A New Blood Pressure Estimation Approach Using PPG Sensors: Subject Specific Evaluation over a Long-term Period

Franck Mouney[1,2(✉)], Teodor Tiplica[1], Jean-Baptiste Fasquel[1], Magid Hallab[2], and Mickael Dinomais[3]

[1] LARIS Laboratory, Angers University, Angers, France
[2] Clinique Bizet, Service de Cardiologie, 23 Rue Georges Bizet, 75116 Paris, France
[3] CHU Angers, Angers, France

Abstract. In this paper, a new approach for predicting the blood pressure (BP) from the photoplethysmogram (PPG) signal is proposed related with a new original public dataset. The originality of the dataset is based on the fact that subjects are periodically monitored over weeks, while public datasets consider short acquisition periods. The proposed BP estimation approach uses key frequencies in the spectrum of the PPG signal isolated using the LASSO algorithm, then a predictive model is constructed as a patient-specific BP estimation model. The efficiency of the proposed methodology is evaluated on experimental data recorded over a long time period. Moreover, an evaluation of the various temporal markers of the PPG signal that have been proposed in the literature is conducted on the same data set. It is showed that only few of these temporal markers are useful for the estimation of the systolic and diastolic blood pressures. The results highlight that better blood pressure estimations are obtained when using the spectrum of the PPG signal rather than optimally selected temporal markers.

Keywords: Blood pressure estimation · LASSO · PPG · spectrum · Temporal markers · Dataset

1 Introduction

Arterial BP has always been a key physiological measurement in the frame of medical examination, being one of the most important bio-markers in clinical evaluation. Thus, continuously monitoring the BP in order to predict cardiovascular diseases is one of the major challenges for the next years.

Some recent surveys focus on the use of Ballistocardiogram signals [10], Electrocardiography (ECG) signals [20], or on both ECG and photoplethysmogram (PPG) signals [14,24] to predict BP. Many researchers focus their works on predicting cuff-less BP measurement from PPG signals only [16,19]. In this paper,

S. Paiva et al. (Eds.): SmartCity360° 2020, LNICST 372, pp. 45–63, 2021.
https://doi.org/10.1007/978-3-030-76063-2_4

one considers PPG-based BP estimation, because such an approach enables a non-invasive and continuous measurement from wearable sensors.

PPG based methods uses optical properties and the relationship between synchronous changes in the blood volume and the arterial BP [7]. Red or near infrared wavelengths are used for the PPG light source as they allow to correctly evaluate blood volume changes through absorption and reflection phenomenons of blood cells. Figure 1 illustrates the use of such sensor integrated in a wearable device (watch/band) for BP monitoring from a PPG signal.

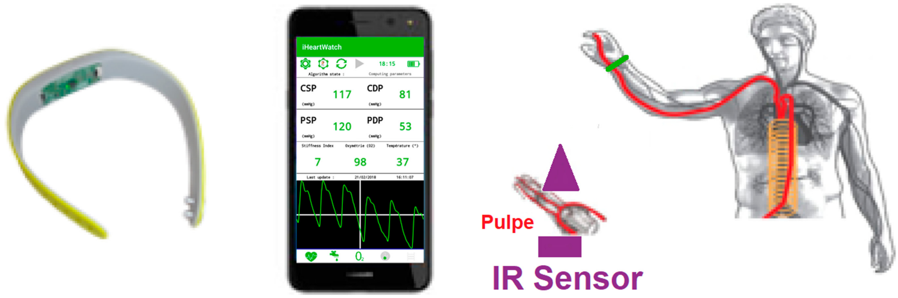

Fig. 1. Example of wearable device (watch) enabling to monitor PPG.

As it appears in the literature, BP estimation from PPG sensors has been mainly studied by considering time domain approaches [5, 13, 25]. Most of them involve signal pre-processing (smoothing, filtering, etc.), extraction of temporal features and finally data-based model for BP estimation.

Frequential features have been investigated for BP estimation. For instance, Fast Fourier Transform (FFT) has been recently considered to extract spectral information to estimate BP [30]. As recently underlined [19], spectral-domain-based techniques are more convenient as there is no need to detect pulse waves, compared to time-domain-based techniques. Note that some recent approaches combine both temporal and frequential information [2].

Besides BP estimation techniques, another crucial aspect concerns the PPG signal collection protocol to estimate the BP. According to Stergiou et al. [21], a new kind of validation protocols will be developed for continuous, cuff-less, and central BP monitors. The IEEE Std 1708$^{\text{TM}}$ [1] was specifically studied to define the objective performance to evaluate wearable cuff-less BP measurement devices with different operating mode (e.g., to measure short-term, long-term, snapshot, continuous, beat(s)-to-beat(s) BP, or BP variability). An important underlying aspect regards the evaluation dataset that has to be representative enough.

In this context, we created a new data set for evaluating blood pressure estimation models. The originality of this dataset is that it concerns the long time period follow-up of a subject, with PPG and BP acquisitions (snapshot) regularly taken over several weeks. Existing datasets do not consider such acquisition protocol and are limited to short time acquisitions.

The originality of this paper concerns the evaluation of all existing temporal features on our original dataset and the proposition of a new efficient frequency-based BP estimation method.

Section 2 describes both the material and the proposed approach, including the methodology that is considered for comparing with time-based features. Section 3 presents the results. Before concluding, our work is discussed in Sect. 4.

2 Material and Method

2.1 Material

The acquisition protocol has been drawn in order to obtain clean and replicable data and most importantly the same physiological condition over several weeks. This protocol implies that the subject lies down for 5 min with the medical devices already positioned on him (avoiding physiological variations in the microvascular capillaries resulting from PPG clip positioned on the finger). Then, after 5 min, three consecutive measurements are made with 1 min interval between them. Both PPG signal and blood pressure are recorded at the same time.

