accuracy Calibration-Free Blood Pressure Estimation Using Pulse Transit Time," in *IEEE International Symposium on Circuits and Systems (ISCAS)*, 2015: IEEE, pp. 1006-1009.

[6] X. Xing and M. Sun, "Optical blood pressure estimation with photoplethysmography and FFT-based neural networks," *Biomedical Optics Express,* vol. 7, no. 8, pp. 3007-3020, 2016.

[7] S. S. Mousavi, M. Hemmati, M. Charmi, M. Moghadam, M. Firouzmand, and Y. Ghorbani, "Cuff-less blood pressure estimation using only the ECG signal in frequency domain," in *8th International Conference on Computer and Knowledge Engineering (ICCKE)*, 2018: IEEE, pp. 147-152.

[8] S. S. Mousavi, M. Hemmati, M. Charmi, M. Moghadam, M. Firouzmand, and Y. Ghorbani, "Cuff-Less blood pressure estimation using only the ECG signal in frequency domain," in *2018 8th International Conference on Computer and Knowledge Engineering (ICCKE)*, 2018: IEEE, pp. 147-152.

[9] F. Miao *et al.*, "A Novel Continuous Blood Pressure Estimation Approach Based on Data Mining Techniques," *IEEE Journal of Biomedical and Health Informatics,* vol. 21, no. 6, pp. 1730-1740, 2017.

[10] A. E. Johnson *et al.*, "MIMIC-III, a freely accessible critical care database," *Scientific Data,* vol. 3, no. 1, pp. 1-9, 2016.

[11] M. Elgendi, "On the analysis of fingertip photoplethysmogram signals," *Current Cardiology Reviews,* vol. 8, no. 1, pp. 14-25, 2012.

[12] M. Kachuee, M. M. Kiani, H. Mohammadzade, and M. Shabany, "Cuff-less high-accuracy calibration-free blood pressure estimation using pulse transit time," in *Circuits and Systems (ISCAS), 2015 IEEE International Symposium on*, 2015: IEEE, pp. 1006-1009.

[13] 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 Transactions on Biomedical Engineering,* vol. 63, no. 5, pp. 964-972, 2016.

[14] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and Computing,* vol. 14, no. 3, pp. 199-222, 2004.

[15] I. A. Naguib, "Improved predictions of non-linear support vector regression and artificial neural network models via preprocessing of data with orthogonal projection to latent structures: A case study," *Bulletin of Faculty of Pharmacy, Cairo University,* vol. 55, no. 2, pp. 287-291, 2017.

[16] B. Zhang, H. Ren, G. Huang, Y. Cheng, and C. Hu, "Predicting blood pressure from physiological index data using the SVR algorithm," *BMC Bioinformatics,* vol. 20, no. 1, p. 109, 2019.

[17] G. R. Nidigattu, G. Mattela, and S. Jana, "Non-invasive modeling of heart rate and blood pressure from a photoplethysmography by using machine learning techniques," in *International Conference on COMmunication Systems & NETworkS (COMSNETS)*, 2020, pp. 7-12.

[18] Y.-Z. Yoon *et al.*, "Cuff-less blood pressure estimation using pulse waveform analysis and pulse arrival time," *IEEE Journal of Biomedical and Health Informatics,* vol. 22, no. 4, pp. 1068-1074, 2018.