Cuffless Blood Pressure Estimation for Continuous 24/7 Patient Monitoring

Maria Margarida Almeirão Brites
mariambrites@tecnico.ulisboa.pt

Instituto Superior Técnico, Lisboa, Portugal

January 2021

Abstract

Blood pressure (BP) is an important factor in the monitoring of patients admitted to the general ward, which is the focus of the Wireless Assessment of Respiratory and Circulatory Distress (WARD) clinical support system. Current practice relies on intermittent cuff based measurements, based on the oscillometric method. This method has major drawbacks such as the low frequency of evaluation and discomfort for the patients. Continuous cuffless blood pressure estimation methods have been explored in the literature as solutions for these problems. This master thesis aims at addressing cuffless blood pressure estimation using a data-driven method based on a machine learning algorithm (Random Forest).

Several morphological features and pulse arrival time features were extracted from the photopletymogram (PPG) waveform, its derivative and second derivative and from the ECG waveform. The set of features was used to train two Random Forest Regression models to estimate systolic blood pressure (SBP) and diastolic blood pressure (DBP), respectively. The BP estimation algorithm was fist trained and tested on data publicly available at the Multi-Parameter Intelligent Monitoring for Intensive Care II (MIMIC II) database. On a second stage, the solution was applied to data from the WARD project.

Although signal quality created difficulties in achieving results that compare to those in the literature, in a small subset of high quality data from the MIMIC II database, it was possible to obtain SBP and DBP estimations with a mean error of $5.20 \pm 5.13 mmHg$ and $1.70 \pm 7.98 mmHg$, respectively. The results obtained in the WARD data suggest the signal pre-processing and cleaning pipeline should be improved to meet the clinical standards. Despite the limitations, a machine learning method based on PPG and ECG features shows potential for the estimation of blood pressure without a cuff.

Keywords: Blood Pressure, Cuffless, Patient monitoring, Photoplethysmography, Electrocardiogram, Random Forest

1. Introduction

Blood pressure is a key hemodynamic variable, since subtle changes of its values, together with changes in other biosignals, are early signs of clinical deterioration eventually leading to adverse events [10]. Blood pressure is often monitored continuously in critically ill patients. It is common procedure to monitor arterial blood pressure (ABP) by means of an arterial catheter during high-risk surgery, in the postoperative and in the intensive care unit (ICU). The advantages of this monitoring technique include instantaneous detection of pressure changes and accuracy [15], as this is the gold standard method to monitor BP [17]. However, this procedure is invasive, and may lead to complications so it is not suitable for general ward. Alternatively, blood pressure is often monitored using auscultation or oscillometry methods, which are non-invasive methods that employ an inflatable cuff [17].

Blood pressure is one of the biosignals being monitored in the Wireless Accessment of Respiratory and Circulatory Distress (WARD) project. The WARD project is a collaboration between the Technological University of Denmark (DTU), Rigshospitalet and Bispebjerg hospital. It aims at conducting a continuous fully automatic assessment of vital signal being monitored in high-risk patients and concerns general ward patients. In this project, signals are continuously measured from inpatients in the post operative period after major surgery and by patients admitted due to chronic obstructive pulmonary disease (COPD) exacerbation, using wear and forget devices. At this point, 500 patients in the first situation and 200 in the latter have been monitored, meaning that a large data set is now available as a basis for algorithm development. The biosignals acquired include electrocardiogram (ECG), photopletysmogram (PPG) and blood pressure, among others.

Currently in the WARD project, systolic blood pressure (SBP) and diastolic blood pressure (DBP) are measured every 30 or 15 min during daytime and

every hour or half hour from 10PM to 7AM using a cuff-based ambulatory blood pressure monitor. However, these measurements may cause discomfort to the patients and are disruptive during sleeping. A cuffless system for blood pressure estimation would allow unobtrusive and continuous measurements, providing more information about blood pressure variation.

Several systems have been proposed for continuous and cuffless blood pressure estimation. Many rely on pulse transit time (PTT) or pulse arrival time measures (PAT), which consist of time differences obtained from two simultaneously acquired pulse signals representing the activity of the heart [17]. The electrocardiogram and photopletismogram waveforms acquired in the scope of the WARD project are commonly used to obtain such measures.

Other systems which rely on features from the ECG and PPG waveforms separately have also been proposed. However, integration into clinical practice has not yet been achieved, possibly because the methods proposed either lack accuracy or validation over a sufficiently large population [7] [22]. Therefore, it is important to further improve and test these blood pressure estimation methods.

The main aim of the project is to investigate a method to estimate systolic and diastolic BP based on biosignals acquired in the scope of the WARD project, in a continuous or semi-continuous and unobtrusive manner.

2. State of the art on cuffless BP estimation

Cuffless blood pressure estimation has been extensively studied over the last years. The classical BP estimation models consist of mapping indicators that can reveal the BP changes, such as PTT and PAT to the BP values [22]. Several approaches have been attempted to model the relationship between BP indicator variables and BP and they can be roughly divided in two groups. One of the possible approaches consists of using theory-based mathematical models to map the relationship between these indicators and the BP, requiring expert knowledge of the underlying physiologically processes. The alternative are data-driven approaches, based on machine learning techniques.

2.1. Model-Based Methods

Early studies such as those by Chen et al. and Poon et al., two of the most cited works in the field, developed PTT models for BP estimation based on the Moens-Korteweg equation. This equation correlates pulse wave velocity (PWV), the velocity at which an arterial pressure wave propagates along the walls of the arterial tree, with the modulus of elasticity of the arterial wall [22]. Although their results demonstrated that PTT is able to track BP, both models have shortcomings such as limited accuracy and short calibration intervals [13]. More recently in 2015, the photoplethysmogram intensity ratio (PIR) was proposed as a new indicator for BP estimation in a study by Ding et al. [4]. The inclusion of this parameter, which can be affected by changes in the arterial diameter, improved the estimation when compared to BP models based only on PTT.

The theory-driven model-based methods have the advantages of being generalizable and interpretable. However, the relationship between BP indicators has been demonstrated to be more complicated than a simple linear or nonlinear regression model. In the case of PTT, it has been shown to have distinct correlation with BP among different individuals [22]. There are many factors affecting blood pressure, such as age, temperature, mental stress, and different behaviour pattern, which are not reflected in this indicator.

2.2. Data-driven Methods

Machine learning methods are particularly valuable for cuffless BP estimation due to their ability to constantly learn from data. Also, machine learning techniques allow the use of multiple indicators for the estimation, which would be difficult to integrate in a physiological mathematical model, and may prevent BP prediction from being affected by confounding factors or noise in a single indicator [18].