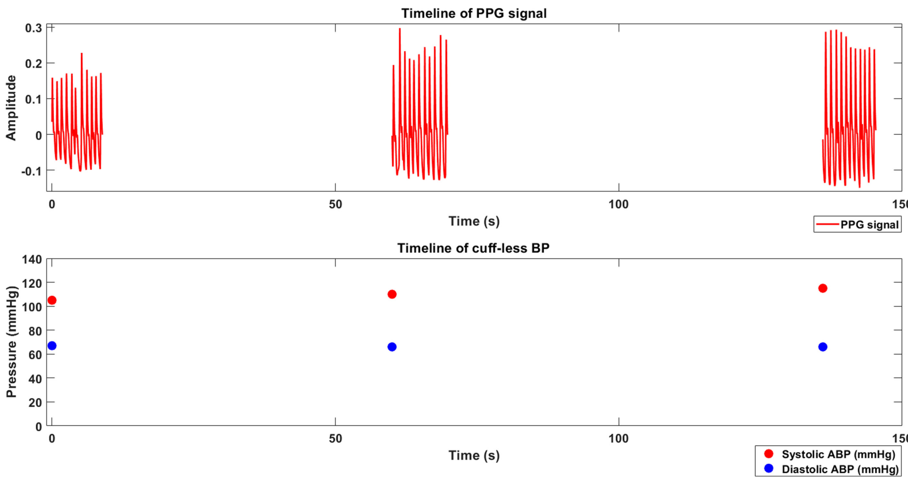

Fig. 2. Timeline of one day record.

We used a certified BP cuff Cardio Maxi[1] and a classic PPG sensors plug via pOpmetre device[2]:

[1] https://mon-materiel-medical-en-pharmacie.fr/tensiometres/669-tensiometre-cardio-maxi.html.

[2] https://www.axelife.fr/le-popmetre-en/.

- The blood pressure cuff is positioned on the left arm of the subject.
- The PPG clamp is positioned on the thumb of the subject's right hand.

We assume that there is no difference of PPG signals with respect to the selected arm, as studied by [9].

The measurements are usually done at the same time every day just before lunch in order to avoid other disturbance of the daily life (e.g. lunch, coffee, etc.). Every day, during 6 weeks, three consecutive measurements of 10 s PPG signals were recorded (sampling frequency 1000 Hz) together with the corresponding cuff-less BP (see Fig. 2). As a result, 84 PPG signals and their corresponding systolic and diastolic blood pressures were recorded on a period of 40 days. To capture the intrinsic blood pressure variation, even if all acquisitions are achieved in the same conditions, a long period monitoring is needed. This aspect is illustrated and discussed in Sect. 3, providing an overview of the distribution of measured blood pressures for the considered subject.

2.2 Method

The methodology used in order to predict the blood pressure from the PPG signal is illustrated by Fig. 3. The three main steps of the proposed method are detailed hereafter. Note that the proposed frequency-based method corresponds to the right part of Fig. 3, while the left part deals with the comparison with existing temporal features.

Data Processing: First, each PPG signal is standardized by substracting its mean value and by dividing by its standard deviation (Std), leading to PPG_s defined by:

$$PPG_s(t) = \frac{PPG(t) - Mean[PPG(t)]}{Std[PPG(t)]} \tag{1}$$

Then, the data processing is conducted separately in the time and frequency domains.

Time Domain: As previously mentioned, most of related works consider a pre-processing step (filtering) before extracting temporal features.

The pre-processing step is required to correct artefacts resulting from the acquisition [15,16]: motion artefact, loose of skin contact, power-line interference, irregular cardio-vascular rhythm, detection of signal saturation, etc.... In our case, no artefacts were detected during acquisition.

Frequential filtering is mostly used with cutoff frequency band usually set to [0.1–10] Hz [16]. However, in order to filter unrelated artefacts from BP variations, Kalman filtering can be considered [11]. The motivation for such an adaptive filtering is that PPG signal is subject related. Indeed, according to [11], non-stationary effects caused by breathing artefacts are affecting the PPG signal. In such a case, an adaptative filtering is required to take into account the

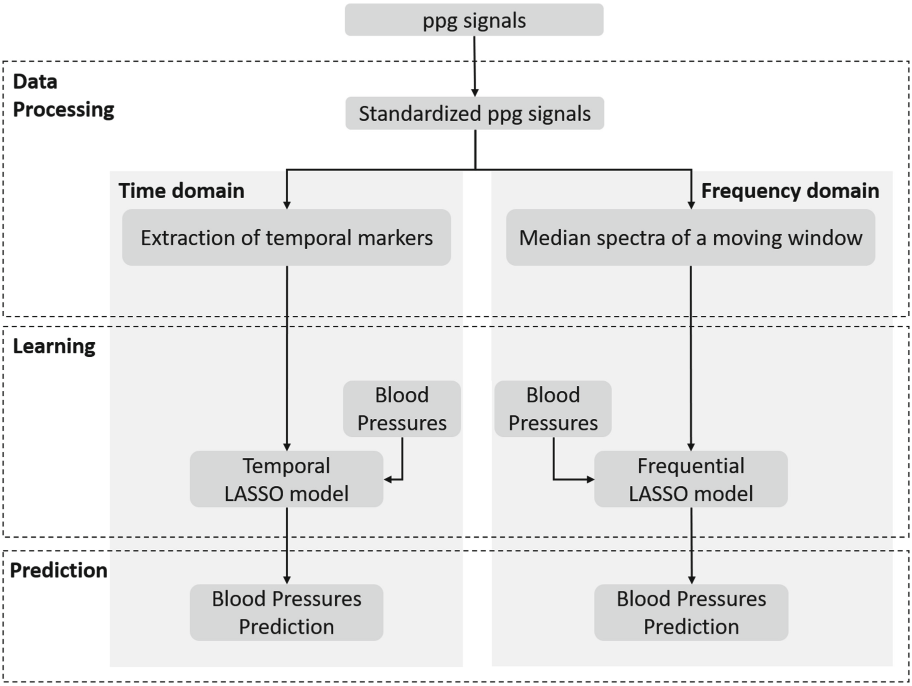

Fig. 3. Methodology used for predicting the blood pressure: the proposed frequency-based method corresponds to the right part, while the left one concerns the comparison with the use of temporal markers.

breathing time variation. Thus, the pre-processing filtering needs the ability to adapt its parameter over-time.

In [3], authors considered a classical band-pass filtering from 0.5 Hz to 5 Hz followed by a baseline drift removal. They used a sum of 2 Gaussian distributions to approximate the shape of each pulse (see Fig. 5). It allows the reduction of irregularities in the shape of the pulse captured by a smartphone. Note that [13] do not use any pre-processing at signal level but tried to reduce the noise during the recording from the camera of the smartphone by considering brightness, skin colour and position of the finger on the camera.

We use a low pass filter to suppress high frequencies in the signal with cutoff frequency 10 Hz which allows to suppress any power-line interference (50 Hz, 60 Hz, etc.). A smoothing is then applied by means of a digital filter with a polynomial order of three in order to have good quality of 1st and 2nd derivative of the signal.

