2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) | 979-8-3503-2447-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/EMBC40787.2023.10340929

Generalization Error of a Regression Model for Non-Invasive Blood Pressure Monitoring using a Single Photoplethysmography (PPG) Signal

Marek Żyliński, Edoardo Occhipinti, and Danilo Mandic

***Abstract*—Photoplethysmography (PPG) sensors integrated in wearable devices offer the potential to monitor arterial blood pressure (ABP) in patients. Such cuffless, non-invasive, and continuous solution is suitable for remote and ambulatory monitoring. A machine learning model based on PPG signal can be used to detect hypertension, estimate beat-by-beat ABP values, and even reconstruct the shape of the ABP. Overall, models presented in literature have shown good performance, but there is a gap between research and potential real-world use cases. Usually, models are trained and tested on data from the same dataset and same subjects, which may lead to overestimating their accuracy. In this paper: we compare cross-validation, where the test data are from the same dataset as training data, and external validation, where the model is tested on samples from a new dataset, on a regression model which predicts diastolic blood pressure from PPG features. The results show that, in the cross-validation, the predicted and the real values are linearly dependent, while in the external validation, the predicted values are not related to the real ones, but probably just through an average value.**

I. Introduction

Cuffless estimation of arterial blood pressure (ABP) with photoplethysmography (PPG) is an ongoing topic of research. PPG sensors integrated in wearable devices offer the potential to monitor ABP in patients, with a cuffless, non-invasive, and continuous solution for remote and ambulatory monitoring. An end-to-end solution using a single-location PPG sensor can be used to detect hypertension, estimate ABP beat-by-beat and even reconstruct the shape of the ABP signal. Cano *et al.* [1] showed that a simple k-nearest neighbor (k-NN) classifier, which uses features extracted from ECG and PPG signals, can be used for binary classification of patients as normotensive or hypertensive group, with 88.7% accuracy, and increasing to 97% accuracy after calibration [2],[3].

Previous works [4,5] have attempted to estimate blood pressure values (systolic - SBP, diastolic - DBP and mean blood pressure - MAP) using regression models built on beat-by-beat PPG features. In [3] DBP, MAP, and SBP were estimated with the mean absolute error (MAE) of 5.6±5.1 mmHg, 5.2±5.2 mmHg, and 7.4±7.3 mmHg, respectively. Dias *et al.* [5] employed the *Category Boosting* algorithm and obtained even better model prediction scores with a mean error prediction of 0.02±3.77 mmHg and the Pearson's Correlation Coefficient (R-value) of 0.93 for DBP, mean error of 0.05±7.84 mmHg and the R-value of 0.93 for SBP. Wang *et al.* [5] employed a neural network model and achieved a mean absolute error of 4.02 ± 2.79 mmHg for SBP and 2.27 ± 1.82 mmHg for DBP.

In [7], a convolutional neural network with calibration predicting ABP exhibited MAE of 3.20 mmHg and the root-mean-square error (RMSE) of 4.38 mmHg. Overall, models presented in the literature have shown good performance, but also that there is a gap between research and potential real-world use cases. Usually, these models are trained and tested on data from the same dataset and subjects, which may lead to overestimating their accuracy.

Padovano *et al.* [8] pointed out that cross-validation can overestimate performance of a model. They have compared the accuracy of different models for detection of obstructive sleep apnea in two scenarios: 10-fold cross-validation and external validation, where the model was tested on a different database from the one used as the training set. The authors found that, regardless of the employed classifier, cross-validation presents more optimistic results compared to those obtained under the external validation approach.

Indeed, the generalization capability of a model to adapt to new, unseen data, is a prerequisite to employing this model in real-life scenarios [8]. To this end, in this paper, we compare cross-validation and external validation of the regression model for the prediction of DBP from PPG features. We show that in the cross-validation, the predicted and the real values are linearly dependent, while in the external validation, the predicted values are not related to the real ones, but probably just through an average value.

II. Methods

A. Data preparation and feature extraction

The dataset used was the "Continuous Cuffless Monitoring of Arterial Blood Pressure via Graphene Bioimpedance Tattoos" [9], available on PhysioNet. It contains recordings

This work is supported by USSOCOM grant (EESB P85655). E.O. is supported by the UKRI Centre for Doctoral Training in AI for Healthcare grant number EP/S023283/1.

