Generalization Capability of a Neural Network for Blood Pressure Estimation from Photoplethysmography

Clémentine Aguet[1,2], Jérôme Van Zaen[1], Martin Proença[1], Guillaume Bonnier[1], Pascal Frossard[2], Mathieu Lemay[1]

[1]Swiss Center for Electronics and Microtechnology (CSEM), Neuchâtel, Switzerland
[2]Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

Abstract

Photoplethysmography (PPG) is one of the most promising alternatives for non-invasive and cuffless blood pressure (BP) monitoring. In recent years, several machine learning approaches have been considered for this task: either feature engineering-based to map PPG-derived features into BP values or feature learning-based with an automated feature extraction process. The generalization capability, namely the ability of a model to adapt to unseen data, is an important aspect of such data-driven models due to the heterogeneity of PPG waveforms. However, in published studies, this point is generally omitted. Therefore, we propose to assess the generalization capability of a feature learning model built to estimate BP from PPG signals by comparing the model accuracy (bias) and precision (standard deviation of the error) on two datasets with different recording protocols. On unseen subjects from the training dataset, the proposed model achieved mean and standard deviation errors of -0.88±10.29 mmHg for systolic BP (SBP) and –0.60±5.76 mmHg for diastolic BP (DBP). Whereas, on the other dataset, the same metrics were 1.13±12.86 mmHg for SBP and -0.44±6.99 mmHg for DBP. Taken together, these results show that a feature learning model can extract feature representation that are generalizable over different populations and different PPG sensors.

1. Introduction

Hypertension is a serious condition that greatly increases the risk of developing cardiovascular diseases [1]. It affects about one third of the adult population. Moreover, one of the primary concerns with hypertension is the absence of perceptible symptoms in approximately 46% of cases. Continuous blood pressure (BP) monitoring allows the identification of abnormal fluctuations. It would therefore help early detection of hypertension and prevent potential complications. In the effort of developing non-invasive continuous and cuffless BP monitoring devices, photoplethysmography (PPG) has recently gained increased interest. This simple and low-cost optical technology detects blood volume variations in the microvascular bed of tissues. The analysis of the pulse morphology and its derivatives, known as pulse wave analysis (PWA), has revealed characteristic points correlated to cardiovascular parameters and useful in the modelling of BP [2], [3]. Various supervised regression machine learning (ML) methods have already been investigated to translate the PWA-based features into BP values [2], [4]–[6]. In [7], we highlighted the potential of a feature learning approach for BP estimation from PPG signal. With an automatic feature extraction process, we have addressed some of the limitations of feature engineering, such as the computational complexity of some features and the requirement of expert knowledge. The proposed model based on convolutional layers combines the feature learning task with the regression task in a single architecture. It extracts representative information from an ensemble average (EA) pulse computed over PPG windows and estimates systolic (SBP) and diastolic BP (DBP) accordingly.

However, ML methods, such as neural networks (NNs), are data-driven, i.e. they are built based on data available during the training of the model. Such models raise the question of their generalization capability, which describes the ability of a model to adapt to unseen data. This aspect is particularly relevant with the heterogeneity of PPG waveforms, which typically vary according to individual-specific characteristics, the PPG sensor used, or the measurement site. Although crucial to the success of the model and its application, this aspect is, most of the time, not addressed in published studies.

The present work proposes to assess the generalization capability of a PPG-based BP estimation model primarily introduced in [7]. Such a study could help to fill the gap between research and potential real-world use case. To reach this goal, we use PPG signals with associated BP readings from two datasets. Both collected during non-cardiac surgical interventions but with different PPG sensors and including different target populations.

2. Materials and Methods

2.1. Datasets

In this study, two datasets are used to evaluate the generalization capability of the proposed BP estimation model in a context of high BP variability induced by the administration of anaesthetic, analgesic and anti-hypotensive agents. They were collected using different protocols, including various target population (Asian vs. Caucasian) and PPG sensors (finger-clip in transmission mode vs. smartphone in reflectance mode). They were both recorded in operating room (OR), allowing to have the radial arterial BP (ABP) from an invasive catheter as reference.

The largest dataset is retrieved from the open-access VitalDB data bank [8]. It includes the recordings of 6'388 patients admitted at Seoul National University Hospital (Seoul, Republic of Korea) that underwent non-cardiac surgical interventions between June 2016 and July 2017. The study was approved by the local ethics committee (H-1408-101-605, NCT02914444 at ClinicalTrials.gov). We select patients including the original waveforms of ABP and PPG, which results in 3'326 patients. This dataset is referred to as VitalDB in this paper.