One of the first to study this approach was Monte-Moreno, who combined a set of features describing several PPG characteristics in several machine learning techniques to predict continuous SBP [18]. Ridge linear regression, a multilayer perceptron neural network, support vector machines and random forests where tested, and the best performance was obtained with the Random Forest Tree method, which has resulted in a coefficient of determination between the reference and the prediction of 0.91 and 0.89 for SBP and DBP, respectively [22].

Later, in a study by Ruiz-Rodriguez [19], a neural network based method based on PPG features was studied on patients undergoing continuous invasive BP measurement with an arterial catheter. The validation group included 47 patients and the results obtained were not satisfactory to allow clinical application [19].

In 2013, Kurylyak et al. also used an Artificial Neural Network and a set of 21 features extracted from the PPG waveform to estimate SBP and DBP. Training and test data were extracted from the MIMIC database and the estimation results presented a mean absolute error of $3.80 \pm 3.46 mmHg$ for SBP and $2.21 \pm 2.09 mmHg$ for DBP. However, no information is given on the amount of subjects in which it was tested [11].

Recently, more and more researchers have attempted machine learning methods for cuffless BP estimation. For instance, Kachuee et al. extracted multiple physiological parameters from ECG and PPG signals and used several regression algorithms to estimate BP [9]. Also, Xing et al. developed a method to measure blood pressure without calibration and based only on the PPG signal and height, using a random

forest algorithm [25].

With recent developments in machine learning, more advanced methods such as Deep Neural Networks have also been used to model the nonlinear relationship between the BP predictors and BP measurements. Temporal dependencies between the raw input signals and blood pressure have also been model using recurrent neural networks [22].

3. Methodology

3.1. Dataset

The data used are from the Multi-Parameter Intelligent Monitoring for Intensive Care II (MIMIC II) database. MIMIC II is a freely available database that contains data from more than 25000 ICU patients who stayed in critical care units of the Beth Israel Deaconess Medical Center in Boston, Massachusetts. Data were collected over a seven year period, beginning in 2001 [2]. The database contains both clinical data stored in a relational database and waveform data recorded by the bedside monitors. The waveform data (MIMIC-II Waveform Database) contains records from a subset of patients and includes physiological signals such as electrocardiogram (ECG), photopletysmogram (PPG) and intra-arterial blood pressure (ABP) [2].

The particular version of the database used (MIMIC II Waveform Database Matched Subset, version 3.1, which can be found at https://archive.physionet.org/physiobank/database/mimic2wdb/matched/) contains 4,897 waveform records and 5,266 numerics records matched with 2,809 MIMIC II Clinical Database records. Not all signals are available for all of the subjects in this subset. Therefore, records containing simultaneous PPG, ABP and ECG lead II waveforms were selected. Abnormal (low quality) signal identified in a previous work by Liang et al. [12] were also excluded from the present analysis. The records selected are from a total of 146 patients and have durations from 10 minutes to 1 hour. The distribution of systolic and diastolic blood pressure records is shown in figure 1.

The steps implemented to estimate blood pressure from the data extracted from the MIMIC II database are described in figure 2, and will be described in detail in the following sections.

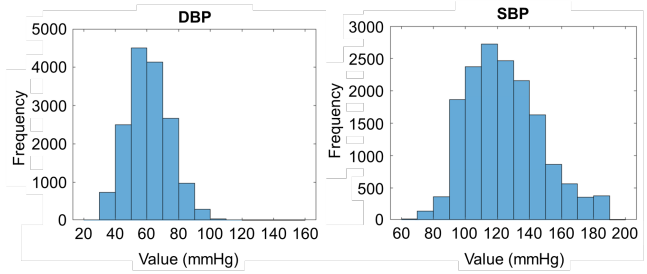

Figure 1: Distribution of parameters in the records selected from the MIMIC II database: systolic blood pressure (SBP) and diastolic blood pressure (DBP)

3.1.1 State of the art method implementation

To establish a basis for comparison, the method by Kachuee et al. [9] was implemented. This method was selected based on the possibility to apply it to the particular biosignals collected in the WARD project as well as on performance results, number of subjects in which it was evaluated and in the absence of calibration. This paper suggests two different implementations. In the first, features extracted from the PPG and ECG signals are based on physiological parameters. On the alternative, a whole-based representation of vital signals is used. The first was chosen since it yielded better BP estimation results. The model was trained and tested on the MIMIC II data described in the previous section. The Random Forest Regression was the algorithm selected, as it led to the best estimation results in the paper, and feature vector used as input contained the features described in figure 4 and table 1.

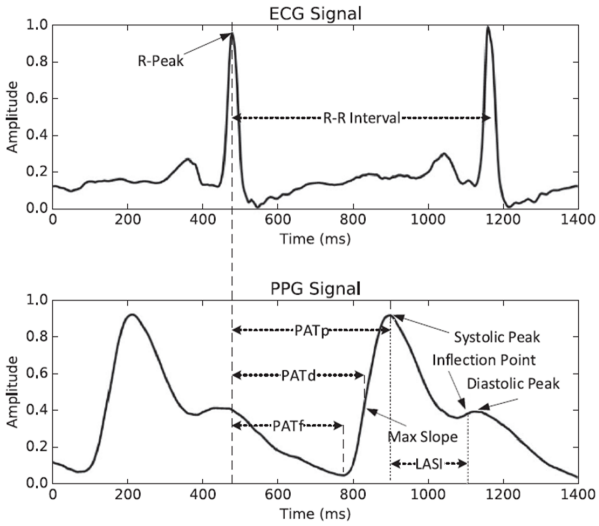

Figure 4: PPG and ECG features. The features are further described in table 1. Adapted from [9]

3.1.2 Pre-processing

In the method developed, the ABP, PPG and ECG signals are first resampled from 125 Hz to 1000Hz using linear interpolation. The PPG and ECG signals are then filtered and denoised using discrete wavelet transform (DWT), following a similar method to [9]. The signal is decomposed to level 10 using DWT with the Daubechies 8 (db8) mother wavelet. Both low frequencies (from 0 to 0.98Hz) and high frequencies (from 250 to 500 Hz) are removed by zeroing the respective decomposition coefficients. Wavelet denoising is then performed on the remaining coefficients using *MATLAB* function `wdenoise`.

