

# Blood pressure estimation with complexity features from electrocardiogram and photoplethysmogram signals

Sen Yang , Wan Suhaimizan Wan Zaki, Stephen P. Morgan, Siu-Yeung Cho, Ricardo Correia, Yaping Zhang



**University of  
Nottingham**

UK | CHINA | MALAYSIA

Faculty of Science and Engineering, University of Nottingham Ningbo  
China, 199 Taikang East Road, Ningbo, 315100, Zhejiang, China.

First published 2020

This work is made available under the terms of the Creative Commons  
Attribution 4.0 International License:

<http://creativecommons.org/licenses/by/4.0>

The work is licenced to the University of Nottingham Ningbo China  
under the Global University Publication Licence:

<https://www.nottingham.edu.cn/en/library/documents/research-support/global-university-publications-licence.pdf>



**University of  
Nottingham**

UK | CHINA | MALAYSIA

# Blood Pressure Estimation with Complexity Features from Electrocardiogram and Photoplethysmogram Signals

*SenYang<sup>1</sup>, Wan Suhaimizan Wan Zaki<sup>2</sup>, Stephen P. Morgan<sup>2</sup>, Siu-Yeung Cho<sup>1</sup>, Ricardo Correia<sup>2</sup>, Yaping Zhang<sup>1\*</sup>*

<sup>1</sup>International Doctoral Innovation Centre, University of Nottingham Ningbo China, 199 Taikang East Road, Ningbo, China

<sup>2</sup>Optics and Photonics Research Group, University of Nottingham, Nottingham, United Kingdom

\* Corresponding author: yaping.zhang@nottingham.edu.cn

## Abstract

A novel method for the continual, cuff-less estimation of the systolic blood pressure (SBP) and diastolic blood pressure (DBP) values based on signal complexity analysis of the photoplethysmogram (PPG) and the electrocardiogram (ECG) is reported. The proposed framework estimates the blood pressure (BP) values obtained from signals generated from 14 volunteers subjected to a series of exercise routines. Herein, the physiological signals were first pre-processed, followed by the extraction of complexity features from both the PPG and ECG. Subsequently the complexity features were used in regression models (artificial neural network (ANN), support vector machine (SVM) and LASSO) to predict the BP. The performance of the approach was evaluated by calculating the mean absolute error (MAE) and the standard deviation (STD) of the predicted results and compared with the recommendations made by the British Hypertension Society (BHS) and Association for the Advancement of Medical Instrumentation (AAMI). Complexity features from the ECG and PPG were investigated independently, along with the combined dataset. It was observed that the complexity features obtained from the combination of ECG and PPG signals resulted to an improved estimation accuracy for the BP. The most accurate DBP result of  $5.15 \pm 6.46$  mmHg was obtained from ANN model, and SVM generated the most accurate prediction for the SBP which was estimated as  $7.33 \pm 9.53$  mmHg. Results for DBP fall within recommended performance of the BHS but SBP is outside the range. Although initial results are promising, further improvements are required before the potential of this approach is fully realised.

**Key words** blood pressure (BP) · complexity analysis · photoplethysmogram (PPG) · electrocardiogram (ECG) · machine learning

## 1 Introduction

Unhealthy lifestyle choices and dietary habits have led to the prevalence of rising blood pressure (BP) which could result in hypertension and increase the risks of cardiovascular diseases. One-third of global deaths is related to cardiovascular diseases of which hypertension has become a highly prevalent risk factor and a worldwide health problem (Organization, Group 2003). One important method of preventing hypertension both clinically and through self-administration is the continuous monitoring of BP, and with the ubiquity of smartphones and wearable devices, these procedures could make use of the internet of things (IoT) to make healthcare delivery more effective (Qi et al. 2017). Thus, the increasing need for optimal mobile and user-friendly diagnostic tools for monitoring ailments such as hypertension has

promoted the development and adoption of non-invasive cuff-less continuous BP measurement and methods.

Most research studies aimed at developing ambulatory and continual BP monitoring systems adopt a combination of electrocardiogram (ECG) and photoplethysmogram (PPG) (Buxi et al. 2015; Mukkamala et al. 2015; Sun et al. 2016; Zaki et al. 2016; Zhang et al. 2017). Pulse wave velocity (PWV), pulse arrival time (PAT) and pulse transit time (PTT) all rely on the theory that pressure waves propagating along arterial vessels are related to the vessel wall properties. In general, PTT and BP values have been parameterised through linear, logarithmic or exponential relationships (Sharma et al. 2017). In principle, PAT-based approaches for cuff-less BP monitoring require two simultaneously collected physiological signals (e.g. ECG and PPG) (Mukkamala et al. 2015). Moreover, the measurement and collection of two physiological signals requires more complex instrumentation, which are susceptible to measurement errors. For instance, in the detailed procedures to estimate BP by PAT using PPG and ECG signals, the calculated systolic BP (SBP) and diastolic BP (DBP) errors for normal and healthy patients were deduced to be in the range -18.3 to 14.8 mmHg and -12.1 to 8.2 mmHg respectively (Goli, Jayanthi 2014). However, this wide range of errors might not be effective for diagnosis in practice.

The additional complexity of measuring 2 physiological signals simultaneously can be mitigated by estimating BP values based on either ECG or PPG alone. BP estimation can be facilitated using machine learning and artificial intelligence. In practice, the collection of ECG signal usually involves the application of electrodes to the body which is inconvenient for continuous measurement, as such most investigators usually utilise PPG signals to estimate BP mainly because of the design simplicity and ease of use (Wang et al. 2018; Addison 2016; Suzuki 2015; Visvanathan et al. 2013). This technique exploits various machine learning and signal processing algorithms to estimate BP. Herein, features related to the BP which are obtained after pre-processing the PPG signal are fed into machine learning models which subsequently generates an estimate of the BP values (Kachuee et al. 2016). Recently, a method for estimating SBP and DBP using features extracted solely from the PPG signal and an artificial neural network (ANN) computed a mean absolute error (MAE) of  $4.20 \pm 2.79$  mmHg and  $2.27 \pm 1.82$  mmHg for SBP and DBP respectively (Wang et al. 2018).

