A Machine Learning Approach to Predict Arterial Blood Pressure from Photoplethysmography Signal

Felipe M Dias[1,2], Thiago B S Costa[1,3], Diego A C Cardenas[1],
Marcelo A F Toledo[1], Jose E Krieger[1] and Marco A Gutierrez[1]

[1] Heart Institute, Clinics Hospital, University of Sao Paulo Medical School
[2] Polytechnique School, University of Sao Paulo
[3] Federal University of ABC

Abstract

Blood pressure (BP) monitoring is a basic procedure for the physiological measurement of the cardiovascular system, especially because high BP, although preventable, is a major risk for stroke, heart failure, and other serious conditions. Photoplethysmography (PPG) is a promising technology developed to allow non-invasive, regular, or even continuous measurement of blood volume variation. Recently, some works have tried to use PPG signals to estimate BP. In this work, we propose a regression model based on the Category Boosting algorithm (CatBoost) that uses 133 morphological and temporal features from the PPG signal to estimate the corresponding diastolic and systolic BP. We processed and selected a total of 50,182 windows of 1,000 samples (sampling rate of 125Hz during 8 seconds) of PPG and BP signals from the MIMIC-II dataset, distributed into training and test sets. Three different data cross-validation schemes were adopted. The model prediction metrics were evaluated by Mean Error and standard deviation (ME[STD]), and Pearson's Correlation Coefficient (R-value). For one of the validation schemes, we obtained, for the diastolic BP, 0.02[3.77] mmHg with an R-value of 0.93; and for systolic BP: 0.05[7.84] mmHg with an R-value of 0.93. Our results meet the AAMI standard and are comparable to the state of the art. However, we show that these results rely on a specific validation scheme.

1. Introduction

Hypertension is a major health issue, being among the highest-ranking death causes worldwide [1], increasing the risk of heart failure, stroke, and myocardial infarction. As these outcomes often follow a long period of hypertension development [2], arterial blood pressure (BP) assessment is not only critical for diagnosis but its monitoring enables early lifestyle changes and preventive clinical treatment, reducing the risks of such outcomes. The gold standard for BP assessment is the use of a sphygmomanometer by trained clinical professionals. For non-technical measurements, there are many automatically cuff-based equipment available, enabling people to have measurements at home. However, cuff-based devices can cause discomfort if used frequently and are unfeasible for continuous monitoring.

Photoplethysmography (PPG) has originally been developed for measuring tissue blood volume, and later adapted for SpO_2 measurement (oximetry) [3]. However, a direct way to obtain BP from the PPG signal alone is still unknown and its use for BP estimation is still a challenge.

Several machine learning approaches have been proposed for assessing BP from PPG. Some methods deal with the raw PPG signal after filtering, usually using 1D convolutional neural networks [4, 5] or time-series processing architectures, such as LSTM [6]. Other methods rely on manual feature extraction and a regression method, e.g., multilayer perceptron [7] and random forest [4]. Results are encouraging, however, data distribution and validation conditions are unclear in many works in the literature.

In order to estimate BP from the PPG signal, we proposed a regression model based on the Category Boosting Algorithm (CatBoost) using a combination of 133 morphological and temporal features and evaluated the performance according to three different validation schemes.

2. Materials And Methods

The Multi-Parameter Intelligent Monitoring in Intensive Care II (MIMIC-II) database [8] contains continuous recordings of several physiological signals obtained from Intensive Care Unit (ICU) patients of the Beth Israel Deaconess Medical Center (BIDMC) from 2001 to 2008.

Kachuee *et al* [9] proposed a subset of the MIMIC II dataset: Cuff-Less Blood Pressure Estimation Data Set (CLBPE). This dataset was specifically designed for BP estimation purposes, containing PPG, Arterial Blood Pres-

sure (ABP), and Electrocardiogram (ECG) signals sampled at 125 Hz. In this dataset, as described in [9], all signals were pre-processed, and regions with unrealistic BP or heart rate values were discarded. There are 12,000 recordings that were equally split in four files ("Part_1.mat", "Part_2.mat", "Part_3.mat", "Part_4.mat"). A major limitation of this dataset is the absence of patient information. It can be inferred from Kachuee *et al.* [9,10] that the CLBPE dataset contains 851 patients, but there is no information on patient identification of each record in the dataset.