BP signal cleaning Systolic blood pressure (SBP) and diastolic blood pressure (DBP) are computed for

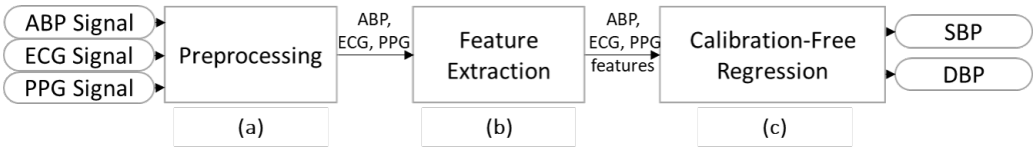

Figure 2: Block diagram of the proposed cuffless BP estimation method, where SBP and DBP features used as label are extracted from the ABP signal. The steps in blocks (a), (b) and (c) are described in sections 5.3, 5.4 and 5.5, respectively.

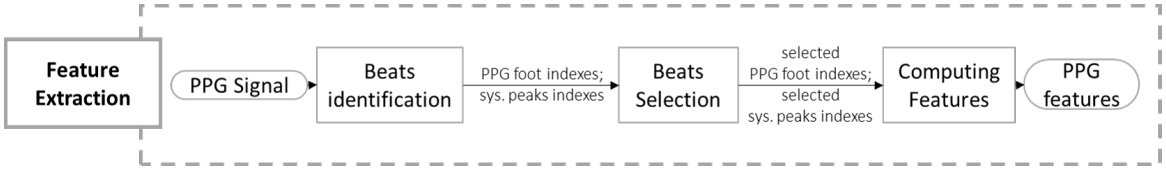

Figure 3: Feature extraction process, corresponding to step (b) in the block diagram in figure 2

Table 1: Features computed from the PPG and ECG for each heart cycle. AI-Augmentation Index; LASI-Large Artery Stiffness Index

Feature	Description
PATf	time interval between the ECG R-peak and the PPG foot
PATd	time interval between the ECG R-peak and the PPG derivative maximum
PATp	time interval between the ECG R-peak and the PPG systolic peak
RR interval	Time interval between consecutive ECG R peaks
AI	ratio of the height of the diastolic peak (x) to the systolic upeak (y) in the pulse $AI = \frac{x}{y}$
LASI	time delay between the systolic peak and the point of inflection
S1	Area under the curve (AUC) from start of cycle to max derivative point
S2	AUC from max derivative point to systolic peak
S3	AUC from systolic peak to diastolic rise
S4	AUC from diastolic rise to end of cycle

each heart cycle. Abnormal SBP and DBP values are discarded, using a method based on the signal abnormality index (SAI) proposed by Sun et al [24]. The BP data exclusion criteria are shown in table 2. The mean arterial pressure (MAP) is defined as:

$$MAP = \frac{2DBP + SBP}{3} \tag{1}$$

Table 2: Features computed from the ABP signal and respective abnormality criteria. Adapted from Sun et al.[24]

Feature	Description	Abnormality Criteria
SBP	Systolic blood pressure	$SBP > 300mmHg$
DBP	Diastolic blood pressure	$DBP < 20mmHg$
PP	Pulse pressure	$PP < 20mmHg$
MAP	Mean arterial pressure	$MAP < 30mmHg$ or $MAP > 200mmHg$
T	Duration of each beat	
f	Heart rate (60/T)	$f < 20$ or $f > 200bpm$
ΔSBP	$SBP[k] - SBP[k-1]$	$\lvert\Delta SBP\rvert > 20mmHg$
ΔDBP	$DBP[k] - DBP[k-1]$	$\lvert\Delta DBP\rvert > 20mmHg$
ΔT	$T[k] - T[k-1]$	$\lvert\Delta T\rvert > 2/3sec$

PPG Signal Cleaning The PPG signal quality is evaluated based on a signal quality index proposed by Elgendi [6]. Elgendi compared eight different signal quality indices, and for lengths of PPG waveforms between 2 s and 30 s, the skewness Signal Quality Index (sSQI) method demonstrated better performance than others. Skewness is a measure of the symmetry of a probability distribution and sSQI is defined as:

$$S_{SQI} = 1/N \sum_{i=i}^{N} [x_i - \hat{\mu}_x/\sigma]^3 \tag{2}$$

where $\hat{\mu}_x/$ and σ are the empirical estimate of the mean and standard deviation of x_i, respectively, and N is the number of samples in the PPG signal [6]. Using this method, the classifications of excellent waveforms versus acceptable or unfit were best when 5s was used as the window of the PPG waveform segment. Therefore, the sSQI is computed for each 5 second segment of the PPG signal. Segments with a value below zero are discarded. Flat segments where no signal is present are also removed.

3.1.3 Feature Extraction

The methodology followed to estimate blood pressure relies on the extraction of several features from the ECG and PPG signals, which have been used in previous studies with the same goal. All the features are described in tables 1 and 3, and in figures 4, 5a and 5b. While many of these parameters are proposed in literature without explicit meaning, others have physiological meanings which have been described.

Pulse arrival time (PAT) corresponds to the time delay between the electrical activity of the heart and a peripheral pulse measured in a peripheral point in the arterial tree. [22]. It indirectly approximates the PWV, which reflects the properties of the arterial wall, namely arterial stiffness [22]. The PAT features are defined in table 1 and figure 4.

Large artery stiffness index (LASI) is related to the transit time of pressure waves from the root of the subclavian artery to the apparent site of reflection and back to the subclavian artery [5]. This metric is computed as the time difference between systolic peak

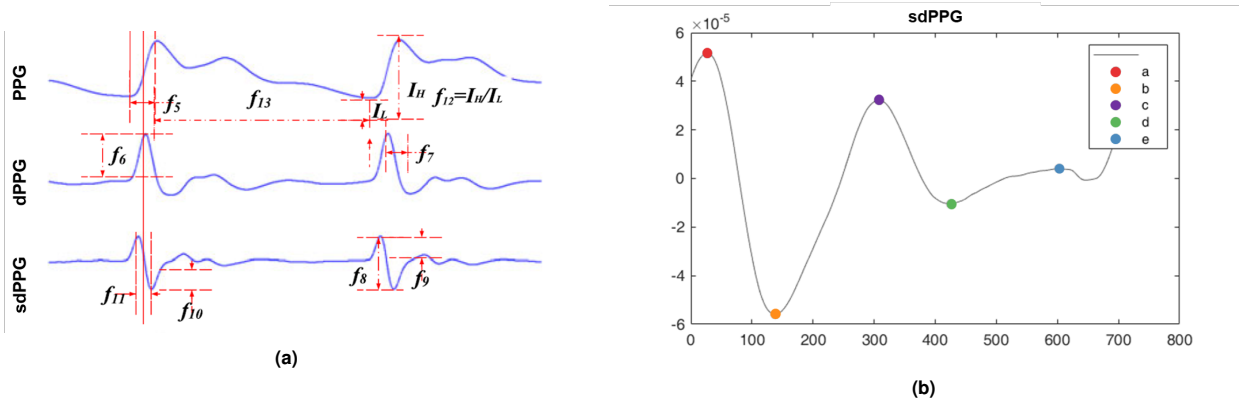

Figure 5: (a) PPG signal (top), PPG signal derivative (middle) and PPG signal second derivative (bottom) with features identified: f5 - Systolic time; f6 – dPPGHeight; f7 – dPPGWidth; f8 – ppgSecondDeriHeight; f9 – sdPPGPeakHeight; f10 – sdPPGFootHeight; f11 – sdPPGWidth; f11 – sdPPGDeriWidth; f12 – PIR; f13 - Diastolic time, adapted from [14]; (b) Waves in the PPG signal second derivative (sdPPG), namely a-wave (early systolic positive wave), b-wave (early systolic negative wave), c-wave (late systolic reincreasing wave), d-wave (late systolic redecreasing wave) and e-wave (early diastolic positive wave). The e-wave represents the dicrotic notch [5]