It is important to note that although the use of PPG signals to predict BP is simple and convenient, experimental artefacts introduced from motion and ambient noise usually affects the accuracy. Thus, other studies have also investigated methods to estimate BP using only the ECG (Simjanoska et al. 2018; Yang et al. 2017; Wu et al. 2016; Monroy Estrada et al. 2014). In an earlier report, a relationship between BP and ECG was demonstrated by resolving the systolic and diastolic portions of the ECG signals and the obtained average error were 2.70% for systole and 4.35% for diastole (Monroy Estrada et al. 2014). Also, the morphological features of ECG signals have been used to calculate BP values through the use of ANN models. In this research, the average error rate of the BP measurement is lower than 5% (1.96%/2.14% which is approximately  $2.38 \pm 3.66$  mmHg/ $1.57 \pm 4.17$  mmHg for SBP/DBP) when compared with a mercury sphygmomanometer (Wu et al. 2016). Hitherto, different from utilising the morphology of the ECG signals, a recent report obtained BP values by combining complexity analysis of the ECG with machine learning models. In this case, the MAE for the SBP and DBP were respectively obtained as 7.72 mmHg and 9.45 mmHg whilst a value of 8.13 mmHg was obtained for the mean arterial pressure (Simjanoska et al. 2018).

The research reported in this paper is based on the extraction of different complexity features from both PPG and ECG signals such as: autoregressive (AR) coefficients, and Hjorth parameters. Some of the features adopted in this paper have been used for recognition of heart

rhythm, or the measurement of heart rate variability during myocardial ischaemia (Zhao,Zhang 2005; Leonarduzzi et al. 2010; Li,Zhou 2016) but not for BP estimation. These parameters are then used to create a regression model such as an ANN or SVM in order to estimate BP. This novel approach is in contrast to earlier studies which utilised the PTT or morphological features of the ECG and/or PPG to predict BP (Zhang et al. 2017; Zaki et al. 2016; Kachuee et al. 2016; Millasseau et al. 2006). This paper is to our knowledge the first use of complexity features from PPG signals to determine BP, and also the first to compare this approach and ECG based methods.

## 2 Background Theory

The following subsections describe the parameters extracted to develop the BP estimation methods.

### 2.1 PTT

BP is the pressure created from circulating blood on vessel walls that aids the flow of blood from the ventricle to the aorta and peripheral vessels. The pressure wave along the arterial vessels and its speed depend on arterial properties such as: elasticity, stiffness, thickness of the artery wall, and size of the artery. PTT is related to the propagation speed of the pressure wave, which can be explained by the Moens-Korteweg equation (Mukkamala et al. 2015):

$$PWV = \frac{D}{PTT} = \sqrt{\frac{Eh}{2r\rho}} \quad (1)$$

where  $PWV$  is the velocity of the pulse wave (m/s),  $D$  is the length of the vessel (m),  $PTT$  is the pulse transit time (s),  $E$  is Young's modulus of the arterial wall (Pa),  $\rho$  is the blood density ( $\text{kg/m}^3$ ),  $h$  is the thickness of the vessel wall (m), and  $r$  is the radius of vessel (m).

Young's modulus of the arterial wall  $E$  is not a constant value, which follows the empirical exponential relationship with the fluid pressure  $P$  according to (Hughes et al. 1978):

$$E = E_0 e^{\gamma P} \quad (2)$$

where  $E_0$  and  $\gamma$  are subject-specific parameters which depend on the thoracic aorta and abdominal aorta,  $e = 2.718$  is Euler's number and  $P$  is the BP (mmHg) within the vessel. In practice, it is assumed that the change in elasticity is much greater than the change in the radius of vessels or wall thickness. Subsequently, the elasticity increases with the rise of BP, which leads to an increase in the PWV and hence decrease in PTT. Although there are some limitations, the PAT (time delay between the the PPG and ECG signals) is often used as a surrogate for the PTT.

### 2.2 Other features

#### *Autoregressive model coefficients*

AR coefficients are utilised to classify and extract features from signals. Due to the reduced dimensions of feature vectors, the classification can be done quickly. An AR model is a representation of a type of random process and it indicates that the output variable depends linearly on its own previous values and on a stochastic term. An AR model with order  $p$ , and

signal  $x[n]$  at time constant  $n$  can be represented as a linear combination of  $p$  previous values of the same signal. Accordingly, the AR ( $p$ ) model is defined as:

$$x[n] = \sum_{i=1}^p a[i]x[n-i] + e[n] \quad (3)$$

where  $a[i]$  is the  $i$ -th coefficient of the model,  $e[n]$  corresponds to a white noise source with mean value of zero, and  $p$  is an arbitrary order. There are different methods to estimate the coefficients of an AR model, in this work, an 8-order AR model was adopted to represent each heartbeat according to the ARfit method (Zhao,Zhang 2005). Thus, there are 8 features extracted from the AR model coefficients utilised in this research.

#### *Multifractal wavelet leader*

Multifractal analysis can be used to study scaling phenomena and long-term correlations in time series which in turn can be used to quantify the distribution of their singularities. In practice, this can be used to analyse the fluctuations of heart rate variability caused by myocardial ischaemia. The wavelet leader based multifractal formalism is a new approach for multifractal analysis which shows improvement on myocardial ischaemia detection. In this feature extraction, multifractal 1-D wavelet leader estimate is applied to find the Holder exponents and the second cumulant of the scaling exponents. Specifically, the Holder exponents and second cumulant characterizes the signal regularity and departure from linearity respectively (Leonarduzzi et al. 2010).

#### *Shannon Entropy (SE) values*

SE is used to measure the uncertainty of the information content in given systems. Daubechies wavelet with 4 vanishing moments (db4) were found to perform well in the ECG classification (Li,Zhou 2016). In this research, this wavelet is also adopted to extract features from the maximal overlap discrete wavelet packet transform. In terms of decomposition level, the signal cannot be expressed well when the level is too small and a very large one can lead to a large number of features with many values close to zero. Accordingly, the decomposition level is set to range from 4 to 6. It is found that the best results are achieved at level 5 and they are presented in the following sections. Due to the space limitation, the results for decomposition level 4 and 6 are not presented here but they can be provided upon request. SE can be calculated from Equation (4) based on the probability distribution of the energy (Li,Zhou 2016):

$$SE_{i,j} = - \sum_{k=1}^N p_{i,j,k} * \log (p_{i,j,k}) \quad (4)$$

where  $p_{i,j,k}$  is the  $k$ -th coefficient at the  $i$ -th level for the  $j$ -th node. A total of 32 features are extracted from the SE model.