Regarding time domain features present in the literature, there are different categories that have been tested to predict blood pressure (see Fig. 4). Note that related works use a combination of such temporal features. The categories and the corresponding published papers are:

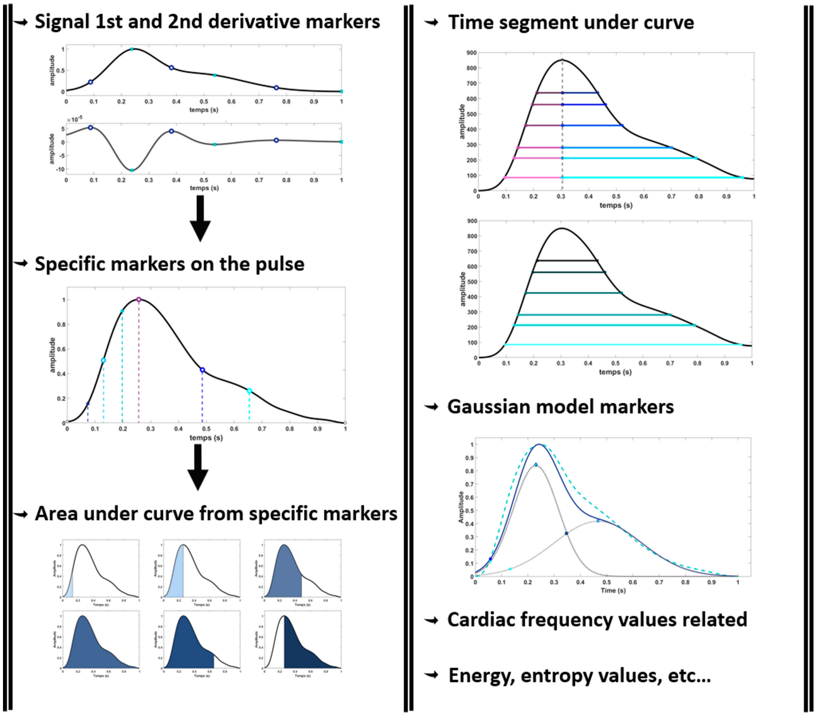

Fig. 4. Gaussian model method

- Specific markers on the pulse: [23,29,31].
- Area under curve from specific markers: [28,29,31].
- 1st and 2nd derivatives: [17,22,31].
- Time segment under curve: [11–13,17,25,29].
- Cardiac frequency values related: [5,11–13,28,31].
- Energy, entropy, skewness coefficient, etc.: [18].
- Gaussian model parameters: similar to [3].

We introduced new temporal features: 21 features from Gaussian model (see Fig. 5), energy of the signal, the slope of the upstroke, the skewness and kurtosis of the pulse. The 21 features built from the fitted Gaussian model [4] correspond to specific points, some of them being indicated in Fig. 5. Regarding the Gaussian model developed in this article, we extracted 21 parameters from this method which are mostly taken from some specific coordinate points (see Fig. 5). We use the method of maximum likelihood and the expectation maximization (EM) algorithm to make the sum of the two Gaussian to fit the closest possible the raw pulse [4]. Note that a similar approach has been considered for the analysis of noisy signals [3]. For sake of clarity, all features are not exhaustively described in this paper.

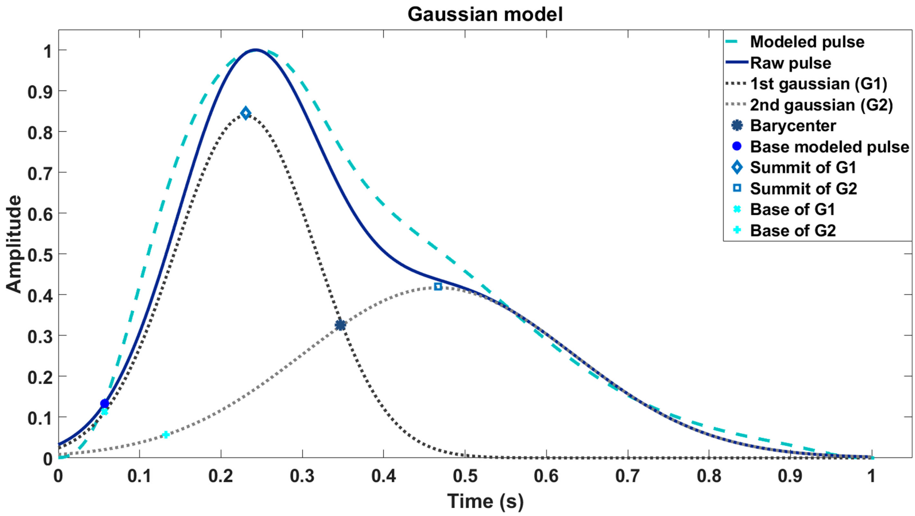

Fig. 5. Gaussian model and some key markers of the model.

In this study, we consider 153 temporal features (markers), including the newly introduced ones. Only relevant ones (leading to the best estimation) will be summarized in Sect. 3.

Frequency Domain: Given that acquisition of the PPG signals is not yet automatic (i.e. with a fixed and controled duration), the time length of the various PPG signals lies between 4.47 and 32.98 s. This is troublesome because the various spectra of the PPG signals depict different spectral resolutions. In order to overcome this problem, the PPG spectrum results from the computation of the median spectrum over sliding windows of 10 s, as illustrated by Fig. 6. Thus, the PPG signals having less than 10 s are not used in this study. That means that in our case only 56 PPG signals out of 84 are retained and their median spectra computed as described before.

Moreover, to increase the resolution of the PPG signal spectra, the PPG signal has been padded (adding zero values at end of the signal) up to 10^5 (equivalent of a temporal window of 100 s at 1 KHz sampling rate).