and inflection point, as defined in table 1 and figure 4.

The augmentation pressure, measured by the augmentation index (AI), is the measure of the contribution that the wave reflection makes to the SBP, and it is obtained by measuring the reflected wave coming from the arterial tree periphery to the centre [5]. This metric is computed as the amplitude ratio of the inflection point to the systolic peak, as defined in table 1 and figure 4.

Systolic time, which is also known as crest time, has been proved useful for cardiovascular disease classification [5]. It is computed as the time difference between systolic peak and the PPG foot, as defined in table 3 and figure 5a. The PPG characteristic value (PPGk) in table 3 is defined as

$$PPGk = \frac{p_m - p_f}{p_s - p_f} \tag{3}$$

where $p_m = \frac{1}{T}\int p_t dt$ and p_t gives the values of one cycle of the PPG signal as a function of time t, T is the duration of the cycle, p_s is the value at the systolic peak and p_f is the value at the foot of the waveform. PPGk was found to be a relevant feature for BP estimation by Miao et al. [14].

The second derivative of the PPG signal (sdPPG), an indicator of the acceleration of the blood in the finger, also contains commonly used features. The sdPPG signal is characterized by several peaks and valleys respectively designated as a-wave (early systolic positive wave), b-wave (early systolic negative wave), c-wave (late systolic reincreasing wave), d-wave (late systolic redecreasing wave) and e-wave (early diastolic positive wave) [5]. The e-wave represents the dicrotic notch [5]. Analysis is usually performed in terms of the amplitudes of the b-, c-, d-, and e-waves with respect to the a-wave amplitude, illustrated in figure 5b. The ratios computed are described in table 3.

The feature extraction process requires several sub-steps, as illustrated in figure 3, which will be described next.

Table 3: Features computed from the PPG for each heart cycle

Feature	Description
Systolic time	Ascending time from PPG foot to PPG peak
dPPGHeight	Intensity of the first derivate of the PPG waveform
dPPGWidth	Time width of the first derivate of the PPG waveform
sdPPGHeight	Total intensity of the second derivate of the PPG waveform
sdPPGPeakHeight	Peak intensity of the second derivate of the PPG waveform
sdPPGFootHeight	Foot intensity of the second derivate PPG waveform
ppgSecondDeriWidth	Time width of the second derivate of the PPG waveform
PIR	Ratio of PPG peak intensity to PPG bottom intensity
Diastolic time	Descending time from PPG peak to PPG foot
PPGk	PPG characteristic value
b/a	ratio of the b-wave to the a-wave in the PPG second derivative (sdPPG)
c/a	ratio of the c-wave to the a-wave in the sdPPG signal
d/a	ratio of the d-wave to the a-wave in the sdPPG signal
e/a	ratio of the e-wave to the a-wave in the sdPPG signal

Beats Identification Systolic peaks are first computed using an automatic multiscale-based peak detection (AMPD) algorithm [20]. The next step is to identify the points that mark the beginning of the PPG cycle (PPG foot), which are defined as the minimum value between two consecutive systolic peaks.

Beats Selection Since some parts of the signal may contain low quality PPG cycles, a method to detect and remove them is necessary.

For this process, the signal is divided in 1 min segments, and the following steps are followed:

- Template creation: a template is computed as average of 30 seconds of beats (T1); The beats of that 30 seconds window that have a correlation coefficient lower than 0.85 with the template are excluded
  - If more than $\frac{1}{4}$ of the beats are excluded, the template is not used (If a previous template is available it is used instead; Else, a new one is computed from the next 30 seconds)
  - Otherwise, a new template is computed with the remaining beats (T2);
- Template comparison: All the beats in the 1 minute window are compared to the template and if the correlation coefficient is lower than 0.95 they are excluded

Computing features To compute the features described above, several fiducial points need to be computed from the ECG, PPG, PPG derivative (dPPG) and sdPPG signals. These include the R peaks in the ECG, peaks in the ECG, which are detected using the Pan-Thompkins algorithm, the PPG foot, the systolic peaks, the max derivative points (dPPG maximum) and inflection points in the PPG signal.

The max derivative point is defined as the time value with greater gradient between the cycle start and the systolic peak.

The inflection point is defined as the largest peak in the PPG gradient between the systolic peak and the start of the next cycle. An example of the points computed is shown in figure 6.

Finally, to avoid the effects of noise, all features are averaged over 10 seconds segments, including blood pressure labels, so that only one value of each systolic blood pressure (SBP) and diastolic blood pressure (DBP) is estimated for each segment.

3.2. Regression algorithm

Random Forest Random forests are an ensemble method which can be used for both classification and regression problems. The algorithm is based on a collection of decision trees [8].

In order to split the data, simple decision rules are learned from the data features. A tree is trained by choosing the best feature to split each node, starting from the root. The best feature to split node m is defined as the one that maximizes the impurity drop of the children nodes, with respect to node m. Several impurity measures are available, such as entropy and Gini index. The one used in this work is the Gini index, according to which the impurity of node m is defined as:

$$i(m) = -\sum_{k=1}^{K} P(k|m)(1 - P(k|m)) \qquad (4)$$

where $P(k|m)$ is the proportion of class k observations in node m [8].