#### *Hjorth parameters*

Signal mobility and signal complexity are parameters computed by using the Hjorth parameters method (Kugiumtzis,Tsimpiris 2010). They are indicators of statistical properties used for signal processing in the time domain. The mobility parameter represents the mean frequency or the proportion of standard deviation of the power spectrum. This is defined as the square root of the variance of the first derivative of the signal divided by its variance. The signal complexity parameter represents the change in frequency. This parameter compares the signal's similarity to a pure sine wave. Signals that are identical to a sine wave have complexity parameter of unity (Simjanoska et al. 2018).

#### *Heart rate*

Heart rate is the number of times a person's heart beats per minute. At rest, the heart rate for adults ranges from 60 to 100 beats per minute (Hori, Okamoto 2012). Heart rate depends on the individual, weight, age, heart conditions, use of medication, and activity state. Heart rate is used as a feature to predict BP.

### **3 Methods**

#### **3.1 Experiments**

Ethical approval was provided by the Ethics Committee, Faculty of Engineering, University of Nottingham.

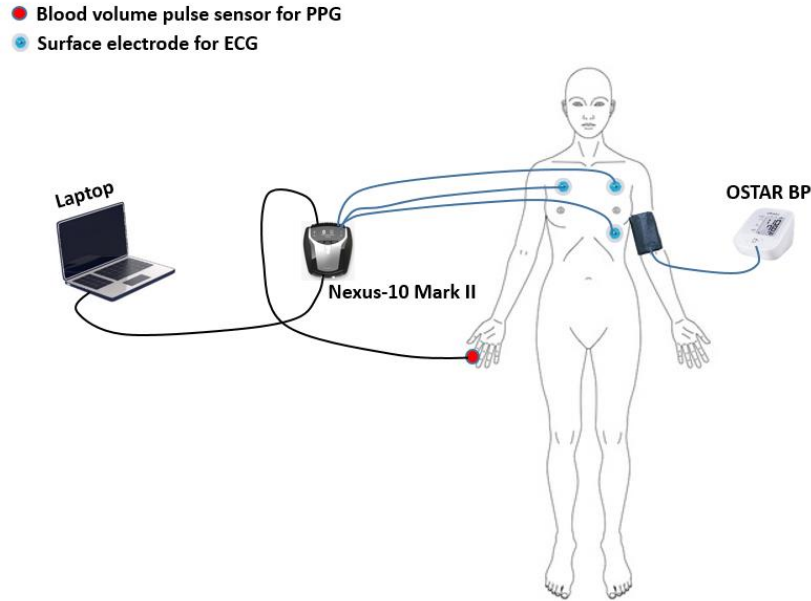
##### **3.1.1 Subjects and devices**

14 male volunteers (age:  $30 \pm 13$  years; height:  $166 \pm 15$  cm; weight:  $62 \pm 13$  kg) participated in this study. All of the subjects are healthy with no clinically apparent arterial disease or physical abnormality. Four hours prior to the experiment, the subject was asked to avoid drinks that contained caffeine or a heavy meal to prevent a large variability in blood pressure.

A data acquisition device (Nexus-10 Mark II) converted the analogue signal into digital form before transmitting to a laptop for further processing via Bluetooth. The device is battery-operated to reduce electrical hazard and has a built-in amplifier with low input noise, high input impedance and high common mode rejection ratio. In this research, one bipolar channel was used for ECG measurement with surface electrodes according to lead II configuration. The PPG signal from the fingertip was measured by a Nexus BVP (NX-BVP1A) sensor which consists of red and infra-red light sources and a photodetector to measure relative blood volume through the blood vessels in the finger.

##### **3.1.2 Data acquisition**

The PPG sensor was clamped to subject's right index finger and ECG electrodes were attached on the chest using adhesive pads (Figure 1). A sphygmomanometer (OSTAR, P2) was placed on the subject's left arm.



**Fig. 1** Sensor placement on subject for PPG and ECG signal measurement

The PPG, ECG and BP were measured simultaneously from subjects at successive intervals between pedalling on an exercise bike, this physical exertion would likely result in an increase in BP as the heart rate increases. The sampling frequency for the data acquisition was 256 samples/second with 24 bit resolution. Each subject was asked to sit on an exercise bike (Domyos VM740) with their hands on the armrests. As PPG signals can be sensitive to motion, the subjects were asked to stay still during the measurement period. Each measurement experiment took 26 minutes and involved three parts:

(I) the subject stayed still for 6 minutes and the PPG, ECG, and BP were recorded every 2 minutes.

(II) the subject cycled at a fixed intensity (108 W with 60 turns per minute) for 1 minute and rested for another minute. This process was repeated 4 times whilst the measurements were taken during the rest periods.

(III) the subject cycled with the same intensity as in (II) for 6 minutes and measurements were recorded every 2 minutes three times after the exercise is completed.

Hence, there are 10 sets of data collected from each subject under different states and all data were monitored and recorded through BioTrace+ software (Mind Media B.V. Netherland). Accordingly, there was a total number of 140 sets of data corresponding to PPG, ECG and BP values that were simultaneously recorded in the course of this experiment.

## 3.2 BP estimation

### 3.2.1 Signal pre-processing

A total of 140 sets of data corresponding to ECG, PPG, SBP and DBP values, were collected from 14 people in various scenarios after different duration of exercise. However, a few datasets were invalid as they either failed to record certain signals (e.g. missing certain BP



values) or included unmanageable outliers. After screening the results, 129 groups from the dataset were retained for the next step of signal pre-processing. These datasets contained complete ECG, PPG, SBP and DBP values.

Since the duration of the signal recordings ranged from 14 seconds to more than 1 minute according to BP acquisition time, 10 seconds of signals are selected for each group of data and there are 129 groups in total. The acquired raw PPG signal was then pre-processed by a Chebyshev II bandpass filter with the lower and upper cut off frequencies of 0.5 and 10 Hz respectively. This procedure was used to remove the noise from the raw PPG signal which in turn increases the signal quality index (SQI) (Liang et al. 2018). The raw ECG signal is processed by a Butterworth bandpass filter with cut off frequencies of 0.5 and 40 Hz to remove the baseline wander, and to suppress spurious data fluctuations due to motion (Shin et al. 2010). Subsequently, the PPG and ECG signals were normalized and their peaks in each period were ascertained. The most stable period of the both signals were chosen by selecting the highest cross-correlation coefficient between cycles, these were defined by neighbouring peaks and were used in the succeeding feature extraction.