Learning: At this step, the temporal markers and respectively the median spectra of the PPG signals together with their corresponding systolic and diastolic blood pressures are considered to build a predictive model of the subject blood pressure. In both cases (temporal and frequential), the LASSO (Least Absolute Shrinkage and Selection Operator) algorithm [26] is used in order to build the predictive model. The LASSO is an L1 penalized regression technique that enjoys some of the proprieties of both subset selection and ridge regression. The main reason that motivated the choice of the LASSO algorithm in our study is the high number of predictive variables (153 temporal markers and respectively

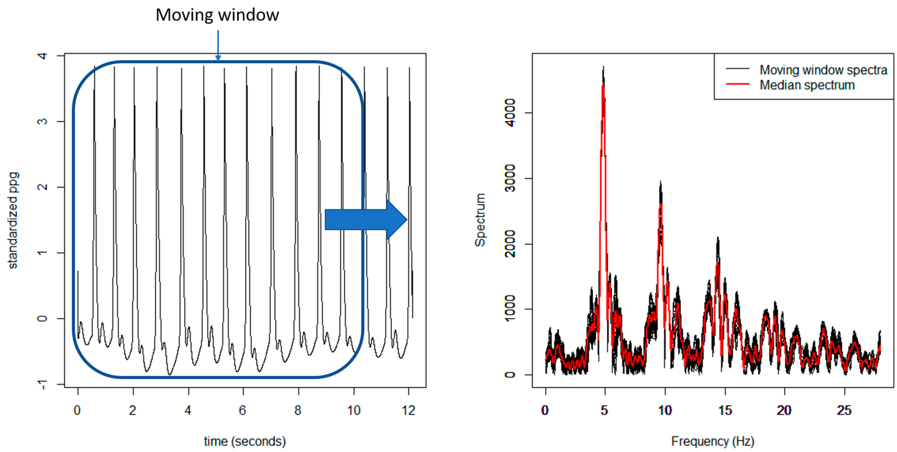

Fig. 6. Computing the median spectrum of a moving window on the PPG signal

700 frequencies) and the few number of observations (84 in the case of temporal markers and respectively 56 in the case of spectral descriptors). Indeed, the ordinary least squares estimates in a regression model are obtained by minimizing the residual squared error. This usually generates low bias estimators but having in counterparts large variance and thus poor estimation accuracy. The LASSO algorithm sets to zero some coefficients in order to reduce the variance of the predicted values and thus to improve the overall estimation accuracy. When a large number of predictors are initially taken into account in the predictive model, the LASSO algorithm keeps a smaller subset that exhibits the strongest effects and thus helps to interpret the resulting model easier. The interested readers can consult the following references [6,27] for more information on the LASSO algorithm. In this paper, we used the R software and the **lars** package (version: 1.2) proposed by Trevor Hastie and Brad Efron for computing the various results obtained.

Estimation: For the estimation of the blood pressure, the trained LASSO model is used. The LASSO model inputs takes the given temporal markers or the given median spectra of the PPG signal and the model realises the blood pressure estimation. The efficiency of the predictive model is evaluated through the Mean Square Error (MSE) indicator obtained by cross-validation. The leave-one-out cross validation method is used in this study. The graphics relevant to the predicted versus real BP additionally give an illustration about the quality of the predictive model, this being detailed in next Section.

3 Results

For the subject taken in consideration in this study, a summary of statistics concerning his blood pressure variation over the survey period of 6 weeks (28 days) is given in Table 1. The patient is a healthy male of 21 years old (height 1.75 m and weight 50 kgs). Note that Q_{25} and Q_{75} represent the 1st and 3rd quartiles of the blood pressure distribution.

Table 1. Summary of statistics of the subject's blood pressure (in mm Hg)

Blood pressure	Min	Q_{25}	Median	Mean	Q_{75}	Max
Systolic	86	95	96.5	97.07	100.25	107
Diastolic	53	57	59	59.04	60.25	67

Figure 7 shows the spread of the blood pressures recorded over the 6 weeks of the survey. This illustrates the previously mentioned variability that can be observed over long time period, even though acquisitions are performed in the same conditions.

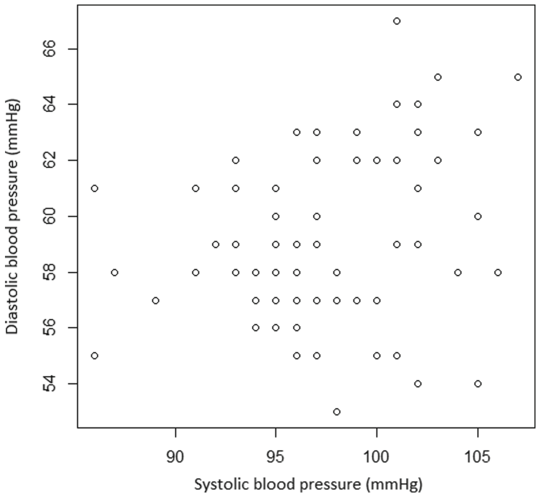

Fig. 7. Subject blood pressure variation over the survey period of 6 weeks

The results concerning the temporal and frequential LASSO models, respectively obtained from temporal markers and median spectra are presented in next Sections.

3.1 Temporal LASSO Model

The LASSO algorithm is used to identify the temporal markers that are useful for the blood pressure estimation. Figure 8 reports the leave-one-out cross-validated

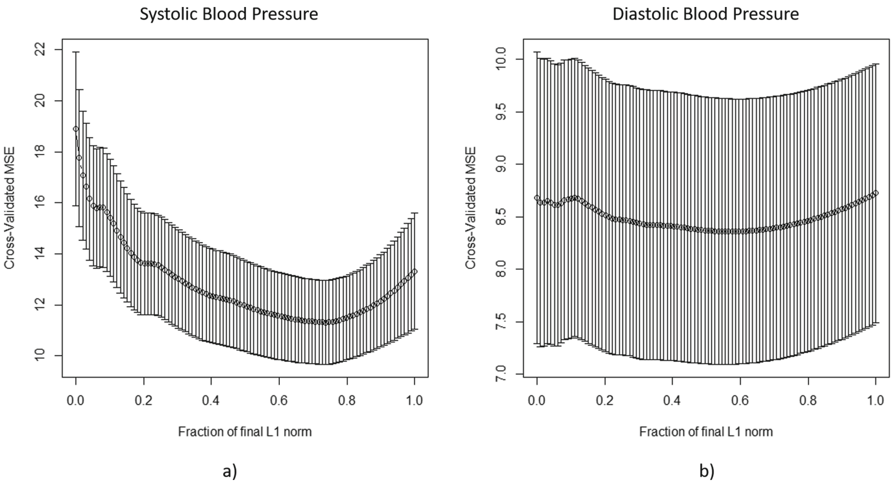

Fig. 8. The cross-validated mean squared estimation error of the lasso algorithm when using the temporal markers of the PPG signal: case of the systolic blood pressure (a) and the diastolic blood pressure (b)

mean squared estimation error of the lasso algorithm, for both systolic and diastolic blood pressures.

The lowest MSE of the blood pressure estimation is obtained with two sets of temporal markers, corresponding to either the systolic or the diastolic pressures.