2.1. Signal Pre-processing

In addition to the signal pre-processing performed on the CLBPE dataset [9], we have filtered the PPG signals using a Type-II Chebyshev band-pass filter (0.5Hz-10Hz) of order 4. Also, we removed recordings that were smaller than 8 minutes. Next, we segmented each recording into windows of 1,000 samples (sampling rate of 125 Hz during 8 seconds) without overlap. Then, we used a signal quality analysis of each PPG and ABP windows. For this, in each window, we detected every beat and estimated a "template" average beat. We calculated Pearson's correlation coefficient between each beat and the template. Then, we excluded the window if its mean correlation was lower than a threshold (0.9). Next, we removed all windows for which (a) systolic pressure was $\leq$70 mmHg, (b) diastolic pressure was $\geq$141 mmHg or (c) the difference between systolic and diastolic pressure $\leq$10 mmHg.

We obtained the systolic and diastolic pressure from the ABP signal in each window using a peak and trough detector [11]. The median of the peak and trough values of the ABP signal were considered the label for systolic and diastolic pressure, respectively. After these steps, we obtained a total of 50,182 PPG windows with an associated systolic and diastolic pressure.

2.2. Feature extraction

Several characteristic points of each 8-second PPG window —and its first and second derivatives—are detected through a procedure adapted from [12]. These are used to calculate morphological and temporal attributes, which represent slopes, areas, intensities, amplitudes, and time spans, according to descriptions provided by [13], resulting in 133 features for each PPG window.

2.3. Method Validation

We employed two boost-based machine learning methods for estimating the BP of each window: CatBoost and XGBoost.

In the literature, numerous works employed 10-fold cross-validation strategies to evaluate their BP estimation algorithms using the CLBPE dataset [6, 14]. However, as aforementioned, this dataset does not contain patient identification and a single patient can have several recordings. Therefore, regardless of the fold validation strategies employed, the data distribution can present samples from the same patient both in the training and testing sets, i.e., data leakage. On the other hand, the CLBPE dataset is split into 4 different ".mat" files. Assuming that a patient can not appear in multiple parts (".mat" files), we can use a 4-fold cross-validation scheme, using each file as a fold, to diminish the data leakage problem.

Hence, three validation schemes were proposed. In the first one, we randomly split the windows in a 10-fold cross-validation (WINDOW). In the second, we split the windows into 10-folds taking care that windows from the same recording did not appear in different folds (RECORD). In the last one, we split the windows into 4 folds in a way that windows from the same ".mat" did not appear in different folds (PART).

3. Results and Discussion

The BP estimation performance was evaluated using Pearson's Correlation Coefficient (ρ), Mean Absolute Error (MAE), Mean Error (ME) and its standard deviation of the error (STD) metrics (Table 1). According to the AAMI standard [15], ME and STD should be lower than 5 mmHg and 8 mmHg, respectively.

Furthermore, we used the British Hypertension Society (BHS) guideline [16] (Table 2). In such metric, we evaluated the percentage of the MAE error that falls below 5 mmHg, 10 mmHg, and 15 mmHg. A grade (A, B, C, or D) is given to the estimation method depending on how well it performs.

Figure 1 shows the scattering of the target versus prediction, (a) and (b), as well as the Bland-Altman plot, (c) and (d), for SBP and DBP estimation using CatBoost, by considering PART, RECORD, and WINDOW validation schemes.

With regard to the AAMI standard and MAE (Table 1), both algorithms achieved similar performances in different validation schemes, even though CatBoost had a slight advantage. Nevertheless, they also exhibited a substantial drop in performance as the validation scheme goes from WINDOW to PART. With CatBoost, for instance, using the WINDOW validation scheme, the STDs for DBP and SBP estimation were 3.77 and 7.88, respectively. However, if we change to the PART validation scheme, the STDs for DBP and SBP improves to 9.78 and 22.37, respectively.

In accordance with the BHS standard (Table 2), the results confirm that XGBoost and CatBoost are indeed analogous regarding the validation scheme, with the latter showing slightly better values than the former. However, both reveal, again, a decrease in performance as our val-

Table 1: Evaluation of BP estimation in the MIMIC-II dataset using the AAMI standard.

Validation scheme	Algorithm	Diastolic blood pressure				Systolic blod pressure			
		ME	ρ	**STD**	**MAE**	**ME**	ρ	**STD**	**MAE**
WINDOW	XGBoost	**0.012**	0.891	4.470	3.091	0.059	0.893	9.485	6.721
	CatBoost	0.022	**0.925**	**3.767**	**2.521**	**0.050**	**0.929**	**7.837**	**5.368**
RECORD	XGBoost	**-0.173**	0.604	7.962	5.653	**-0.852**	0.622	16.648	12.463
	CatBoost	-0.194	**0.638**	**7.618**	**5.418**	-1.166	**0.661**	**15.822**	**11.774**
PART	XGBoost	-0.381	0.297	10.095	7.658	**-0.746**	**0.219**	22.778	18.495
	CatBoost	**-0.257**	**0.306**	**9.784**	**7.524**	-1.230	0.218	**22.368**	**18.209**
AAMI standard		$\leq$ 5 mmHg	-	$\leq$ 8 mmHg	-	$\leq$ 5 mmHg	-	$\leq$ 8 mmHg	-

Table 2: Evaluation of BP estimation in the MIMIC-II dataset with regard to BHS standard.