### 3.2.2 Feature extraction, portioning and model construction

Table 1 summaries all the features extracted and used for BP estimation in this work. PAT and heart rate were extracted from the peaks of the PPG and ECG signals, and other features were obtained from complexity analysis. The PAT value is calculated from the time interval between the R-peak of the ECG and the systolic peak of the PPG within one heartbeat. Heart rate is calculated from the measurement of the peak-to-peak time interval of the PPG or ECG signals. Apart from the PAT and heart rate, the remaining features were extracted separately from both PPG and ECG signals. Hence, the number of complexity features is doubled, and there are 90 features in total from each group of dataset.

**Table 1** Extracted features from ECG and PPG signals

| Feature name   |                                      | Number of features |
|--|--------------------------------------|--------------------|
| Pulse arrival time (PAT)   |                                      | 1                  |
| Autoregressive (AR) model coefficients of order 8  |                                      | 8                  |
| Multifractal wavelet leader  | Second cumulant of scaling exponents | 1                  |
|  | Holder exponents                     | 1                  |
| Shannon Entropy (SE) values for the maximal overlap discrete wavelet packet transform at level 5 |                                      | 32                 |
| Hjorth parameters  | Signal mobility                      | 1                  |
|  | Signal complexity                    | 1                  |
| Heart rate   |                                      | 1                  |

Due to the limited number of subjects participating in the experiment, in the first stage, samples from all volunteers were merged and there are 129 samples in total. Then they are divided into training and testing set with the ratio of 7:3 following earlier studies. Since the purpose of this

research was to investigate the performance of machine learning models using complexity analysis, three cases were proposed to predict BP using the features from Table 1: 1) ECG features alone; 2) PPG features alone; 3) features from both ECG and PPG as well as PAT.

It is ideal to separate the training and testing samples from the same set of subjects, which is often referred to as inter-patient scheme. However, it is often not practical in this research area because of the limited sample data when the researchers choose to recruit the subjects and collect the data by themselves (Zhang et al. 2019; Li et al. 2018; Islam et al. 2017). Nevertheless, this study also applies the inter-patient scheme and aims to reinforce the results in the first stage. In our proposed work, 5-fold and 3-fold cross validations are employed. With respect to 5-fold cross validation, 14 volunteers are split into 4 folds of 3 subjects and 1 fold of 2 subjects. And in the case of 3-fold cross validation, 14 volunteers are split into 2 folds of 5 subjects and 1 fold of 4 subjects. Each cross validation experiment is repeated 20 times to achieve the final result.

In addition to the comparison between using complexity features from ECG and PPG, this study also presents a comparison between the performances of various learning methods in utilising these features. In this comparison, a simple linear regression model (i.e., Least Absolute Shrinkage and Selection Operator (LASSO) regression model), a strong and nonlinear model (i.e., Support Vector Machine (SVM)) as well as an ANN model were included. LASSO regression was used for both variable selection and regularization in order to improve the prediction accuracy and interpretability of the statistical model it produces (Fu 1998). LASSO regression is included in this research to investigate the performance of linear regression on BP estimation. SVM is a suitable and powerful learning algorithm in terms of generating strong models with minimal training resources and a high noise tolerance (Kachuee et al. 2016). The parameters of SVM, such as kernel function and kernel scale, are optimized using “OptimizeHyperparameters” option of “fitsvm” function in MATLAB in training. The optimization procedure involves using a 5-fold cross validation process for the training data. Although the same features are used for the estimation of SBP and DBP, two separate models were trained for each of these targets. The optimized parameters of SVM for different settings in the first stage are listed as follows:

**Table 2** Kernel function and kernel scale for SVM models

|                      |                | PPG and ECG | PPG only                | ECG only |
|----------------------|----------------|-------------|-------------------------|----------|
| SVM Parameters (SBP) | Kenel Function | Gaussian    | Polynomial with order 3 | Linear   |
|                      | Kernel Scale   | 6.8909      | 2.9690                  | 4.6193   |
|                      |                |             |                         |          |
| SVM Parameters (DBP) | Kenel Function | Gaussian    | Linear                  | Linear   |
|                      | Kernel Scale   | 6.4089      | 3.8306                  | 4.6391   |

ANN is a bio-inspired system based on the structure and functions of biological neural networks, which has been used on a variety of tasks (Maind,Wankar 2014). The structure of ANN is determined empirically, including the number of hidden layers (1 and 2), number of hidden nodes (5, 10 and 15) and training algorithm (Levenberg-Marquardt, Bayesian Regularization and BFGS Quasi-Newton). The results indicate that the combination of 1 hidden layer, 10 hidden nodes and Bayesian Regularization algorithm generate most accurate prediction. The default hyperbolic tangent sigmoid transfer function is used for the hidden layer

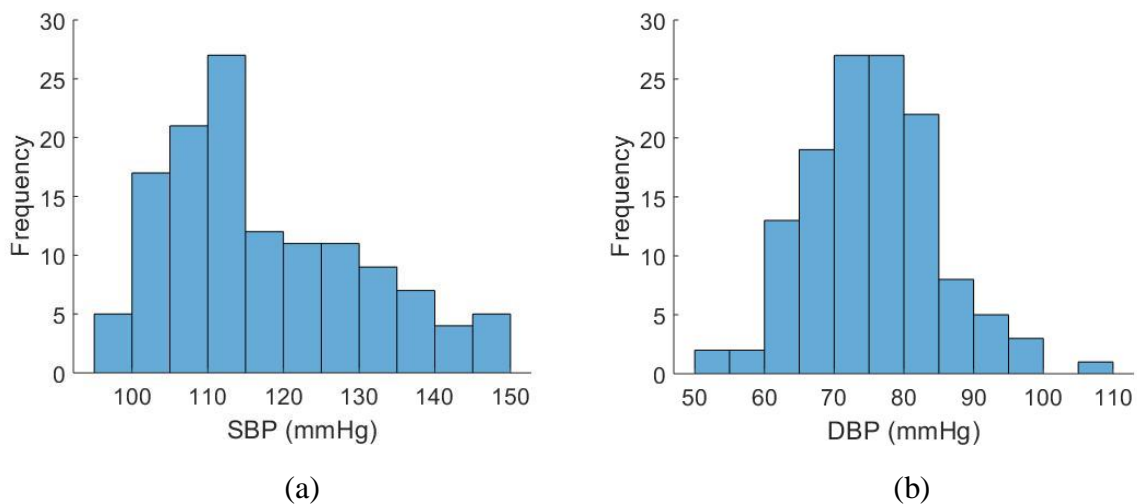
and it can be written as  $f(x) = (\exp(x) - \exp(-x))/(\exp(x) + \exp(-x))$ . A linear transfer function is used for the output layer since the blood pressure is in the form of continuous target values.