In the case of the systolic blood pressure (newly introduced markers are those without reference in the literature):

- $ATG1$: Base amplitude of the first Gaussian model.
- ASM: Area under curve from start to summit of the pulse [31].
- $NCCP$: Number of consecutive pulse detected to evaluate cardiac frequency.
- $TUTE90$: Time width between up-stroke and down-stroke at 90% of amplitude under curve [3,5,25].
- TE: Time width mean of pulses detected [3,28,31].
- $Aa2D$: Amplitude of the 1st top of the signal's second derivative [17].
- HV: Variance of the spectral entropy of the signal. Note that the spectral entropy H has been used by [18].
- LgE: Energy of the signal (inspired from [18]).
- $KTEM$: Mean of the Kaiser-Teager energy of the signal [18].
- $SQIM$: Mean of the Skewness of the pulses in the signal [15].

In the case of the diastolic blood pressure (newly introduced markers are those without reference in the literature):

- $NCCP$: Number of consecutive pulse detected to evaluate cardiac frequency.
- TDP: Time width between start to diastolic peak (2nd maximum point) of the pulse [29].

– AbP: Amplitude of the 2nd minimum point of the signal's second derivative reported the pulse [17,22].
– $ASMMG$: Mean area under curve between start of the pulse to maximum amplitude extracted from Gaussian model (inspired from [31]).
– $AMDP$: Mean area under curve between start to diastolic peak (2nd maximum point) of the pulse [29].
– $Te2D$: Time width between start to the third top of the signal's second derivative of the pulse [17].
– $Tf2D$: Time width between start to the fifth top of the signal's second derivative of the pulse (Inspired from [17] and [22]).

The final predictive LASSO models for both the systolic and diastolic blood pressures are reported in Table 2.

Table 2. LASSO models for the estimation of the systolic (SBP) and diastolic (DBP) blood pressure when using temporal markers extracted from the PPG signal

SBP	ATG1	ASM	NCCP	TUTE90	TE	Aa2D	HV	LgE	KTEM	SQIM
	−5.56	−1460.55	2.36	−0.01	65.40	4331.39	−2.1e+07	−42.64	6.36e+05	−30.19
DBP	NCCP	TDP	AbP	ASMMG	AMDP	Te2D	Tf2D	-	-	-
	1.23	−39.26	−11294.59	988.39	220.86	0.05	−0.04	-	-	-

Figure 9 illustrates the predicted BP versus real BP when using the LASSO predictive models given in Table 2. The leave-one-out cross validated MSE for respectively predicting the systolic blood pressure and the diastolic blood pressure are respectively 11.31 and 8.36.

Blue points in Fig. 9 are the 56 measures that are retained in the frequential analysis (records where the duration of the PPG signal is more than 10 s). These points are highlighted for comparative purposes with the frequential approach detailed in next Section.

3.2 Frequential LASSO Model

Hereafter, the LASSO algorithm is used to identify frequencies of the PPG spectra that are useful for the blood pressure estimation. Figure 10 gives the leave-one-out cross-validated mean squared estimation error of the lasso algorithm for both systolic and diastolic blood pressures.

The LASSO model with the lowest MSE (18.07) used for the estimation of the systolic blood pressure estimation is given in Table 3.

The LASSO model with the lowest MSE (7.95) used for the estimation of the diastolic blood pressure is given in Table 4.

Figure 11 reports predicted versus real blood pressures when using LASSO predictive models given in Tables 3 and 4.

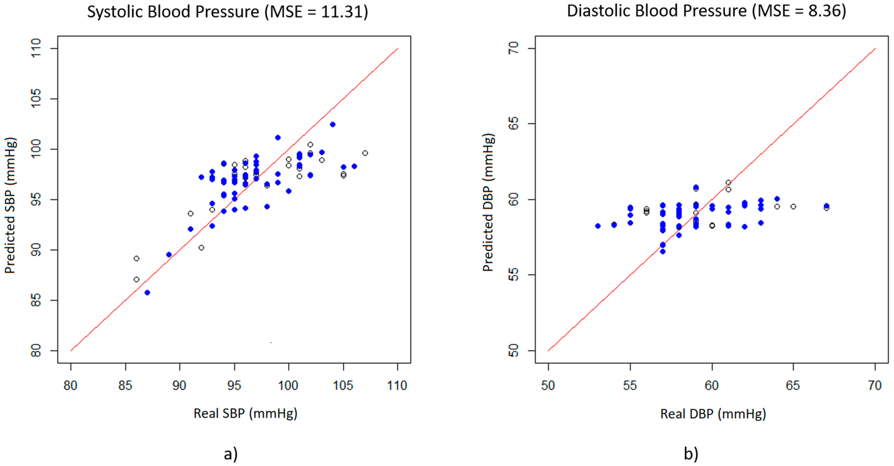

Fig. 9. Predicted versus real systolic blood pressure (a) and diastolic blood pressure (b) for the temporal LASSO model. Blue points correspond to the 56 measurements used for the frequential analysis.

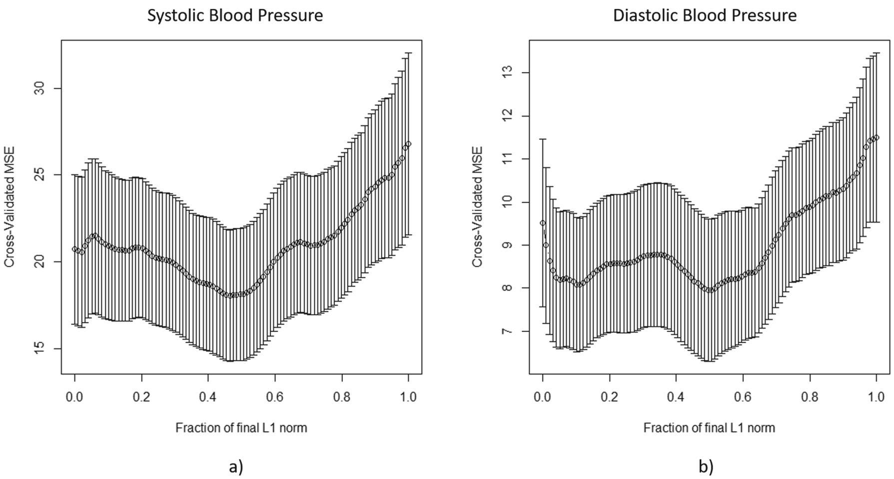

Fig. 10. The cross-validated mean squared estimation error of the lasso algorithm when using the spectra of the PPG signal: case of the systolic blood pressure (a) and the diastolic blood pressure (b)

Table 3. Coefficient of the LASSO model for the estimation of the systolic blood pressure when using the PPG signal spectrum