M. Z. and D.M are with the Electrical and Electronic Engineering Department, Imperial College London, United Kingdom. (email: m.zylinski@imperial.ac.uk; d.mandic@imperial.ac.uk).

E.O. is with the UKRI Centre for Doctoral Training in AI for Healthcare, Imperial College London, United Kingdom (email: edoardo.occhipinti16@imperial.ac.uk).

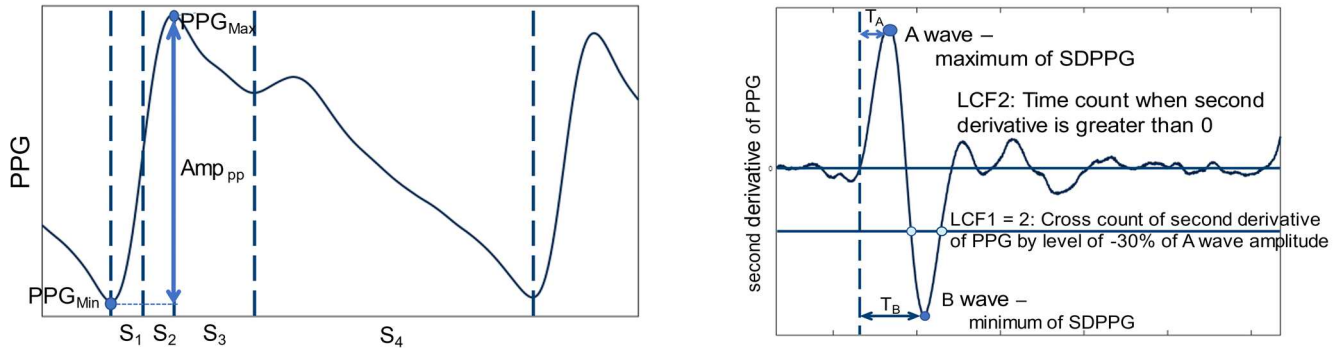

Figure 1. Visualization of features extracted from Left) PPG signal; Right) Second derivative of the PPG signal.

from 7 subjects who performed several experimental maneuvers to alter ABP, such as handgrip, cold pressor test, Valsalva maneuver, and cycling exercises. A recording in resting conditions was also taken before starting the experimental protocol. Each recording contains a PPG signal measurement from the fingertip and a continuous blood pressure signal measured from the finger cuff with the volume-clamp method (Finapres NOVA device).

Before building the regression model we extracted the relevant features from the raw PPG signal. According to [11], we extracted 16 different features:

- Minimum, maximum, peak-to-peak amplitude of PPG.

- PPG intensity ratio (PIR) - the ratio of maximum and minimum value of the PPG signal in each heartbeat, which corresponds to the changes in artery volume [12].

- Areas under the PPG [3]: S1: from onset to maximum slope during systole, S2: from maximum slope to PPG maximum, S3: from peak to inflection point. S4: area during diastole from inflection point to the end of the beat. We also calculated the ratio K of the systolic (S1 + S2 + S3) and diastolic (S4) area under the curve, which can be used to estimate artery compliance [13],[14].

- Amplitude and time of A and B wave in the second derivative of the PPG, LCF1 - count of signal crossing level equals to 30% of A peak amplitude; LCF2 – time duration of a positive derivative [15]. These parameters give information about cardiac contraction and pressure wave propagation.

Fig. 1 visualizes these features in a PPG cycle and its second derivative.

The algorithm described in [16] was used to separate the recordings into individual heartbeats, and the signal quality index (SQI) for each heartbeat was estimated using built-in function from the PhysioNet Cardiovascular Signal Toolbox [17]. The beats which were identified as *poor* using built-in toolbox criteria were excluded from the analysis. The analysis was performed in MATLAB R2022a.

B. Regression model

The dataset was split into 4 parts of equal size (I-IV), and 4 different models were trained for each part. The split was performed in two ways: *random* split and *sequential* split. The random split was equivalent to a 4-fold cross-validation scenario, where data from different subjects are shuffled so that training and testing data contain examples from each subject. Since data from the subjects are ordered in the sequential split, the model was tested on a mix of the seen and unseen subjects.