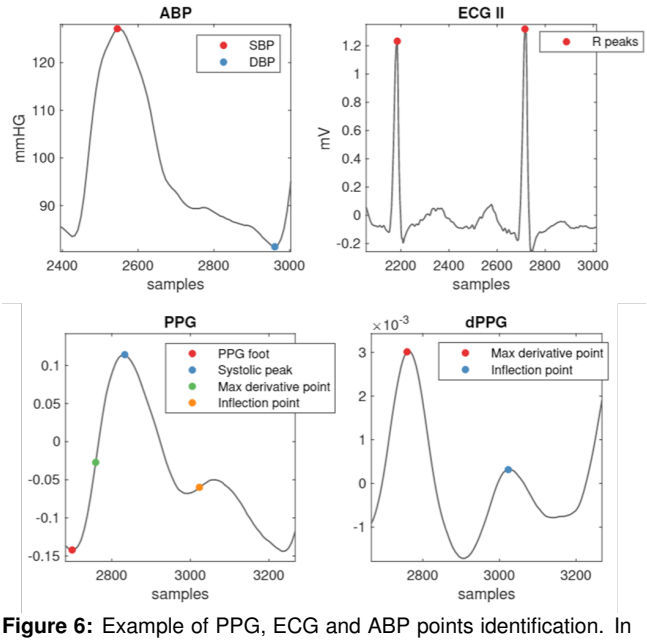

Figure 6: Example of PPG, ECG and ABP points identification. In ABP the systolic (SBP) and diastolic (DBP) BP values are identified. In the ECG signal, the R peaks are identified and in the PPG signal, the PPG foot, Systolic Peak, maximum of the first derivative (Max Derivative point) and Inflection point are identified. In the PPG first derivative signal, the points identified to detect the maximum of the first derivative (Max Derivative point) and Inflection point in the PPG signal are also shown.

In random forests, each tree in the ensemble is built from a sample of the training set generated by bootstrap, that is, by sampling the training set with replacement. If k decision trees are generated, the random forest predictor is then formed by taking the average over k of the trees [1]. When splitting each node during the construction of a tree, the best split is found from a random subset of features. In this implementation, the number of features in the random subset is equal to one third of the total number of features. The purpose of these two sources of randomness is to decrease the variance of the forest estimator, since individual decision trees typically exhibit high variance and tend to overfit. The randomness injected should create decision trees with prediction errors that can be cancelled out by taking an average of the respective predictions.

The Random Forest Regression algorithm was chosen for this implementation due to its robustness to noisy features and outliers. In addition, it has yielded better results than other regression algorithms in studies which implemented several methods for comparison, such as those by Kachuee et al. [9] and Monte-Moreno [16]. Two estimation models were generated, by setting the training targets as SBP and DBP separately. A bagged regression tree algorithm was then used to generate a model, with the min leaf size set to 10 and the number of predictors to 100.

3.3. Model Evaluation

Since the amount of data used is limited, a 7-fold split was performed to evaluate the model. This is a trade-off between the typical test and train set split and a leave-one-subject-out (LOSO) experiment. The latter consists of, in each of n iterations (n = the number of subjects), using the data of n-1 subjects for training, while the data of the left out subject is reserved for testing [21]. The advantages of this method is that the results should not depend on the split choice, which may lead to particularly optimistic or pessimistic results [21]. On the other hand, it highly increases computational complexity. Therefore, the data was instead separated in 7 groups, ensuring that data from each subject was only present in one of the groups. The model was then trained with 6 of the groups created and tested on the remaining group. This procedure is repeated 7 times, allowing all the data to be used for testing.

The agreement of estimated BP with the reference BP was evaluated using Bland–Altman plots, commonly used metrics such as the mean error, mean absolute error and root mean square error and blood pressure measurement standards.

Bland-Altman plots The Bland-Altman plot describes the agreement between two quantitative measurements, which in this case correspond to the true and estimated blood pressure measurements. It consists of a scatter plot where the difference of the two measurements is plotted against their mean value. Its analysis allows the quantification of the agreement between two measurements by studying the mean difference and constructing an agreement interval, within which 95% of the differences between the estimated and true blood pressure measurements fall.

Metrics The mean error (ME), mean absolute error (MAE) and root mean square error (RMSE) are respectively defined as:

$$ME = \frac{1}{n}\sum_{i=1}^{n} y - y^* \tag{5}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y - y^*| \tag{6}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y - y^*)^2} \tag{7}$$

where n is the number of instances, y the true BP value and y^* is the estimated value.

Blood pressure measurement standards Performance was also evaluated based on the Advancement of Medical Instrumentation (AAMI) Standard, which requires BP measurement devices to have ME and standard deviation (STD) values lower than 5 and 8 mmHg, respectively [23]. It additionally requires devices to be evaluated on a statistical population of at least 85 subjects.

4. Results

4.1. State of the art method implementation

The results of implementing the state of the art method by [9] showed a performance inferior to that obtained in such study, as observed in table 5. In this method, 146 subjects from the MIMIC-II database were used for training. For DBP the mean error obtained was $-1.35 \pm 10.72 mmHg$ compared to $0.36 \pm 5.70 mmHg$ and for SBP it was $2.82 \pm 21.96 mmHg$ compared to $-0.06 \pm 9.88 mmHg$ in the original paper. The main reason behind this is the lack of reliability in the feature extraction process for signals with abnormal morphologies. This study does not disclaim how the unsuitable signals are removed, so in the version implemented, all the data described obtained from the MIMIC database were used. It also does not describe any method to reject noisy parts of the signals in which attempting to extract features will lead to abnormal values.

Table 5: Results of implementing the state of the art method for cufless BP estimation by Kachuee et al. [9]

		ME (mmHg)	STD (mmHg)	Subjects
Kachuee [9]	DBP	0.36	5.70	942
	SBP	-0.06	9.88	942
Implemented Version	DBP	-1.3519	10.7213	98
	SBP	2.8224	21.9617	98

4.2. MIMIC II database Total dataset

The regression model was trained and tested on the total data extracted from 146 subjects from the MIMIC database using the 7-fold cross validation process described in the previous section. The data was first trained using only the 10 features described in [9] and in table 1 and figure 4. On a second stage, additional features inspired by other literature studies and physiological parameters that can be extracted from the PPG signal were included, namely those described in table 3. All the results are presented in table 4 and figure 7. All the performance metrics except the RMSE slighly decreased (ME, STD and MAE), pointing that including these extra features improves the algorithm performance.

When comparing the STD value for DBP in table 4 (13.35 mmHg) to that in the implemented version of the algorithm by Kachuee et al., in table 5 (10.72 mmHg), an increase deviation is observed, which is due to the fact that in the original algorithm, only DBP values in the range $60 mmHg \geq DBP \geq 130 mmHg$ are considered, which is a limitation.