The second dataset is from a study at the Lausanne University Hospital (CHUV, Lausanne, Switzerland) in collaboration with the Geneva University Hospital (HUG, Geneva, Switzerland) [9] approved by the local ethics committee (CER-VD no. 2018-01656; NCT03875248 at ClinicalTrials.gov). It includes the recordings of 121 adult patients enrolled from April 2019 to November 2019 and undergoing general anesthesia for diverse non-cardiac surgical reasons. The main difference with VitalDB is in the PPG sensor. In this study, instead of a finger-clip sensor, the camera of a Samsung Galaxy S7 smartphone is used. This dataset is therefore referred to as OR_S7 in the paper. The ABP recording starts at general anesthesia induction and lasts for 20 minutes. For each patient, ten 60-second PPG segments were recorded at 120-second intervals.

2.2. Preprocessing

Due to missing or unreliable values, artifacts, as well as poor-quality signals in the datasets, the following preprocessing steps are applied. The first step consists of removing non-physiological or disconnected ABP signals. Based on [10], each 60-second PPG segment is then aggregated into an EA pulse and the first and second derivatives are computed. The reference SBP and DBP values are obtained by calculating the median over each corresponding 60-second segment of the ABP signal. The segments with too much distortion or noise are excluded based on an EA quality index (≥ 75%) as well as indicators of reference variability (≤ 10%) and reliability (≥ 50%). The EA pulses are finally zero-padded up to the maximum observed cardiac period and resampled to a length of 256.

Some BP characteristics and the patient demographics of the two datasets after preprocessing are summarized in Table 1. For the VitalDB dataset, we end up with 1'492 patients and 166'966 EA pulses, while 101 patients and 2'238 EA pulses remain for the OR_S7 dataset.

Table 1. Demographic and BP characteristics.

Characteristics	VitalDB N = 1492	OR_S7 N = 101
Age *(y)*	59.4 ± 14.0 (18 – 92)	57.8 ± 14.2 (24 – 87)
Height *(cm)*	162.7 ± 8.7 (134.1 – 188.6)	169.8 ± 9.4 (143 – 190)
Weight *(kg)*	61.6 ± 11.5 (35.4 – 139.7)	76.3 ± 17.0 (45 – 152)
Gender, male	821 (55)	51 (50)
SBP average *(mmHg)*	117.2 ± 11.7 (89.5 – 169.1)	117.1 ± 20.3 (84.1 – 199.4)
DBP average *(mmHg)*	62.4 ± 7.6 (41.8 – 90.7)	60.7 ± 9.6 (37.6 – 87.3)
SBP std *(mmHg)*	14.8 ± 4.9 (2.7 – 37.4)	13.3 ± 9.2 (1.0 – 39.9)
DBP std *(mmHg)*	7.9 ± 2.5 (0.9 – 18.7)	6.5 ± 4.1 (0.3 – 22.0)

Data are presented as Mean ± STD (Range) or count (%).

2.3. Model

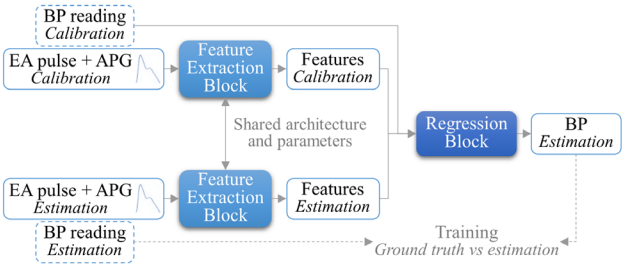

Figure 1. Schematic description of the feature learning model with a PPG-BP calibration measure.

For cuffless BP estimation monitoring devices, a calibration of the system usually helps to adjust for inter-subject variability of the PPG waveform. As illustrated in Figure 1, we opt for a calibration process based on an initial PPG measure with its associated reference BP reading. Each input comprises the PPG measure from which to estimate BP as well as a calibration PPG-BP measure. And each PPG measure combines the EA pulse and its second derivative, which is also called acceleration plethysmogram (APG). To be more representative of a real use case, the calibration measure is always taken from the same recording, within the same inter-flush segment and before the estimation measure. Both feature learning and regression tasks are integrated into the same model to simultaneously extract relevant information from the PPG

and provide a BP estimate.

The two feature extraction blocks in Figure 1 are identical and take in parallel the estimation and calibration measures. Such a block is composed of 4 one-dimensional convolution layers. After each convolution, a rectified linear unit (ReLU) activation function is applied, and a one-dimensional max-pooling layer follows. We then concatenate the features extracted from both estimation and calibration measures together with the calibration BP values. All features are passed to the regression block, which outputs a BP estimation value through a series of 2 fully connected layer with ReLU activation function.