Each data part I-IV was then randomly split into a training (70%) and validation dataset (30%). Next, we generated multiple ensemble tree models which were trained on the training data from each part, validated on the validation dataset, and tested on the validation dataset of other parts. The performance of each model was quantified with the Mean Absolute Error (MAE).

III. RESULTS

Fig. 2 shows the predicted DBP against the real one measured from the finger cuff with the *random* and *sequential* split. Observe that in the random split, the predicted values and the real ones are linearly correlated while in the sequential one, the predicted values are not related to the real ones, but probably just an average value which minimizes the MAE.

Table 1 compares the performance of the different generated models. Even when looking at the MAE scores, we can observe that the in random split, the MAEs, are in the range of 7.4-7.7 mmHg, while in the case of sequential split, the generalization ability of the model was poor, when testing on a different part of the dataset, with MAEs in the range of 4.3-35.2 mmHg.

Table 1. Mean absolute error (MAE) for the ensemble tree regression model for the random and sequential split method.

	Random split					*Sequential split*			
MAE	**I**	**II**	**III**	**IV**	**MAE**	**I**	**II**	**III**	**IV**
I	7.7	7.4	7.2	7.6	**I**	4.9	15.8	9.2	17.1
II	7.5	7.4	7.5	7.6	**II**	25.8	6.8	15.0	23.3
III	7.7	7.4	7.5	7.6	**III**	12.1	13.0	4.3	17.5
IV	7.7	7.5	7.5	7.6	**IV**	35.2	17.3	10.9	6.2

IV. DISCUSSION

The generalization error of a machine learning model represents the difference between the empirical loss for the training set and the expected loss of a test set [18]. Zhang *et al.* [19] demonstrated that complex deep neural networks easily fit random labelled datasets, but when evaluated on

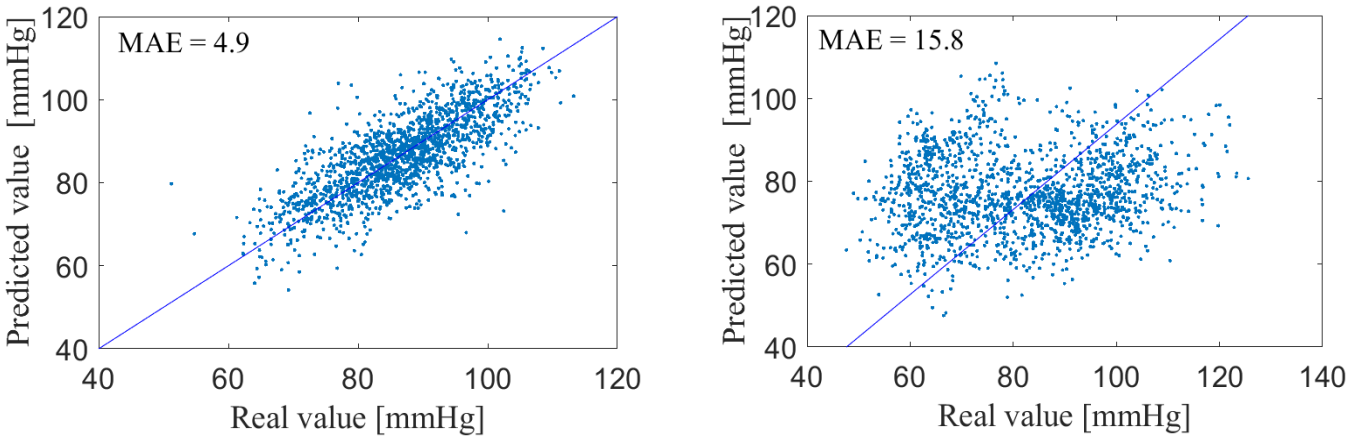

Figure 2. Prediction performance of the model. Left) the predicted result for the model trained on the first part and tested on the same part of the dataset (I-I). Right) plot for the model trained for the first part of the dataset and tested on the second part of the dataset (I-II).

new observations, they produce scores which are no better than random chance. From the results observed in Fig. 3 and Table 1, we can conclude that our proposed regression model does not generalize to unseen subjects. We hypothesize that poor performance of the model may be caused partially by the small size of the dataset used, but mostly by the lack of necessary information in PPG signal to estimate blood pressure.