Frequency (Hz)	0.36	2.08	5.20	6.00	6.28	8.60	10.12	10.48
LASSO coefficients ($\times 10^{-4}$)	−31.32	6.66	7.10	5.54	31.10	−6.11	−25.08	64.51
Frequency (Hz)	10.88	10.92	14.64	15.68	16.52	17.20	17.24	17.76
LASSO coefficients ($\times 10^{-4}$)	−2.21	−22.47	−29.31	15.28	5.11	−5.97	−10.97	−54.22
Frequency (Hz)	18.08	18.36	19.88	20.00	20.24	20.56	20.60	22.60
LASSO coefficients ($\times 10^{-4}$)	22.98	−56.71	10.07	62.41	−35.46	6.20	20.26	20.74
Frequency (Hz)	22.64	25.32	25.56	26.60	27.20	27.56		
LASSO coefficients ($\times 10^{-4}$)	16.95	12.93	−32.07	65.43	24.03	22.43		

Table 4. LASSO model for the estimation of the diastolic blood pressure when using the PPG signal spectrum

Frequency (Hz)	0.04	1.76	2.12	2.60	2.88	5.24	5.76	5.80
LASSO coefficients ($\times 10^{-4}$)	−16.01	34.04	7.10	16.85	0.47	9.92	5.76	6.24
Frequency (Hz)	6.00	6.76	7.08	8.44	8.60	8.76	13.08	13.60
LASSO coefficients ($\times 10^{-4}$)	16.27	−1.24	3.08	−3.25	−0.01	−3.17	−8.06	−17.80
Frequency (Hz)	14.40	15.04	15.76	15.80	17.16	18.84	18.88	19.20
LASSO coefficients ($\times 10^{-4}$)	−2.99	−12.37	−13.61	−1.46	−27.77	−1.05	−7.54	−28.36
Frequency (Hz)	19.88	19.92	21.88	22.92	23.96	24.88	25.08	27.52
LASSO coefficients ($\times 10^{-4}$)	42.60	1.05	48.18	−6.56	21.96	38.08	15.75	198.89

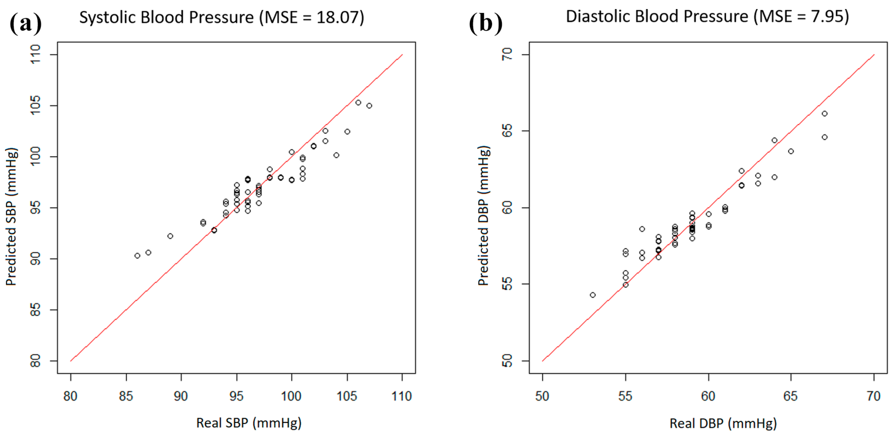

Fig. 11. Predicted versus real systolic blood pressure (a) and diastolic blood pressure (b) for the frequential LASSO model

4 Discussion

According to Table 1 and Fig. 7, it appears that the blood pressure variation is quite consequent even if the subject is always in the same position (lying down). Indeed, we noticed that the systolic blood pressure vary between 86 mmHg and 107 mmHg while the diastolic blood pressure fluctuates between 53 mmHg and 67 mmHg. Thus, a gradient of 21 mmHg on the systolic blood pressure and respectively of 14 mmHg on the diastolic blood pressure is observed over the 6 weeks of daily records. Therefore BP estimation methods should integrate this variability on the evaluation protocol, requiring long term datasets.

The first aspect that we want to discuss here aspect concerns the temporal markers of the PPG signal that are related to the blood pressure estimation. Despite the high number of temporal markers that have been proposed in the literature for predicting the blood pressure, to our knowledge, there is no comparative study of their performances. Indeed, in order to make an objective comparison of these temporal markers, an essential aspect that must be taken into account is obviously to use the same data set. Some public databases that can be used for the blood pressure estimation from the PPG signals like MIMIC [8] or Elgendi [15] have some drawbacks. Concerning the Elgendi dataset, very short PPG signals (2.1 s) are recorded, and only once. Regarding the MIMIC dataset, signals are recorded on subjects in intensive care unit, meaning that subjects might have been under medication or in very unstable states. Although signals are recorded over hours, we cannot extract the same type of information compared to daily records over several weeks as in our case.

The LASSO algorithm found that only 10 temporal markers are useful for the estimation of the systolic blood pressure (see Table 3). The estimation of the systolic blood pressure seems to fit quite well for low blood pressures (less than 95 mmHg) - see Fig. 9a. But, for higher pressures, a saturation effect appears in the estimation (the model can't correctly predict higher blood pressures and the output of the model seems to be a random variation around 97 mmHg). Concerning the diastolic blood pressure, only 7 out of 153 temporal markers seem to be related to the blood pressure. When looking to the Fig. 9b, one can easily notice that the diastolic blood pressure estimation is very poor even if the MSE seems to be acceptable. Indeed, the model can't predict efficiently the diastolic blood pressure variation and the estimations look like a random variation around 59 mmHg. Figure 12 better illustrates this observation over the different records (sessions). The real blood pressures are represented by the black points. The blue points are the estimations when the blood pressure is underestimated and the red points are the estimations when the blood pressure is overestimated. These points are linked to the black points by vertical blue and respectively red lines. It is obvious that less higher are the vertical lines and better is the estimation. It can be noticed that a lot of variation in the blood pressure remains unexplained especially for the diastolic blood pressure. This clearly shows that using just some indicators like MSE, RMSE (Root Mean Square Error), MAE (Mean Average Error), etc., is not sufficient in order to evaluate the quality of a estimation.