4.3. MIMIC II database Good Quality Subset

Although the method developed is able to identify noisy sections of signals, it does not identify the PPG signal portions which do not have the typical shape,

Table 4: Results of implementing the blood pressure estimation method developed using the data obtained from the MIMIC II database

			ME (mmHg)	STD (mmHg)	MAE (mmHg)	RMSE (mmHg)
Total Dataset	10 features (table 1)	DBP	0.76	13.35	10.92	13.52
		SBP	1.22	19.92	16.52	20.30
	24 features (tables 1 and 3)	DBP	0.48	12.76	10.81	13.55
		SBP	1.13	19.53	16.63	20.32
Good Quality Subset	10 features (table 1)	DBP	11.55	6.34	15.03	16.18
		SBP	0.51	9.27	20.58	22.44
	24 features (tables 1 and 3)	DBP	5.20	5.13	10.61	11.82
		SBP	1.70	7.98	17.38	19.18

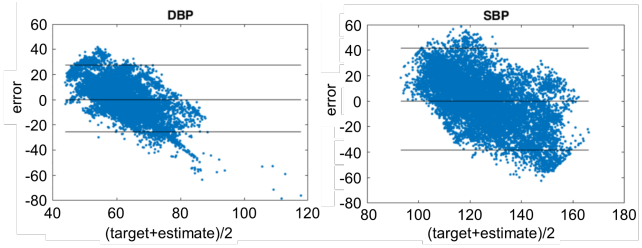

Figure 7: Bland-Altman plot for SBP and DBP estimation using the 24-features model on 146 subjects data from the MIMIC II database

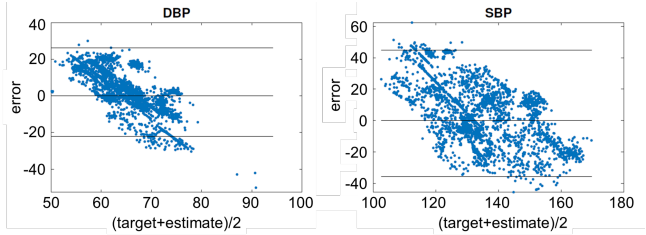

Figure 9: Bland-Altman plot for SBP and DBP estimation using the 24-features model on 21 subjects data from the MIMIC II database

and the second wave and dichrotic notch are not detectable. This typically occurs in older patients, due to the increased artery stiffness, which are commonly monitored in the general ward. An example of this type of PPG signal is shown if figure 10b. It can be observed that there is no prominent maximum between two consecutive *Max derivative* points in the dPPG signal in figure 10b. Therefore, a substudy was conducted by training and testing the model on a small subset of data from 21 subjects, in which the sinals were visually considered to have good quality for feature extraction, as the one shown in figure 10a. The distribution of parameters in this subset is shown in figure 8. The model was trained using the LOSO approach previously described, due to the reduced number of subjects present in the subset.

The model was trained with the initial 10 features in table 1 and with both those features added to the ones in table 3 (the 24 feature model used for the total dataset). The results obtained are shown in table 4 and the Bland-Altman plots in figure 9 for the best performing method, with 24 features.

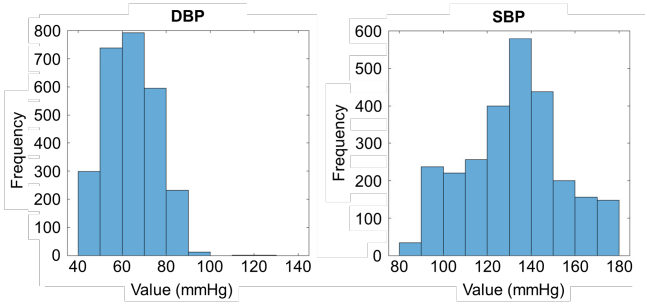

Figure 8: Distribution of parameters for 21 subjects: systolic blood pressure (SBP) and diastolic blood pressure (DBP)

4.4. WARD data

Similar steps to those described in section 3 were also applied to WARD project data. Since invasive ABP measurements are not recorded in the context of the WARD project, standard cuff-based blood pressure measurements are used instead as label for the blood pressure estimation algorithm. There were two sets of PPG data available. The first, corresponding to 231 subjects, was acquired with 8-bit ADC, whereas the second, corresponding to other 108, was acquired with 12-bit ADC.

As in the total MIMIC II dataset, the STDs of the estimation errors did not comply with the AAMI sandard for SBP and DBP estimation. However, the PPG data with higher resolution (12 bit ADC) allowed improved results compared to the PPG data with lower resolution (8 bit ADC), as expected. In addition, the application to the WARD data of the data selection and cleaning procedures described resulted in a large decrease in the number of measurements available for training and testing the BP estimation algorithm. This was less accentuated in the higher resolution dataset.

5. Discussion

5.1. Available data

Data from the MIMIC database has the advantage of having invasive ABP waveforms available, from which SBP and DBP values can be extracted with high precision given that it is the standard for validation of new BP measurement technologies. However, this data is mainly obtained from critically ill patients, having possible medical conditions that alter the PPG waveform, making the extraction process of a number of features less accurate. In particular, for people with arterial stiffness or stage III hypertension as may occur in the older population, the PPG waveform may lose some important features such as its dicrotic notches, making the estimation less precise [25]. As observed in

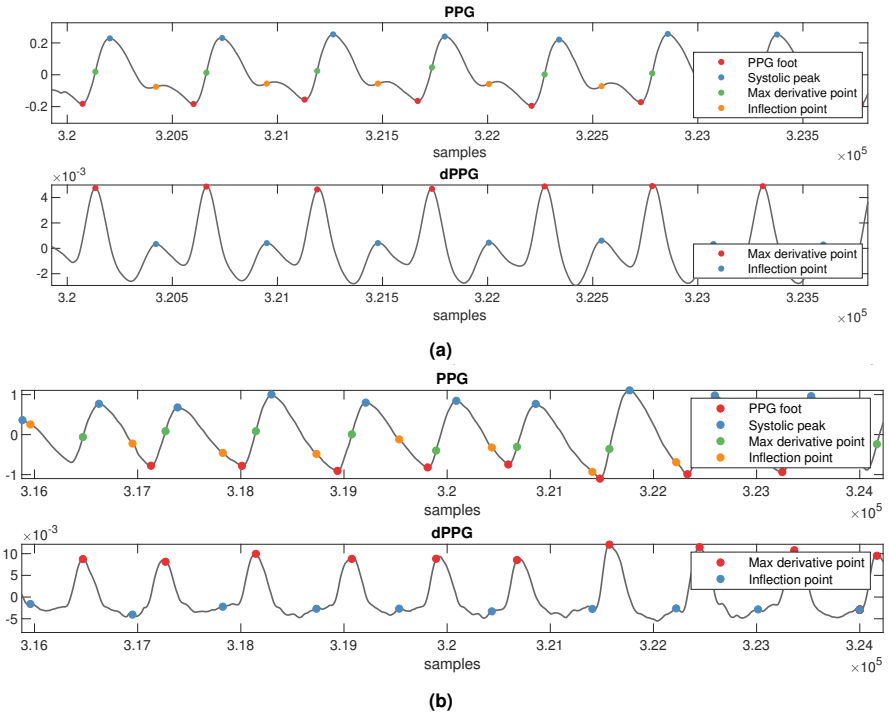

Figure 10: Patient records containing a segment of PPG signal and the respective first derivative (dPPG). The first derivative of the PPG signal shows the peaks corresponding to the *Inflection point* and *Max derivative point*, the maximum of the first derivative in each PPG cycle; (a) good quality segment, in which the dichrotic notch in the PPG signal is easily detected. (b) abnormal record, in which the dichrotic notch in the PPG signal is not easily distinguished.