It is well known that the light intensity measured by PPG is inversely correlated to the blood volume in the tissue [20], and often it is assumed that plethysmography measures the sum of volume changes in all blood vessels (large and small arteries, arterioles, venules, and veins), with the arterial pulsations dominating [21]. However, the PPG signal might not be accurate for the measurement of the arterial volume [22]. Wang and Zheng [22] compared reflection-mode PPG with A-mode ultrasound on the wrist artery and found that there were obvious differences between the shapes of these two signals. There are multiple factors that influence the PPG signal, such as the change in arterial blood volume, the mechanical properties of capillaries, and the erythrocyte movement [23]. The significance of these factors may vary depending on the sensor configuration and wavelength used, e.g. the green light in PPG, may not reach deep arteries hence the reflected signal may be generated by the dermis deformation due to the changing blood pressure [24].

Mounting of the sensor can also alter the PPG measurement, as the force applied to the PPG sensor may change the correlation between the PPG and blood pressure waveforms [25], thus altering the PPG waveform amplitude [26]. The blood pressure waveform may be assumed as a sum of several components: volume-related reservoir pressure, proportional to changes in arterial volume; forward propagation pressure wave, proportional to blood flow in arteries [27]; and reflected backward pressure wave [28]. Therefore, the information about changes in the artery volume may be not sufficient for estimating DBP. The impact of these components in blood pressure may vary between subjects, resulting on different morphologies of blood pressure waveforms [29]. Moreover, the amplitude of the pressure pulse is amplified towards peripheral arteries, and also pulse pressure is substantially higher in peripheral vessels, such as the brachial and radial arteries, compared to the central aorta [30]. In addition, peripheral pulse pressure augmentation is dependent on an individual cardiovascular characteristic, so a generalized function describing the relationship between the pressure waveform at different arterial sites may be inaccurate [31]. Arteries compliance may vary substantially in population [32], as compliance increases with age [33]. Inter arterial pressure may also cause changes in arteries compliance through activation of in arteries wall smooth muscle [34].

In this work, the features extracted from PPG were selected to address different cardiovascular phenomena: artery volume changes, pressure wave propagation, and artery compliance. During training, the model had to learn the relationship between the features and DBP, however, this relationship may be different for every subject. The training dataset must be sufficiently large and well spread over sample space to achieve a good generalization of a model [18]. In our case, especially when sequential split was performed, the trained model was unable to learn the individual subject relationship between DBP and PPG. On the other hand, models trained on a smaller subset of the dataset may show poor generalization ability [35].

We also desire that our model works in different test conditions: during rest, exercises, and autonomic nervous system test (handgrip and Valsalva maneuver). Such changes of the condition may not only change DBP but also arterial compliance and compromise model accuracy. Also, the employed dataset may be insufficient for proper training of a regression model. The proposed model therefore needs certain improvements: it should be trained on a larger dataset, individually calibrated, together with better pre-processing in order to extract only artery-related information. For this, a more sophisticated algorithm may be required.

V. Conclusion

We have investigated the use of single PPG signal with a regression type machine learning model, in order to monitor ABP. This may provide a cuffless, non-invasive, and continuous solution for remote and ambulatory monitoring of ABP and leverage usage of wearables devices but first poor

generalization and decrease of model performance on new subjects must be addressed.

Overall, although our regression model has exhibited a low MAE in cross-validation, it has shown poor generalization ability in an external-validation scenario. This problem must be fully addressed before the model can be implemented in clinical applications. Further works may include enhanced signal preprocessing and feature extraction, usage of more complex model architecture, and individualized calibration. Furthermore, a combination of features extracted from PPG and ECG may give the model more information toward a more reliable blood pressure estimate.