Validation scheme	Algorithm	DBP cumulative error (%)			SBP cumulative error (%)		
		5mmHg	**10mmHg**	**15mmHg**	**5mmHg**	**10mmHg**	**15mmHg**
WINDOW	XGBoost	81.6	96.5	99.0	52.2	78.5	89.9
	CatBoost	**87.5**	**97.7**	**99.4**	**62.5**	**85.7**	**93.7**
RECORD	XGBoost	57.2	84.6	94.6	29.9	52.4	68.3
	CatBoost	**59.2**	**85.4**	**95.1**	**32.4**	**55.1**	**70.7**
PART	XGBoost	**41.5**	72.0	89.2	16.0	31.7	46.8
	CatBoost	40.7	**73.2**	**90.1**	**16.2**	**32.3**	**47.6**
BHS standard	Grade A	60	85	95	60	85	95
	Grade B	50	75	90	50	75	90
	Grade C	40	65	85	40	65	85
	Grade D	<40	<65	<85	<40	<65	<85

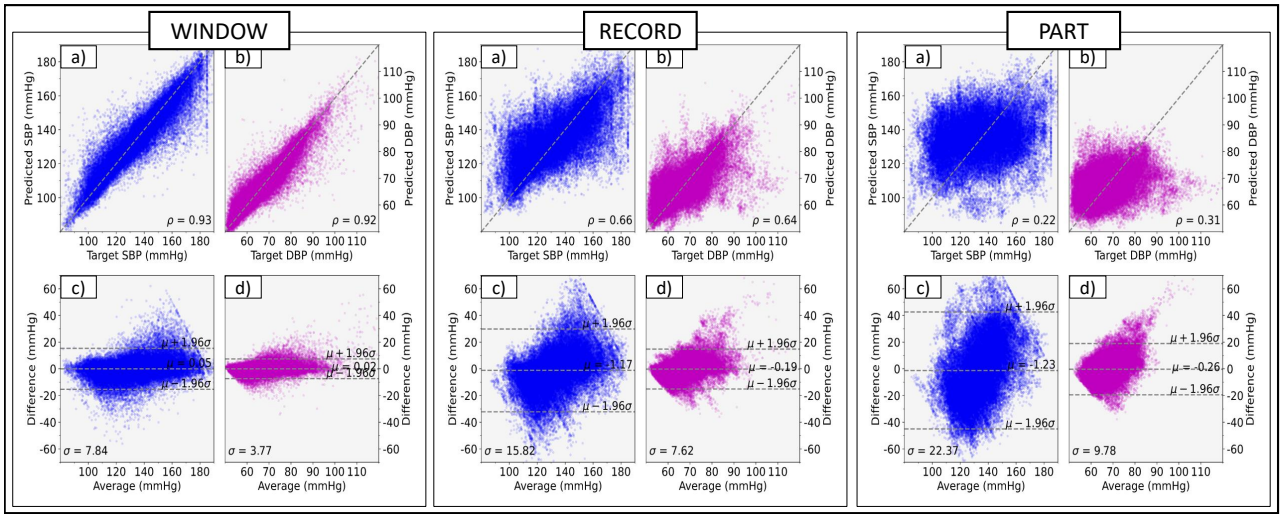

Figure 1: Scattering of the target versus prediction, (a) and (b) and Bland-Altman plot, (c) and (d), for SBP and DBP estimation using CatBoost in the MIMIC-II database, considering validation schemes WINDOW, RECORD, and PART. μ: mean; σ: standard deviation; ρ: correlation coefficient.

idation scheme goes from WINDOW to PART. CatBoost, for example, is classified into Grade A for DBP and Grade B for SBP in WINDOW, but only into Grade C and D in PART. This performance reduction when using the PART validation scheme may indicate that recordings of the same patient are restricted to the same fold and do not appear in the other folds.

The same behavior can be observed in the Pearson's correlation (Figure 1). Although samples of the WINDOW validation scheme have correlation coefficients of 0.93 and 0.92 for SBP and DBP, respectively, samples of PART and RECORD validation schemes have only 0.22 and 0.66 for SBP and 0.31 and 0.64 for DBP. Moreover, the Bland-Altman intervals of agreement get wider as val-

idation scheme goes from WINDOW to PART.