## 4 Results

Each model was implemented to predict DBP and SBP values from a dataset separated into training and testing categories in which overlap between categories is avoided. To execute the main tasks in this investigation, features extracted from either PPG or ECG signals (Table 1) were used to predict BP values, then PAT and features from both PPG and ECG signals were utilized for the BP estimation. Subsequently, LASSO, SVM and ANN models were deployed whilst the calculated MAE and STD were then used for model evaluations. The results from the first stage where the data are merged are given in Section 4.2-4.6 while the results of employing inter-patient scheme are given in Section 4.7.

### 4.1 Measured BP distribution

BP histograms obtained from the sphygmomanometer data are presented in Figure 2. The SBP and DBP ranged from 97 mmHg to 148 mmHg and 53 mmHg to 108 mmHg respectively. These relatively large ranges can be attributed to the different levels of exercise which the subjects were made to undertake. Accordingly, the sensitivity of the results reflects the robustness in the prediction of BP values.



**Fig. 2** Histogram of the BP values measured using: (a) SBP; (b) DBP.

Note that the number of features extracted from either ECG or PPG signals was 45, as described in Sections 3.2.2. Relatedly, the number of features extracted from the combination of ECG and PPG signals was 90 (including PAT and heart rate).

## 4.2 Results using features extracted from only the ECG

Table 3 shows the results from the three machine learning models for DBP and SBP estimation. The features used in this section were extracted solely from the ECG signal. The best performance of the DBP and SBP were respectively obtained from SVM and ANN with corresponding values of  $5.48 \pm 7.07$  mmHg and  $8.08 \pm 10.47$  mmHg.

**Table 3** Comparison of the performance between various machine learning models using complexity features from ECG.

|       | Diastolic Blood Pressure (mmHg) |      | Systolic Blood Pressure (mmHg) |       |
|-------|---------------------------------|------|--------------------------------|-------|
|       | MAE                             | STD  | MAE                            | STD   |
| ANN   | 5.99                            | 7.35 | 8.08                           | 10.47 |
| SVM   | 5.48                            | 7.07 | 8.25                           | 10.74 |
| LASSO | 6.33                            | 7.82 | 8.66                           | 10.85 |

## 4.3 Results using features extracted from only the PPG

Table 4 compares the performance of various machine learning models for features extracted solely from PPG signals. Herein, the predicted values of the DBP obtained from all three models were similar, also, these results were comparatively better than those predicted for the SBP. With respect to the prediction of the SBP, it was observed that the performance of the LASSO was slightly better than those generated by the ANN and SVM models. The predictions obtained from ANN and LASSO resulted in the best performance for DBP and SBP with values of  $5.52 \pm 6.73$  mmHg and  $9.44 \pm 11.63$  mmHg respectively.

**Table 4** Comparison of the performance between various machine learning models using complexity features from PPG

|       | Diastolic Blood Pressure (mmHg) |      | Systolic Blood Pressure (mmHg) |       |
|-------|---------------------------------|------|--------------------------------|-------|
|       | MAE                             | STD  | MAE                            | STD   |
| ANN   | 5.52                            | 6.73 | 13.05                          | 17.77 |
| SVM   | 5.31                            | 7.05 | 10.57                          | 12.86 |
| LASSO | 5.79                            | 7.40 | 9.44                           | 11.63 |

## 4.4 Results using features extracted from both PPG and ECG signals

A comparison between various machine learning models using features extracted from both PPG and ECG are shown in Table 5. In this case, the MAE of DBP is similar for each learning method and ANN resulted in the optimal prediction of  $5.15 \pm 6.46$  mmHg. Also, the predicted value of  $7.33 \pm 9.53$  mmHg generated by the SVM was the optimal result for the SBP. In line with these findings, it was observed that features extracted from both the ECG and PPG signals resulted in higher prediction accuracy compared to those obtained solely from either ECG or PPG signals as shown in Tables 3 and 4.

**Table 5** Comparison of the performance between various machine learning models using PTT and the features extracted from PPG and ECG

|       | Diastolic Blood Pressure (mmHg) |      | Systolic Blood Pressure (mmHg) |       |
|-------|---------------------------------|------|--------------------------------|-------|
|       | MAE                             | STD  | MAE                            | STD   |
| ANN   | 5.15                            | 6.46 | 7.99                           | 10.34 |
| SVM   | 5.22                            | 6.65 | 7.33                           | 9.53  |
| LASSO | 5.73                            | 7.28 | 8.94                           | 11.01 |

It is important to highlight that there is no available literature to compare the results obtained from extracting features from PPG because this work is the first report BP prediction with features extracted from the PPG signal through the use of complexity analysis. Thus, for the sake of consistency, the results obtained from this research were compared with official standards because their stipulated values are general.

#### 4.5 Evaluation of the results obtained from the models relative to the British Hypertension Society (BHS) Standard

The results from the complexity analysis carried out in this work were compared with the BHS standard as shown in Table 6. Generally, the BHS standard grades BP measurement devices based on the cumulative percentage errors under 5, 10 and 15 mmHg (O'Brien et al. 2001). In this work, the predicted values of DBP obtained from the ANN model was consistent with Grade A as stipulated by the BHS standard. Furthermore, the predicted values of the DBP obtained from SVM and LASSO models were in congruence with Grade B of the BHS standard. However, the estimations of the SBP obtained from all three models was not consistent with the BHS standard.