Various aspects of a NN can be adjusted in order to prevent overfitting and improve generalization. Some of them are considered while choosing the model architecture. The first is to constrain the model complexity, as a NN with a high number of parameters, i.e. multiple layers and neurons, is more likely to overfit data. We also apply dropout [11]. The idea is to randomly deactivate a certain number of neurons during the training. At each iteration, the model will learn with a different configuration of neurons. Another technique considered is weight regularization. As a network with large weights is more susceptible to small changes in the input, we constraint the growth of the weights by adding a L2 penalty to the loss function. The last strategy considered is early stopping. One challenge with NNs is to train the model long enough to learn the desired task but not long enough to overfit the training data. By monitoring the performance of the model on the train and validation sets, the idea is to stop before the error on unseen examples starts to increase.

2.4. Training Setting

The model is implemented in Python using the PyTorch deep learning framework. Its hyperparameters are tuned using VitalDB. This dataset is split into train, validation and test sets in a ratio of 6:2:2. To guarantee a similar BP distribution across the three sets, this partitioning is done in a stratified manner, using patient's mean and standard deviation SBP and DBP values. The model parameters are optimized by minimizing the Huber loss function [12] with the Adam algorithm.

2.5. Evaluation

As the main purpose of this study is the evaluation of the generalization capability of the proposed model, the model is evaluated on two aforementioned datasets with different recording protocols, namely the test set of VitalDB and the whole OR_S7 dataset.

The performance of the model in providing an absolute BP estimate is assessed by comparing the proposed feature learning-based method to the invasive reference in a context of general anaesthesia induction. In absence of applicable standard for cuffless BP monitoring, the evaluation is done based on the ISO 81060-2:2018 norm, in terms of accuracy (bias) and precision of agreement (standard deviation of the error). The baseline of comparison is a naïve model assuming no change in BP after calibration.

3. Results

Table 2 and Table 3 summarize the performance of the proposed model on VitalDB dataset for SBP and DBP respectively in terms of ME and STDE in mmHg. The feature learning-based model outperforms the naïve model, with a reduction in STDE of approximately 48.5% for SBP and 48.1% for DBP. These results confirm the model ability to extract relevant information from the PPG and APG waveform relative to BP. The difference of STDE between train and test sets is of 1.24 mmHg for SBP and 1.21 mmHg for DBP. It reveals a reduced overfitting effect and good generalization.

Table 2. SBP estimation performance on VitalDB in mmHg.

	Train		Test	
	ME	STDE	ME	STDE
NN model	-0.95	9.05	-0.88	10.29
Naïve model	-1.35	19.88	-0.80	19.99

Table 3. DBP estimation performance on VitalDB in mmHg.

	Train		Test	
	ME	STDE	ME	STDE
NN model	-0.05	4.55	-0.60	5.76
Naïve model	-0.58	10.16	-0.39	11.09

Table 4 compares the performance of the NN model for SBP and DBP estimation in terms of ME and STDE in mmHg between the two datasets, namely VitalDB and OR_S7. We observe a difference of 2.57 mmHg for SBP and 1.23 mmHg for DBP on the STDE between the two datasets.

Table 4. Evaluation performance of SBP and DBP estimation on the VitalDB and OR_S7 dataset.

	SBP		DBP	
	ME	STDE	ME	STDE
VitalDB	-0.88	10.29	-0.60	5.76
OR_S7	1.13	12.86	-0.44	6.99

4. Discussion

To the best of our knowledge, only a few studies have evaluated the generalization capability of BP estimation models based on PPG signals and ML. In this paper, we propose to verify how such a model would perform on unseen data recorded with different PPG sensors and on different target populations.

The model implemented depends on an EA PPG pulse

and its second derivative. PPG waveforms are affected by many factors. Some individual-specific characteristics or external factors affect its morphology, such as age [13], skin pigmentation, tissue composition or measurement site [14], which makes generalization a crucial aspect of developing PPG-based BP estimation model. Adding a one-time calibration of the system, such as here with an initial PPG-BP spot measurement, helps to cope with the inter-subject variability of the PPG waveform and to improve generalization capability to some extent.

The model's ability to capture relevant and generalizable information depends on the quality of the recorded PPG signals. In this study, the PPG sensors in the two different datasets are both localized at the fingertip. However, they are of different configurations: one in transmission while the other in reflectance. Signals recorded in reflectance are known to be of lower quality than in transmission. High quality PPG signals might be required to properly extract morphological features from the PPG pulse. Lower quality PPG signals recorded in reflectance might affect the performance of the model.