figure 10, this prevents an accurate detection of the inflection points and features based on these will be unreliable. Many elderly patients are also monitored by the WARD system. Therefore these features show little potential for continuous SBP and DBP estimation in the scope of the WARD project, unless an alternative way to estimate the dicrotic notch is developed.

On the other hand, there are few measurements with low or high blood pressure values, as observed in the histogram in figure 1, making it difficult to obtain accurate measurements in the limits of the physiological range. In particular, the ABP data do not provide values above 180 mmHg. Another problem that affects the performance of the BP estimation model is that the ECG and PPG signals are often not synchronized, making the PAT features inaccurate.

5.2. Data pre-processing and shut down algorithm

The use of a clinical database implies some difficulties such as a high presence of noise and motion artifacts, and the possibility that not all of the signals are available for the whole duration of the records. This created the need for several data pre-processing steps. The aim was to maintain only the PPG cycles that are not influenced by motion artifacts and other conditions that change the typical shape of the cycles. It was observed that when features are computed after sSQI selection and template cleaning, less outliers are obtained.

5.3. Comparison with State of the Art Studies

In table 6, a comparison between this work and some state of the art methods is shown. It is generally difficult to compare results from the various state of the art studies, due to different evaluation metrics and varied datasets whose characteristics are often not specified. In addition, it should be taken into account that the methods requiring calibration generally have lower errors. Particularly, the Ding method shows a error of only $-0.40 \pm 7.11 mmHg$ for DBP and $1.17 \pm 5.72 mmHg$, being the only method of those stated that complies with the AAMI standard regarding STD. However, one of the problems of calibration is that its accuracy may deteriorate over time, and the intervals at which a new calibration is necessary are not studied.

Lower errors are also observed in small selected subsets of data while work including large scale data has larger errors, which was also observed in this project. This highlights the difficulty of creating a robust general model on a large dataset, possibly containing subjects with varied characteristics. Another reason for this problem, pointed by Slapnicar et al.[21], is that the large MIMIC II dataset may as well contain data from different PPG and ABP measurement devices.

All the results obtained in this project meet the AAMI standard regarding mean error (ME) except DBP for the good quality subset in the MIMIC II database. In this particular case ME was $5.20 mmH$ which is only slightly above the limit of the standard ($5 mmHg$). This

Table 6: Comparison with state of the art methods.

	DBP			SBP			Number of test subjects
	ME $\pm$ STD (mmHg)	MAE (mmHg)	RMSE (mmHg)	ME $\pm$ STD (mmHg)	MAE (mmHg)	RMSE (mmHg)	
Salpnicar et al. [21]	-	13.62	-	-	18.34	-	510
Xing et al. [25]	2.6 ± 9.3	-	-	5.5 ± 15.5	-	-	1532
Ding et al. [3]	0.40 ± 7.11	-	-	1.17 ± 5.72	-	-	33
Kachuee et al. [9]	0.36 ± 5.70	-	-	-0.06 ± 9.88	-	-	942
Kachuee et al. implemented version	-1.35 ± 10.72	5.83	-	2.82 ± 21.96	11.80	-	942
MIMIC II data	5.20 $\pm$ 5.13	10,61	11.82	1.70 $\pm$ 7.98	17.38	19.18	21
MIMIC II data	0.48 $\pm$ 12,76	10,81	13,55	1.13 $\pm$ 19,53	16,63	20,32	146
AAMI standard	$\leq 5 \pm 8$	-	-	$\leq 5 \pm 8$	-	-	85

could be due to the distribution of DBP measurement values, observed in figure 8. As for standard deviation (STD), only in the small subset the values are inferior to the AAMI standard of $8mmHg$.

6. Conclusions

It was found that there are many recent developments in the field of cuffless blood pressure estimation although the ones validated in a sufficiently large number of subjects do not yet meet the standards for clinical application. In addition, the methods achieving better performances often include a calibration procedure requiring a variation in the subject's blood pressure. This is not feasible in cases such as the one studied in this project, in which the subjects are hospitalized. The most relevant studies were therefore selected based on performance results, number of subjects in which they were evaluated and the absence of calibration.

Although the state of the art methods for cuffless blood pressure estimation methods based on physiological models relating PAT or PTT features to BP have shown promising performance, the implementation of any of these was not feasible due to lack of synchronization between ECG and PPG signals. The main improvements to the current methods implemented and tested were the extra features extracted from the PPG morphology and the steps of data preprocessing allowing to obtain only clean sections of the PPG signal, which lead to an improvement in the estimation performance.

Having a method that allows blood pressure estimation would extremely be valuable, particularly in the setting of the WARD project, in which patients need to have blood pressure evaluated often. The research conducted shows potential for estimation of BP without a cuff based device. However, the event classes requiring BP measurements, namely hypotension, circulatory failure and hypertension are defined based on SBP in the limits of the physiological ranges, where the estimation algorithm shows decreased performance. Therefore, it is necessary to ensure the accuracy of the method, particularly in the extreme values of blood pressure before this algorithm is to be applied to the WARD system

6.1. Future Work

One of the crucial points in the development of an accurate algorithm for the estimation of blood pressure, is to have an effective shut down algorithm. Artifacts are one of the weaknesses in using the PPG for diagnosis, since the noise can limit the reliability and practical implementation of real-time monitoring applications [6]. Artifacts can result in loss of data, inaccurate readings, and false alarms, affecting the accuracy of pulse oximetry.

In the context of blood pressure estimation, it is important to identify and reject noisy and low quality PPG segments to prevent inaccurate BP estimation. Although random forest regression is robust to noisy features, it is important to decrease their occurrence since it affects the performance of the trained model. Other metrics such as kurtosis and perfusion have been proposed for the estimation of signal quality in the PPG signal [6]. Therefore, a combination of these metrics could be a way to improve the detection of low quality segments. It is also important to tune the thresholds of the metrics used to separate good quality and low quality PPG, for which a dataset of PPG segments labeled by experts would be necessary.