**Table 6** Comparison with the BHS standard (O'Brien et al. 2001)

|                        |         | Cumulative Error Percentage |          |          |         |          |          |
|------------------------|---------|-----------------------------|----------|----------|---------|----------|----------|
|                        |         | DBP                         |          |          | SBP     |          |          |
|                        |         | ≤5 mmHg                     | ≤10 mmHg | ≤15 mmHg | ≤5 mmHg | ≤10 mmHg | ≤15 mmHg |
| Including all features | ANN     | 60.53 %                     | 89.47 %  | 97.37 %  | 44.74 % | 71.05 %  | 76.32 %  |
|                        | SVM     | 52.63 %                     | 84.21 %  | 97.37 %  | 47.37 % | 71.05 %  | 84.21 %  |
|                        | LASSO   | 50.00 %                     | 84.21 %  | 94.74 %  | 28.95 % | 68.42 %  | 73.68 %  |
| BHS standard           | Grade A | 60 %                        |          |          | 85 %    |          | 95 %     |
|                        | Grade B | 50 %                        |          |          | 75 %    |          | 90 %     |
|                        | Grade C | 40 %                        |          |          | 65 %    |          | 85 %     |

#### 4.6 Evaluation of the results obtained from the models relative to the Association for the Advancement of Medical Instrumentation (AAMI) Standard

The AAMI standard requires BP measurement devices to achieve MAE and STD values lower than 5 mmHg and 8 mmHg, respectively (Jones, Hall 2002). According to Table 5, the

implemented models for the DBP estimation displayed STD values lower than the maximum acceptable STD values, moreover, the MAE values for the DBP obtained from all models were slightly outside the acceptable limit of the AAMI standard. Also, the MAE and STD values of the SBP obtained from all the three models were also outside the stipulated limits.

#### 4.7 Results of inter-patient scheme using 5-fold and 3-fold cross validation

In order to reinforce the results of the first stage where the data of different subjects are merged before train-test split, inter-patient scheme using 5-fold and 3-fold cross validations are adopted. To reduce variability, 20 rounds of cross validation are performed using different random partitions. The results over different rounds are used to calculate the final MAE and STD, which are shown in Table 7-12.

**Table 7** Average BP prediction results from 5-fold cross validation based on subject division using complexity features from ECG.

|       | Diastolic Blood Pressure (mmHg) |      | Systolic Blood Pressure (mmHg) |       |
|-------|---------------------------------|------|--------------------------------|-------|
|       | MAE                             | STD  | MAE                            | STD   |
| ANN   | 6.22                            | 6.93 | 8.19                           | 10.75 |
| SVM   | 5.69                            | 7.56 | 8.96                           | 11.09 |
| LASSO | 6.95                            | 8.27 | 9.38                           | 11.49 |

**Table 8** Average BP prediction results from 5-fold cross validation based on subject division using complexity features from PPG

|       | Diastolic Blood Pressure (mmHg) |      | Systolic Blood Pressure (mmHg) |       |
|-------|---------------------------------|------|--------------------------------|-------|
|       | MAE                             | STD  | MAE                            | STD   |
| ANN   | 5.63                            | 6.78 | 12.94                          | 13.87 |
| SVM   | 5.75                            | 7.11 | 9.76                           | 13.01 |
| LASSO | 6.10                            | 7.83 | 10.89                          | 12.17 |

**Table 9** Average BP prediction results from 5-fold cross validation based on subject division using complexity features from PPG and ECG

|       | Diastolic Blood Pressure (mmHg) |      | Systolic Blood Pressure (mmHg) |       |
|-------|---------------------------------|------|--------------------------------|-------|
|       | MAE                             | STD  | MAE                            | STD   |
| ANN   | 5.11                            | 5.74 | 8.13                           | 10.17 |
| SVM   | 5.85                            | 7.06 | 8.23                           | 9.94  |
| LASSO | 6.42                            | 7.85 | 9.37                           | 10.56 |

**Table 10** Average BP prediction results from 3-fold cross validation based on subject division using complexity features from ECG.

|       | Diastolic Blood Pressure (mmHg) |      | Systolic Blood Pressure (mmHg) |       |
|-------|---------------------------------|------|--------------------------------|-------|
|       | MAE                             | STD  | MAE                            | STD   |
| ANN   | 6.15                            | 7.09 | 8.15                           | 10.51 |
| SVM   | 5.74                            | 7.39 | 8.74                           | 11.18 |
| LASSO | 6.97                            | 8.43 | 9.43                           | 11.27 |

**Table 11** Average BP prediction results from 3-fold cross validation based on subject division using complexity features from PPG

|       | Diastolic Blood Pressure (mmHg) |      | Systolic Blood Pressure (mmHg) |       |
|-------|---------------------------------|------|--------------------------------|-------|
|       | MAE                             | STD  | MAE                            | STD   |
| ANN   | 5.67                            | 6.77 | 12.47                          | 13.69 |
| SVM   | 5.78                            | 7.24 | 9.83                           | 13.08 |
| LASSO | 6.08                            | 7.99 | 11.20                          | 12.11 |

**Table 12** Average BP prediction results from 3-fold cross validation based on subject division using complexity features from PPG and ECG

|       | Diastolic Blood Pressure (mmHg) |      | Systolic Blood Pressure (mmHg) |       |
|-------|---------------------------------|------|--------------------------------|-------|
|       | MAE                             | STD  | MAE                            | STD   |
| ANN   | 5.12                            | 5.86 | 8.19                           | 10.21 |
| SVM   | 5.81                            | 7.13 | 8.41                           | 10.05 |
| LASSO | 6.49                            | 7.87 | 9.16                           | 10.24 |

It can be seen that 5-fold and 3-fold cross validation experiments generate similar results. In comparison with the results in the first stage, ANN still performs well and there are only very small increases in MAE while STD stays relatively the same. SVM and LASSO seem to generate less accurate predictions in terms of MAE for all three different scenarios. Overall, the results of the inter-patient scheme are consistent with the results in the first stage with slight deterioration with SVM and LASSO models.

## 5 Discussion

With respect to the accuracy of the models in predicting the BP based on the source of the complexity features, it was observed that the ANN model has the highest accuracy for predicting the SBP obtained from ECG. Moreover, with the features extracted from a combination of the ECG and PPG signals, the most accurate prediction of SBP was obtained by the SVM model. The accuracy of all three models improved when the features were extracted from a combination of the ECG and PPG signals.