Concerning other works in the literature that also use the MIMIC II dataset, Panwar et al [6] used a 10-fold cross-validation approach and obtained results that meet the AAMI and the BHS standard criterion. Meanwhile, El-Hajj et al. [14], employed a 70%-15%-15% split for training, validation, and test set, respectively, and got results near the AAMI standard criterion. However, these works do not inform if the data was randomly split or split by recording, which does not allow a direct and fair comparison with our work. We did not find any works in the literature that employed PART cross-validation.

4. Conclusion

We employed a feature-based machine learning method to predict BP from PPG signals and evaluated it using three different cross-validation schemes in a pre-processed MIMIC II dataset (CLBPE). Only in the WINDOW validation scheme, we were able to meet the AAMI standard requirements and grades A and B (BHS standard) for both systolic and diastolic pressure.

Results under validation schemes WINDOW and RECORD reveal a possible undesirable data leakage effect that might lead to over-optimistic results that may not hold in real-life applications. As observed, the pre-exposition of a learning algorithm to information about the same patient can lead to better results. Therefore, we recommend that future works using the CLBPE dataset present their results using our proposed three-scheme cross-validation approach: WINDOW, RECORD, and PART. This would benefit comparison between different methodological approaches, clarifying their contributions and improvements.

Acknowledgments

This work was supported by Foxconn Brazil and Zerbini Foundation as part of the research project "Machine Learning in Cardiovascular Medicine".

References

[1] Zhou B, Carrillo-Larco RM, et al. Worldwide trends in hypertension prevalence and progress in treatment and control from 1990 to 2019: a pooled analysis of 1201 population-representative studies with 104 million participants. The Lancet 2021;398(10304):957–980.

[2] Olsen MH, Angell SY, et al. A call to action and a lifecourse strategy to address the global burden of raised blood pressure on current and future generations: the lancet commission on hypertension. The Lancet 2016;388(10060):2665–2712.

[3] Reisner A, Shaltis PA, et al. Utility of the photoplethysmogram in circulatory monitoring. The Journal of the American Society of Anesthesiologists 2008;108(5):950–958.

[4] Slapničar G, Mlakar N, et al. Blood pressure estimation from photoplethysmogram using a spectro-temporal deep neural network. Sensors 2019;19(15):3420.

[5] Sharma M, Barbosa K, et al. Cuff-less and continuous blood pressure monitoring: a methodological review. Technologies 2017;5(2):21.

[6] Panwar M, Gautam A, et al. Pp-net: A deep learning framework for ppg-based blood pressure and heart rate estimation. IEEE Sensors Journal 2020;20(17):10000–10011.

[7] Kurylyak Y, Lamonaca F, et al. A neural network-based method for continuous blood pressure estimation from a ppg signal. In 2013 IEEE International instrumentation and measurement technology conference (I2MTC). IEEE, 2013; 280–283.

[8] Goldberger AL, Amaral LA, et al. Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. circulation 2000; 101(23):e215–e220.

[9] Kachuee M, Kiani MM, et al. Cuff-less high-accuracy calibration-free blood pressure estimation using pulse transit time. In 2015 IEEE international symposium on circuits and systems (ISCAS). IEEE, 2015; 1006–1009.

[10] Kachuee M, Kiani MM, et al. Cuffless blood pressure estimation algorithms for continuous health-care monitoring. IEEE Transactions on Biomedical Engineering 2016; 64(4):859–869.

[11] Bishop SM, Ercole A. Multi-scale peak and trough detection optimised for periodic and quasi-periodic neuroscience data. In Intracranial Pressure & Neuromonitoring XVI. Springer, 2018; 189–195.

[12] Hsu YC, Li YH, et al. Generalized deep neural network model for cuffless blood pressure estimation with photoplethysmogram signal only. Sensors 2020;20(19):1–19.

[13] Lin WH, Li X, et al. Investigating the physiological mechanisms of the photoplethysmogram features for blood pressure estimation. Physiological Measurement 2020;41(4):1–13.

[14] El-Hajj C, Kyriacou PA. Cuffless blood pressure estimation from ppg signals and its derivatives using deep learning models. Biomedical Signal Processing and Control 2021; 70:102984.

[15] White WB, Berson AS, et al. National standard for measurement of resting and ambulatory blood pressures with automated sphygmomanometers. Hypertension 1993; 21(4):504–509.

[16] O'Brien E, Petrie J, et al. The british hypertension society protocol for the evaluation of blood pressure measuring devices. Journal of Hypertension 1993;.

Address for correspondence:

Felipe M Dias
Heart Institute (InCor) - Av. Dr. Enéas Carvalho de Aguiar, 44 - Cerqueira César, São Paulo - SP, Brazil
f.dias@hc.fm.usp.br