Another point to be improved is the automatic detection of PPG fiductial points. The systolic peaks are detected with the AMPD algorithm, which is robust to artifacts. However, the determination of the remaining fiductial points would be more accurate if more robust methods were used.

References

[1] L. Breiman. Random forests. *Machine Learning*, 45:5–32, 2001.

[2] G. D. Clifford, D. J. Scott, and M. Villarroel. User guide and documentation for the MIMIC II

database. *MIMIC-II database version*, 2(January 2009), 2009.

[3] X. Ding, B. P. Yan, Y. T. Zhang, J. Liu, N. Zhao, and H. K. Tsang. Pulse Transit Time Based Continuous Cuffless Blood Pressure Estimation: A New Extension and A Comprehensive Evaluation. *Scientific Reports*, 7(1):1–11, 2017.

[4] X. R. Ding, Y. T. Zhang, J. Liu, W. X. Dai, and H. K. Tsang. Continuous Cuffless Blood Pressure Estimation Using Pulse Transit Time and Photoplethysmogram Intensity Ratio. *IEEE Transactions on Biomedical Engineering*, 63(5):964–972, 2016.

[5] M. Elgendi. On the Analysis of Fingertip Photoplethysmogram Signals. *Current Cardiology Reviews*, 8(1):14–25, 2012.

[6] M. Elgendi. Optimal signal quality index for photoplethysmogram signals. *Bioengineering*, 3(4):1–15, 2016.

[7] M. Elgendi, R. Fletcher, Y. Liang, N. Howard, N. H. Lovell, D. Abbott, K. Lim, and R. Ward. The use of photoplethysmography for assessing hypertension. *npj Digital Medicine*, 2(1):1–11, 2019.

[8] T. Hastie, J. Friedman, and R. Tisbshirani. *The Elements of statistical learning: data mining, inference, and prediction*. Springer, 2017.

[9] M. Kachuee, M. M. Kiani, H. Mohammadzade, and M. Shabany. Cuffless Blood Pressure Estimation Algorithms for Continuous Health-Care Monitoring. *IEEE Transactions on Biomedical Engineering*, 64(4):859–869, 2017.

[10] A. K. Khanna, P. Hoppe, and B. Saugel. Automated continuous noninvasive ward monitoring: Future directions and challenges. *Critical Care*, 23(1):1–5, 2019.

[11] Y. Kurylyak, F. Lamonaca, and D. Grimaldi. A Neural Network-based method for continuous blood pressure estimation from a PPG signal. *Conference Record - IEEE Instrumentation and Measurement Technology Conference*, pages 280–283, 2013.

[12] Y. Liang, D. Abbott, N. Howard, K. Lim, R. Ward, and M. Elgendi. How Effective Is Pulse Arrival Time for Evaluating Blood Pressure? Challenges and Recommendations from a Study Using the MIMIC Database. *Journal of Clinical Medicine*, 8(3):337, 2019.

[13] B. M. McCarthy, C. J. Vaughan, B. O'Flynn, A. Mathewson, and C. Ó. Mathúna. An examination of calibration intervals required for accurately tracking blood pressure using pulse transit time algorithms. *Journal of Human Hypertension*, 27(12):744–750, 2013.

[14] F. Miao, N. Fu, Y. T. Zhang, X. R. Ding, X. Hong, Q. He, and Y. Li. A novel continuous blood pressure estimation approach based on data mining techniques. *IEEE Journal of Biomedical and Health Informatics*, 21(6):1730–1740, 2017.

[15] F. Michard, D. I. Sessler, and B. Saugel. Non-invasive arterial pressure monitoring revisited. *Intensive Care Medicine*, 44(12):2213–2215, 2018.

[16] E. Monte-Moreno. Non-invasive estimate of blood glucose and blood pressure from a photoplethysmograph by means of machine learning techniques. *Artificial Intelligence in Medicine*, 53(2):127–138, 2011.

[17] R. Mukkamala, J. O. Hahn, O. T. Inan, L. K. Mestha, C. S. Kim, H. Toreyin, and S. Kyal. Toward Ubiquitous Blood Pressure Monitoring via Pulse Transit Time: Theory and Practice. *IEEE Transactions on Biomedical Engineering*, 62(8):1879–1901, 2015.

[18] M. Radha, K. De Groot, N. Rajani, C. C. Wong, N. Kobold, V. Vos, P. Fonseca, N. Mastellos, P. A. Wark, N. Velthoven, R. Haakma, and R. M. Aarts. Estimating blood pressure trends and the nocturnal dip from photoplethysmography. *Physiological Measurement*, 40(2):1–20, 2019.

[19] J. C. Ruiz-Rodríguez, A. Ruiz-Sanmartín, V. Ribas, J. Caballero, A. García-Roche, J. Riera, X. Nuvials, M. De Nadal, O. De Sola-Morales, J. Serra, and J. Rello. Innovative continuous non-invasive cuffless blood pressure monitoring based on photoplethysmography technology. *Intensive Care Medicine*, 39(9):1618–1625, 2013.

[20] F. Scholkmann, J. Boss, and M. Wolf. An efficient algorithm for automatic peak detection in noisy periodic and quasi-periodic signals. *Algorithms*, 5(4):588–603, 2012.

[21] G. Slapni Č Ar, N. Mlakar, and M. Luštrek. Blood pressure estimation from photoplethysmogram using a spectro-temporal deep neural network. *Sensors (Switzerland)*, 19(15), 2019.

[22] J. Solà and R. Delgado-Gonzalo. *The Handbook of Cuffless Blood Pressure Monitoring, A Practical Guide for Clinicians, Researchers, and Engineers*. Springer Nature Switzerland AG, 2019.

[23] G. S. Stergiou, B. Alpert, S. Mieke, R. Asmar, N. Atkins, S. Eckert, G. Frick, B. Friedman, T. Graßl, T. Ichikawa, J. P. Ioannidis, P. Lacy, R. McManus, A. Murray, M. Myers, P. Palatini, G. Parati, D. Quinn, J. Sarkis, A. Shennan, T. Usuda, J. Wang, C. O. Wu, and

E. O'Brien. 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.

[24] J. X. Sun, A. T. Reisner, and R. G. Mark. A signal abnormality index for arterial blood pressure waveforms. *Computers in Cardiology*, 33:13–16, 2006.

[25] X. Xing, Z. Ma, M. Zhang, Y. Zhou, W. Dong, and M. Song. An Unobtrusive and Calibration-free Blood Pressure Estimation Method using Photoplethysmography and Biometrics. *Scientific Reports*, 9(1):1–8, 2019.