From an assessment of the prediction efficacy of the different models, SVM was more accurate in predicting the DBP obtained from the ECG signals, whereas the prediction of DBP values obtained from either the PPG or its combination with the ECG signals more accurately predicted by the ANN model. The higher prediction accuracy displayed by the ANN model could be attributed to its strong ability to learn and optimally resolve non-linear and complex

relationships (Maind, Wankar 2014). The optimal prediction of the DBP from features extracted from the ECG signals as exhibited by the SVM model could be linked to this model's strong ability to resolve non-linear relationships coupled with its minimal requirement for training resources and high noise tolerance (Kachuee et al. 2016). It is interesting to point out that the three different models were all able to predict DBP more accurately than SBP, using all combinations of models and signals.

It is noteworthy to mention that the accuracy of the models in predicting the SBP values generated from features extracted from either or both the ECG and PPG signals are similar to the results in the literature (Sun et al. 2016; Kachuee et al. 2016; Simjanoska et al. 2018). With regards to the DBP estimation, the most accurate predictions were generated from features extracted from the combination of ECG and PPG, and the errors were not significantly large. From a broader perspective, it can be concluded that features extracted from the combination of both signals resulted in accurate prediction of both the DBP and SBP by all three models considered. In line with this, the simplicity, ability to be designed as a wearable component, and other practical advantages of the PPG monitoring unit generally make it a viable system that could be adopted for the continual monitoring of BP.

With respect to the evaluation of the model with the BHS standard, it was observed that the estimation of the DBP values with the aid of the ANN model was consistent with Grade A, whereas the data obtained from the SVM and LASSO models can be classed under Grade B. Although the results of the SBP obtained from all three models deviated from the stipulated the BHS and AAMI standards, nevertheless, the error margin could be narrowed further. For example, in this research, all 14 volunteers are males mainly because the adoption of the chest electrodes would be inconvenient for female volunteers. Accordingly, two major limitations of this study could be linked to the bias in the results obtained from one sex, and the small number of subjects. In line with this, future investigations aimed at further developing this study would include both sexes, and the sample size of subjects would also be increased, so a wide range of different people's data will increase the accuracy of the results. Moreover, the stress induced from the controlled exercise undertaken by the subjects could also introduce experimental artefacts which may affect the acquired physiological signals, thus, the signal pre-processing aspect of this work should be further optimised. Additionally, more accurate results could be obtained by making measurement only at rest, not during exercise, and by training the neural networks using personalized data from individuals over a longer period. This indicates that the adopted models may still be optimised to achieve accurate predictions in accordance with the criteria set out by these standards.

Inter-patient scheme is used by employing 5-fold and 3-fold cross validation on the subjects and its results are overall consistent with the results from the first stage. ANN is proved to provide especially robust results.

Although, the BP estimations within this work were obtained through the use of ANN, SVM and LASSO regression models, it is pertinent to state that the inclusion of more machine learning models with complex structures coupled with the novel approach at feature extraction reported in this work could result in the generation of more accurate BP estimations. Relatedly, even though the two morphological features of PAT and heart rate were added to the signal complexity features, still, future aspects of this research would benefit from incorporating more morphological features in order to further improve the accuracy of the BP estimation.



## 6 Conclusion

In this article, a novel method for BP estimation using complexity analysis to extract features from PPG and ECG signals has been proposed. The participants in this study were a group of 14 male volunteers subjected to different exercise scenarios from which signals of ECG and PPG together with BP values were simultaneously measured. In all, 129 distinct datasets were adopted for pre-processing, and these served as the input for the signal complexity analysis which resulted in the extraction of 90 features from the PPG and ECG signals. Three regression models, ANN, SVM and LASSO, were adopted for the separate prediction of DBP and SBP. Unlike in traditional methods in which the BP values were predicted by processing the morphological features from both the PPG and ECG signals, the approach adopted in this study utilised signal complexity analysis to resolve the information encoded in the PPG or ECG signals thereby making it possible to use machine learning models to predict the SBP and DBP. Accordingly, this work described a novel approach for the estimation of BP. Furthermore, the use of signal complexity analysis demonstrated that in addition to being able to extract features from either PPG or ECG signals with relatively high prediction capabilities of the BP, the extraction of features from a combination of both signals markedly increased the prediction accuracy.

## Acknowledgements

The authors acknowledge the financial support from the International Doctoral Innovation Centre, Ningbo Education Bureau, Ningbo Science and Technology Bureau, and the University of Nottingham. This work was also supported by the UK Engineering and Physical Sciences Research Council [grant numbers EP/G037345/1 and EP/L016362/1] and the Optics and Photonics Research Group in the University of Nottingham.

## References

- Addison, P.S.: Slope transit time (STT): A pulse transit time proxy requiring only a single signal fiducial point. *IEEE Transactions on Biomedical Engineering* **63**(11), 2441-2444 (2016)
- Buxi, D., Redoute, J.-M., Yuce, M.R.: A survey on signals and systems in ambulatory blood pressure monitoring using pulse transit time. *Physiol Meas.* **36**(3), R1-R26 (2015)
- Fu, W.J.: Penalized regressions: the bridge versus the lasso. *Journal of computational and graphical statistics* **7**(3), 397-416 (1998)
- Goli, S., Jayanthi, T.: Cuff less continuous non-invasive blood pressure measurement using pulse transit time measurement. *Int. J. Recent Dev. Eng. Technol* **2**(1), 16-86 (2014)
- Hori, M., Okamoto, H.: Heart rate as a target of treatment of chronic heart failure. *Journal of cardiology* **60**(2), 86-90 (2012)
- Hughes, D., Geddes, L., Babbs, C., Bourland, J.: Measurements of young's modulus of the canine aorta in-vivo with 10 mhz ultrasound. In: 1978 Ultrasonics Symposium 1978, pp. 326-326. IEEE
- Islam, M.T., Zabir, I., Ahamed, S.T., Yasar, M.T., Shahnaz, C., Fattah, S.A.: A time-frequency domain approach of heart rate estimation from photoplethysmographic (PPG) signal. *Biomedical Signal Processing and Control* **36**, 146-154 (2017)
- Jones, D.W., Hall, J.E.: The national high blood pressure education program: thirty years and counting. *Hypertension* **39**(5), 941-942 (2002)