REFERENCES

[1] J. Cano, V. Bertomeu-González, L. Fácia, J. Moreno-Arribas, R. Alcaraz, and J. J. Rieta, "ECG and PPG-Based Hypertension Screening Under Non-Hypertensive Blood Pressure Recordings." *Computing in Cardiology (CinC)*, vol. 49, IEEE, 2022

[2] J. Cano, L. Fácila, J. M. Gracia-Baena, R. Zangróniz, R. Alcaraz, and J. J. Rieta, "The Relevance of Calibration in Machine Learning-Based Hypertension Risk Assessment Combining Photoplethysmography and Electrocardiography," *Biosensors*, vol. 12, no. 5, p. 289, 2022.

[3] J. Cano, L. Fácia, F. Hornero, P. Langley, R. Alcaraz, and J. J. Rieta, "Cuffless Hypertension Risk Assessment and the Significance of Calibration." *Computing in Cardiology (CinC)*, vol. 49, IEEE, 2022

[4] P. Li and T.-M. Laleg-Kirati, "Central blood pressure estimation from distal PPG measurement using semiclassical signal analysis features," *IEEE Access*, vol. 9, pp. 44963-44973, 2021.

[5] F. M. Dias, T. B. Costa, D. A. Cardenas, M. A. Toledo, J. E. Krieger, and M. A. Gutierrez, "A Machine Learning Approach to Predict Arterial Blood Pressure from Photoplethysmography Signal." *Computing in Cardiology (CinC)*, vol. 49, IEEE, 2022

[6] L. Wang, W. Zhou, Y. Xing, and X. Zhou, "A novel neural network model for blood pressure estimation using photoplethesmography without electrocardiogram," *Journal of Healthcare Engineering*, vol. 2018, 2018.

[7] J. Cheng, Y. Xu, R. Song, Y. Liu, C. Li, and X. Chen, "Prediction of arterial blood pressure waveforms from photoplethysmogram signals via fully convolutional neural networks," *Computers in Biology and Medicine*, vol. 138, p. 104877, 2021.

[8] D. Padovano, A. Martinez-Rodrigo, J. M. Pastor, J. J. Rieta, and R. Alcaraz, "Hidden Hazards Beneath Cross-Validation Methods in Machine Learning-Based Sleep Apnea Detection." *Computing in Cardiology (CinC)*, vol. 49, IEEE, 2022

[9] C. Aguet et al. "Generalization Capability of a Neural Network for Blood Pressure Estimation from Photoplethysmography." *Computing in Cardiology (CinC)*, vol. 49, IEEE, 2022

[10] B. Ibrahim, "Continuous Cuffless Monitoring of Arterial Blood Pressure via Graphene Bioimpedance Tattoos." *PhysioNet*, 2022.

[11] S. Haddad, A. Boukhayma, and A. Caizzone, "Continuous ppg-based blood pressure monitoring using multi-linear regression," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 5, pp. 2096-2105, 2021.

[12] 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*, vol. 7, no. 1, pp. 1-11, 2017.

[13] Z. Liu, K. P. Brin, and F. Yin, "Estimation of total arterial compliance: an improved method and evaluation of current methods," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 251, no. 3, pp. H588-H600, 1986.

[14] M. Żyliński, W. Niewiadomski, M. Sadowiec, and G. Cybulski, "Different methods of arterial compliance estimation tested with reservoir-wave model of arterial system," in *World Congress on Medical Physics and Biomedical Engineering 2018*, 2019: Springer, pp. 707-710.

[15] D. Fujita, A. Suzuki, and K. Ryu, "PPG-based systolic blood pressure estimation method using PLS and level-crossing feature," *Applied Sciences*, vol. 9, no. 2, p. 304, 2019.

[16] W. Zong, T. Heldt, G. Moody, and R. Mark, "An open-source algorithm to detect onset of arterial blood pressure pulses," in Computers in Cardiology, 2003, IEEE, pp. 259-262.

[17] A. N. Vest et al., "An open source benchmarked toolbox for cardiovascular waveform and interval analysis," *Physiological Measurement*, vol. 39, no. 10, p. 105004, 2018.

[18] D. Jakubovitz, R. Giryes, and M. R. Rodrigues, "Generalization error in deep learning," in *Compressed Sensing and its Applications:* Springer, 2019, pp. 153-193.