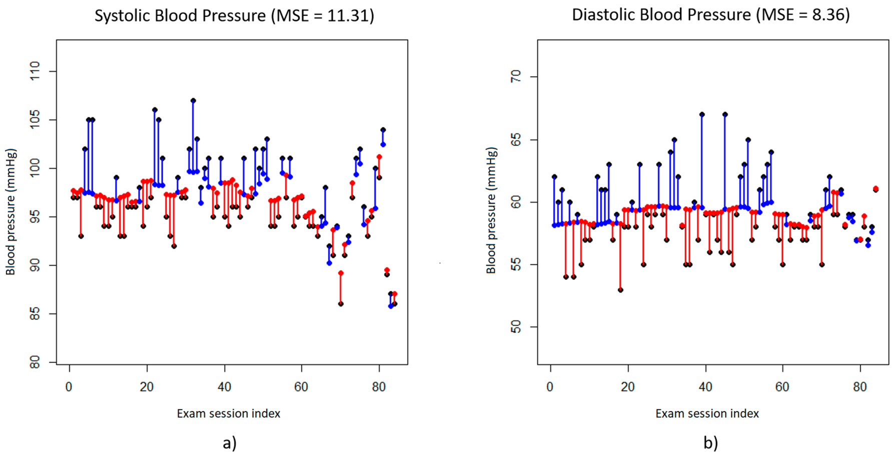

Fig. 12. Predicted systolic blood pressure (a) and diastolic blood pressure (b) for the temporal LASSO model

Concerning the frequential LASSO model used for the estimation of the blood pressure, it has to be noticed that we used only 30 frequencies in the case of the systolic blood pressure (see Table 3) and 32 frequencies in the case of the diastolic blood pressure (see Table 4). It appears that better estimations are obtained than in the case of the temporal markers. Indeed, one can easily visually observe that both predicted systolic and diastolic blood pressures better fit real blood pressures than in the case of temporal markers (see Fig. 11, to be compared with blue points in Fig. 9). It is clear that, in the spectra of the ppg signal, there are some frequencies that can be used to efficiently predict the entire spread of the blood pressure variation. The estimation of the diastolic blood pressure is highly better than in the case of the temporal markers. In the case of the systolic blood pressure, the model seems to slightly underestimate the high blood pressures and to slightly overestimate the low blood pressures. Anyway, the quality of the systolic blood pressure estimation is significantly improved compared to the use temporal markers because the saturation observed when the estimations are done in the temporal domain is not present in the spectral domain. Figure 13 better illustrates the higher quality of estimations in the spectral domain. Compared to the estimations based on temporal markers (see Fig. 12), the height of the vertical lines representing the difference between the real and the predicted is significantly reduced. Regarding the relevance of indicators such as the considered MSE, note that reported values such as 18.07 (frequential LASSO on Systolic pressure) and 11.31 (temporal LASSO on Systolic pressure), are not sufficient to reflect the efficiency of the approach: indeed, according to MSE (computed through a leave-one-out procedure), the temporal approach seems to outperform the frequential one. Nevertheless, as discussed hereabove, it is not the case, as illustrated by Figures (Figures report estimations, using trained models, on the entire dataset).

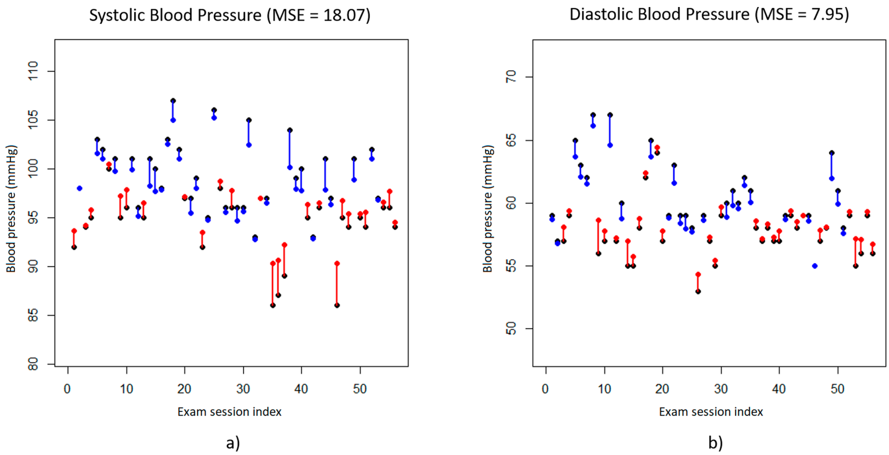

Fig. 13. Predicted systolic blood pressure (a) and diastolic blood pressure (b) for the frequential LASSO model

These results appears accurate enough to integrate the proposed algorithm into a medical device, although additional evaluations have to be performed on an extended dataset.

The last aspect to be pointed out is the size of dataset, limited to one subject with relatively few BP measurements for these preliminary experiments. Further experiments are planned with more subjects and more BP measurements per subject. Therefore, in this work, the limited number of BP measurements may involve a risk of over-fitting, even if one considers a cross-validation. Additionaly, LASSO models resulting from the cross-validation procedure may not keep the same set of (temporal or frequential) features for each cross-validated result: it would be interesting to study, on a larger database, whether the automatically selected set of features remains the same whatever the subdataset used to build the LASSO models. Concerning the evaluation procedure, one considers the standard leave-one-measurement out procedure, usually used in data analysis. It would be interesting to use a leave-one-day out procedure to evaluate the model performance over time, in adequacy with the nature of this dataset and of the considered application.

5 Conclusion

We propose a new data set to be used by researchers working in the field of estimation of the blood pressure from the PPG signal. This data set contains 84 measurements of the systolic and diastolic blood pressures related to PPG signals. Subject has been monitored over a period of 6 weeks, this being not the case any other public data set. We hope that more data set recording subjects over long time period will appear in the next few years as we found out that it enables better blood pressure estimation model to come out and as it enables to compare efficiently estimation approaches. We will add more content to our data set in the future encouraging researcher to do the same. The public data set is available online[3].

We objectively evaluated on this data set various temporal markers of the PPG signal that have been proposed in the literature for the estimation of the blood pressure. Based on our data set, few of them have been found to be useful for the blood pressure estimation. The estimation quality is relatively low, especially for the diastolic blood pressure.

We proposed a new subject specific blood pressure estimation approach based on the spectrum of PPG signals. Our approach exploits the LASSO algorithm for the selection of key frequencies to predict the blood pressure. Our method gives better results than in the case of time markers and variations appear correctly captured over the distribution of recorded blood pressures. An additional strength of this approach is its robustness regarding bad signal quality, compared to temporal approaches that are very sensitive to artefacts.

Some future works concern the test of the proposed approach on an extended dataset involving a new acquisition protocol where the subjects' blood pressures and their corresponding PPG signals are measured in different conditions: at rest lying down (the current dataset); at rest standing up; immediately after a physical effort and finally a short time after the physical effort. We hope that much more variation of the blood pressure will be available in this way. Data issued from this new protocol will be used for different purposes: identification of the subject profile when doing effort; calibration of the predicting model according to the subject profile; better estimation of the blood pressure in conditions that are closer to the real life activities of a person.