- Kachuee, M., Kiani, M.M., Mohammadzade, H., Shabany, M.: Cuffless blood pressure estimation algorithms for continuous health-care monitoring. *IEEE Transactions on Biomedical Engineering* **64**(4), 859-869 (2016)
- Kugiumtzis, D., Tsimpliris, A.: Measures of analysis of time series (MATS): A MATLAB toolkit for computation of multiple measures on time series data bases. *arXiv preprint arXiv:1002.1940* (2010)
- Leonarduzzi, R.F., Schlotthauer, G., Torres, M.E.: Wavelet leader based multifractal analysis of heart rate variability during myocardial ischaemia. In: 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology 2010, pp. 110-113. IEEE
- Li, J.-Q., Li, R., Chen, Z.-Z., Deng, G.-Q., Wang, H., Mavromoustakis, C.X., Song, H., Ming, Z.: Design of a continuous blood pressure measurement system based on pulse wave and ECG signals. *IEEE journal of translational engineering in health and medicine* **6**, 1-14 (2018)
- Li, T., Zhou, M.: ECG classification using wavelet packet entropy and random forests. *Entropy* **18**(285), 1-16 (2016)
- Liang, Y., Elgendi, M., Chen, Z., Ward, R.: An optimal filter for short photoplethysmogram signals. *Scientific data* **5**(180076), 1-12 (2018)
- Maind, S.B., Wankar, P.: Research paper on basic of artificial neural network. *International Journal on Recent and Innovation Trends in Computing and Communication* **2**(1), 96-100 (2014)
- Millasseau, S.C., Ritter, J.M., Takazawa, K., Chowienzyk, P.J.: Contour analysis of the photoplethysmographic pulse measured at the finger. *Journal of hypertension* **24**(8), 1449-1456 (2006)
- Monroy Estrada, G., Mendoza, L.E., Molina, V.: Relationship of blood pressure with the electrical signal of the heart using signal processing. *Tecciencia* **9**(17), 9-14 (2014)
- Mukkamala, R., Hahn, J.-O., Inan, O.T., Mestha, L.K., Kim, C.-S., Töreyn, H., Kyal, S.: Toward ubiquitous blood pressure monitoring via pulse transit time: theory and practice. *IEEE Transactions on Biomedical Engineering* **62**(8), 1879-1901 (2015)
- O'Brien, E., Waeber, B., Parati, G., Staessen, J., Myers, M.G.: Blood pressure measuring devices: recommendations of the European Society of Hypertension. *Bmj* **322**(7285), 531-536 (2001)
- Organization, W.H., Group, I.S.o.H.W.: 2003 World Health Organization (WHO)/International Society of Hypertension (ISH) statement on management of hypertension. *Journal of hypertension* **21**(11), 1983-1992 (2003)
- Qi, J., Yang, P., Min, G., Amft, O., Dong, F., Xu, L.: Advanced internet of things for personalised healthcare systems: A survey. *Pervasive and Mobile Computing* **41**, 132-149 (2017)
- Sharma, M., Barbosa, K., Ho, V., Griggs, D., Ghirmai, T., Krishnan, S.K., Hsiai, T.K., Chiao, J.-C., Cao, H.: Cuff-less and continuous blood pressure monitoring: a methodological review. *Technologies* **5**(2), 21 (2017)
- Shin, W., Cha, Y.D., Yoon, G.: ECG/PPG integer signal processing for a ubiquitous health monitoring system. *Journal of Medical systems* **34**(5), 891-898 (2010)
- Simjanoska, M., Gjoreski, M., Gams, M., Madevska Bogdanova, A.: Non-invasive blood pressure estimation from ECG using machine learning techniques. *Sensors* **18**(4), 1-20 (2018)
- Sun, S., Bezemer, R., Long, X., Muehlsteff, J., Aarts, R.: Systolic blood pressure estimation using PPG and ECG during physical exercise. *Physiological measurement* **37**(12), 2154 -2169 (2016)
- Suzuki, A.: Inverse-model-based cuffless blood pressure estimation using a single photoplethysmography sensor. *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine* **229**(7), 499-505 (2015)
- Visvanathan, A., Sinha, A., Pal, A.: Estimation of blood pressure levels from reflective Photoplethysmograph using smart phones. In: 13th IEEE International Conference on Bioinformatics and BioEngineering 2013, pp. 1-5. IEEE
- Wang, L., Zhou, W., Xing, Y., Zhou, X.: A novel neural network model for blood pressure estimation using photoplethysmography without electrocardiogram. *Journal of healthcare engineering* **2018**, 1-9 (2018)

- Wu, C.-M., Chuang, C.Y., Chen, Y.-J., Chen, S.-C.: A new estimate technology of non-invasive continuous blood pressure measurement based on electrocardiograph. *Advances in Mechanical Engineering* **8**(6), 1-8 (2016)
- Yang, W.-Y., Mujaj, B., Efremov, L., Zhang, Z.-Y., Thijs, L., Wei, F.-F., Huang, Q.-F., Luttun, A., Verhamme, P., Nawrot, T.S.: ECG voltage in relation to peripheral and central ambulatory blood pressure. *American journal of hypertension* **31**(2), 178-187 (2017)
- Zaki, W.S.W., Correia, R., Korposh, S., Hayes-Gill, B.R., Morgan, S.P.: Cuff-Less Continuous Blood Pressure Monitoring System Using Pulse Transit Time Techniques. *International Journal of Integrated Engineering* **8**(1), 51-57 (2016)
- Zhang, B., Ren, H., Huang, G., Cheng, Y., Hu, C.: Predicting blood pressure from physiological index data using the SVR algorithm. *BMC bioinformatics* **20**(1), 1-15 (2019)
- Zhang, Q., Zhou, D., Zeng, X.: Highly wearable cuff-less blood pressure and heart rate monitoring with single-arm electrocardiogram and photoplethysmogram signals. *Biomedical engineering online* **16**(1), 1-20 (2017)
- Zhao, Q., Zhang, L.: ECG feature extraction and classification using wavelet transform and support vector machines. In: *2005 International Conference on Neural Networks and Brain 2005*, pp. 1089-1092. IEEE (2005)