[19] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals, "Understanding deep learning (still) requires rethinking generalization," *Communications of the ACM*, vol. 64, no. 3, pp. 107-115, 2021.

[20] A. A. Kamshilin and N. B. Margaryants, "Origin of photoplethysmographic waveform at green light," *Physics Procedia*, vol. 86, pp. 72-80, 2017.

[21] A. Reisner, P. A. Shaltis, D. McCombie, H. H. Asada, D. S. Warner, and M. A. Warner, "Utility of the photoplethysmogram in circulatory monitoring," *The Journal of the American Society of Anesthesiologists*, vol. 108, no. 5, pp. 950-958, 2008.

[22] C.-Z. Wang and Y.-P. Zheng, "Comparison between reflection-mode photoplethysmography and arterial diameter change detected by ultrasound at the region of radial artery," *Blood Pressure Monitoring*, vol. 15, no. 4, pp. 213-219, 2010.

[23] S. Chatterjee, K. Budidha, and P. A. Kyriacou, "Investigating the origin of photoplethysmography using a multiwavelength Monte Carlo model," Physiological Measurement, vol. 41, no. 8, p. 084001, 2020.

[24] I. S. Sidorov, R. V. Romashko, V. T. Koval, R. Giniatullin, and A. A. Kamshilin, "Origin of infrared light modulation in reflectance-mode photoplethysmography," *PLoS* One, vol. 11, no. 10, p. e0165413, 2016.

[25] H. Hsiu, C.-L. Hsu, and T. Wu, "Effects of different contacting pressure on the transfer function between finger photoplethysmographic and radial blood pressure waveforms," Proceedings of the Institution of Mechanical Engineers, Part H: *Journal of Engineering in Medicine*, vol. 225, no. 6, pp. 575-583, 2011.

[26] A. Chandrasekhar, M. Yavarimanesh, K. Natarajan, J.-O. Hahn, and R. Mukkamala, "PPG sensor contact pressure should be taken into account for cuff-less blood pressure measurement," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 11, pp. 3134-3140, 2020.

[27] J. V. Tyberg, J. C. Bouwmeester, K. H. Parker, N. G. Shrive, and J.-J. Wang, "The case for the reservoir-wave approach," *International Journal of Cardiology*, vol. 172, no. 2, pp. 299-306, 2014.

[28] N. Westerhof, P. Sipkema, G. V. D. Bos, and G. Elzinga, "Forward and backward waves in the arterial system," *Cardiovascular Research*, vol. 6, no. 6, pp. 648-656, 1972.

[29] J. P. Murgo, N. Westerhof, J. P. Giolma, and S. A. Altobelli, "Aortic input impedance in normal man: relationship to pressure wave forms," *Circulation*, vol. 62, no. 1, pp. 105-116, 1980.

[30] P. Segers, D. Mahieu, J. Kips, and L. M. Van Bortel, "The use of a generalized transfer function: different processing, different results!," *Journal of Hypertension*, vol. 25, no. 9, pp. 1783-1787, 2007.

[31] S. A. Hope, I. T. Meredith, D. Tay, and J. D. Cameron, "'Generalizability'of a radial-aortic transfer function for the derivation of central aortic waveform parameters," *Journal of Hypertension*, vol. 25, no. 9, pp. 1812-1820, 2007.

[32] D. Chemla et al., "Total arterial compliance estimated by stroke volume-to-aortic pulse pressure ratio in humans," *American journal of physiology-Heart and Circulatory Physiology*, vol. 274, no. 2, pp. H500-H505, 1998.

[33] L. Van Bortel and J. Spek, "Influence of aging on arterial compliance," *Journal of Human Hypertension*, vol. 12, no. 9, pp. 583-586, 1998.

[34] W. M. Bayliss, "On the local reactions of the arterial wall to changes of internal pressure," *The Journal of Physiology*, vol. 28, no. 3, p. 220, 1902.

[35] S. Durga, R. Iyer, G. Ramakrishnan, and A. De, "Training Data Subset Selection for Regression with Controlled Generalization Error," in Proceedings of the *International Conference on Machine Learning*, 2021: PMLR, pp. 9202-9212.