References

1. Association, I.S., et al.: IEEE standard for wearable Cuffless blood pressure measuring devices. IEEE Std. 1708–2014 (2014)
2. Baek, S., Jang, J., Yoon, S.: End-to-end blood pressure prediction via fully convolutional networks. IEEE Access **7**, 185458–185468 (2019)
3. Banerjee, R., Ghose, A., Choudhury, A.D., Sinha, A., Pal, A.: Noise cleaning and Gaussian modeling of smart phone photoplethysmogram to improve blood pressure estimation. In: 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 967–971. IEEE (2015)
4. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer, New York (2006)
5. Choudhury, A.D., Banerjee, R., Sinha, A., Kundu, S.: Estimating blood pressure using windkessel model on photoplethysmogram. In: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 4567–4570. IEEE (2014)
6. Efron, B., et al.: Least angle regression. Ann. Stat. **32**(2), 407–451 (2004)
7. Elgendi, M.: On the analysis of fingertip photoplethysmogram signals. Curr. Cardiol. Rev. **8**(1), 14–25 (2012)

[3] https://www.kaggle.com/franckycash/cuff-blood-pressure-ppg-over-6-weeks.

8. Goldberger, A.L., et al.: Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. Circulation **101**(23), e215–e220 (2000)
9. Jiang, X., et al.: Change of bilateral difference in radial artery pulse morphology with one-side arm movement. Artery Res. **19**, 1–8 (2017)
10. Kim, C.S., Carek, A.M., Inan, O.T., Mukkamala, R., Hahn, J.O.: Ballistocardiogram-based approach to cuffless blood pressure monitoring: proof of concept and potential challenges. IEEE Trans. Biomed. Eng. **65**(11), 2384–2391 (2018)
11. Kurylyak, Y., Barbe, K., Lamonaca, F., Grimaldi, D., Van Moer, W.: Photoplethysmogram-based blood pressure evaluation using kalman filtering and neural networks. In: 2013 IEEE International Symposium on Medical Measurements and Applications (MeMeA), pp. 170–174. IEEE (2013)
12. Kurylyak, Y., Lamonaca, F., Grimaldi, D.: A neural network-based method for continuous blood pressure estimation from a PPG signal. In: 2013 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), pp. 280–283. IEEE (2013)
13. Lamonaca, F., et al.: Application of the artificial neural network for blood pressure evaluation with smartphones. In: 2013 IEEE 7th International Conference on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS), vol. 1, pp. 408–412. IEEE (2013)
14. Landry, C., Peterson, S.D., Arami, A.: Nonlinear dynamic modeling of blood pressure waveform: Towards an accurate cuffless monitoring system. IEEE Sens. J. **20**(10), 5368–5378 (2020)
15. Liang, Y., Chen, Z., Liu, G., Elgendi, M.: A new, short-recorded photoplethysmogram dataset for blood pressure monitoring in china. Sci. Data **5** (2018)
16. Liang, Y., Elgendi, M., Chen, Z., Ward, R.: An optimal filter for short photoplethysmogram signals. Sci. Data **5** (2018)
17. Liu, M., Po, L.M., Fu, H.: Cuffless blood pressure estimation based on photoplethysmography signal and its second derivative. Int. J. Comput. Theory Eng. **9**(3), 202 (2017)
18. Monte-Moreno, E.: Non-invasive estimate of blood glucose and blood pressure from a photoplethysmograph by means of machine learning techniques. Artif. Intell. Med. **53**(2), 127–138 (2011)
19. Mouney, F., Tiplica, T., Hallab, M., Dinomais, M., Fasquel, J.B.: Towards a smartwatch for cuff-less blood pressure measurement using PPG signal and physiological features. In: International Conference on IoT Technologies for HealthCare (2019)
20. Simjanoska, M., Gjoreski, M., Gams, M., Madevska Bogdanova, A.: Non-invasive blood pressure estimation from ECG using machine learning techniques. Sensors **18**(4), 1160 (2018)
21. Stergiou, G.S., et al.: A universal standard for the validation of blood pressure measuring devices: association for the advancement of medical instrumentation/European society of hypertension/international organization for standardization (AAMI/ESH/ISO) collaboration statement. Hypertension **71**(3), 368–374 (2018)
22. Suzuki, S., Oguri, K.: Cuffless and non-invasive systolic blood pressure estimation for aged class by using a photoplethysmograph. In: 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 1327–1330. IEEE (2008)

23. Suzuki, S., Oguri, K.: Cuffless blood pressure estimation by error-correcting output coding method based on an aggregation of adaboost with a photoplethysmograph sensor. In: 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 6765–6768. IEEE (2009)
24. Tang, Z.: A chair-based unobtrusive cuffless blood pressure monitoring system based on pulse arrival time. IEEE J. Biomed. Health Inform. **21**(5), 1194–1205 (2016)
25. Teng, X., Zhang, Y.: Continuous and noninvasive estimation of arterial blood pressure using a photoplethysmographic approach. In: Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No. 03CH37439), vol. 4, pp. 3153–3156. IEEE (2003)
26. Tibshirani, R.: Regression shrinkage and selection via the lasso. J. Roy. Stat. Soc. B **58**(1), 267–288 (1996)
27. Tibshirani, R.: The lasso method for variable selection in the cox model. Stat. Med. **16**(4), 385–395 (1997). https://doi.org/10.1002/(SICI)1097-0258(19970228)16:4⟨385::AID-SIM380⟩3.0.CO;2-3
28. 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, pp. 1–5. IEEE (2013)
29. Xie, Q., Wang, G., Peng, Z., Lian, Y.: Machine learning methods for real-time blood pressure measurement based on photoplethysmography. In: 2018 IEEE 23rd International Conference on Digital Signal Processing (DSP), pp. 1–5. IEEE (2018)
30. Xing, X., Sun, M.: Optical blood pressure estimation with photoplethysmography and fft-based neural networks. Biomed. Opt. Express **7**(8), 3007–3020 (2016)
31. Yang, S., Zhang, Y., Cho, S.Y., Morgan, S.P., Correia, R., Wen, L.: Cuff-less blood pressure measurement using fingertip photoplethysmogram signals and physiological characteristics. In: Optics in Health Care and Biomedical Optics VIII, vol. 10820, p. 1082036. International Society for Optics and Photonics (2018)