Some aspects related to the NN architecture have already been considered in this study to improve generalization capability. However, a promising solution that could be investigated is domain adaptation. This is a transfer learning scenario in which the target probability distribution differs from the source setting but remains related, such as PPG signals from various sensors or target populations. The model might generalize better, as it learns to recognize representations applicable to different situations. Such an approach has the potential to help account for the PPG waveform variability. It would be interesting to deepen research in this direction.

5. Conclusion

The present study investigates the challenge of building a BP estimation model based on PPG signals and feature learning that generalizes well to unseen data. This task is particularly difficult due to the heterogeneity of PPG waveforms. The results highlight the ability of such a model to adapt well to unseen data from different PPG sensors and target populations.

References

[1] K. T. Mills, A. Stefanescu, and J. He, "The Global Epidemiology of Hypertension," *Nat Rev Nephrol*, vol. 16, no. 4, pp. 223–237, Apr. 2020.

[2] Y. Kurylyak, F. Lamonaca, and D. Grimaldi, "A Neural Network-Based Method for Continuous Blood Pressure Estimation from a PPG Signal," in *2013 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, Minneapolis, MN, USA, May 2013, pp. 280–283.

[3] M. Liu, L.-M. Po, and H. Fu, "Cuffless Blood Pressure Estimation Based on Photoplethysmography Signal and Its Second Derivative," *IJCTE*, vol. 9, no. 3, pp. 202–206, 2017, doi: 10.7763/IJCTE.2017.V9.1138.

[4] S. G. Khalid, J. Zhang, F. Chen, and D. Zheng, "Blood Pressure Estimation Using Photoplethysmography Only: Comparison between Different Machine Learning Approaches," *Journal of Healthcare Engineering*, vol. 2018, pp. 1–13, Oct. 2018, doi: 10.1155/2018/1548647.

[5] Y. Zhang and Z. Feng, "A SVM Method for Continuous Blood Pressure Estimation from a PPG Signal," in *Proceedings of the 9th International Conference on Machine Learning and Computing - ICMLC 2017*, Singapore, Singapore, 2017, pp. 128–132. doi: 10.1145/3055635.3056634.

[6] J. Jorge *et al.*, "Machine Learning Approaches for Improved Continuous, Non-Occlusive Arterial Pressure Monitoring Using Photoplethysmography," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Montreal, QC, Canada, Jul. 2020, pp. 910–913. doi: 10.1109/EMBC44109.2020.9176512.

[7] C. Aguet *et al.*, "Feature Learning for Blood Pressure Estimation from Photoplethysmography," in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, Nov. 2021, pp. 463–466. doi: 10.1109/EMBC46164.2021.9630665.

[8] H.-C. Lee and C.-W. Jung, "Vital Recorder—a Free Research Tool for Automatic Recording of High-Resolution Time-Synchronised Physiological Data from Multiple Anaesthesia Devices," *Sci Rep*, vol. 8, no. 1, p. 1527, Dec. 2018, doi: 10.1038/s41598-018-20062-4.

[9] G. Hofmann *et al.*, "A Novel Smartphone App for Blood Pressure Measurement: A Proof-of-Concept Study Against an Arterial Catheter," *J Clin Monit Comput*, Jun. 2022, doi: 10.1007/s10877-022-00886-2.

[10] Y. Ghamri *et al.*, "Automated Pulse Oximeter Waveform Analysis to Track Changes in Blood Pressure During Anesthesia Induction: A Proof-of-Concept Study," *Anesthesia & Analgesia*, vol. 130, no. 5, pp. 1222–1233, May 2020, doi: 10.1213/ANE.0000000000004678.

[11] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *Journal of Machine Learning Research*, vol. 15, no. 56, pp. 1929–1958, 2014.

[12] P. J. Huber, "Robust Estimation of a Location Parameter," *The Annals of Mathematical Statistics*, vol. 35, no. 1, pp. 73–101, Mar. 1964, doi: 10.1214/aoms/1177703732.

[13] J. Allen and A. Murray, "Age-Related Changes in the Characteristics of the Photoplethysmographic Pulse Shape at Various Body Sites," *Physiol. Meas.*, vol. 24, no. 2, pp. 297–307, May 2003, doi: 10.1088/0967-3334/24/2/306.

[14] V. Hartmann, H. Liu, F. Chen, Q. Qiu, S. Hughes, and D. Zheng, "Quantitative Comparison of Photoplethysmographic Waveform Characteristics: Effect of Measurement Site," *Front. Physiol.*, vol. 10, p. 198, Mar. 2019, doi: 10.3389/fphys.2019.00198.

Address for correspondence:

Clémentine Aguet.
Rue Jaquet-Droz 1, 2000 Neuchâtel, Switzerland.
clementine.aguet@csem.ch