Machine learning techniques applied to in-silico pulse wave velocity estimation based on photoplethysmographic signals

Vivien Debuchy, Mohamed Khalifa, Philippe Tresson, Nadège Thirion-Moreau, Eric Moreau

Machine learning techniques applied to in-silico pulse wave velocity estimation based on photoplethysmographic signals

Vivien Debuchy
Université de Toulon, Aix Marseille Univ
UMR CNRS 7020, LIS
Tecmoled
Toulon, France
vivien.debuchy@tecmoled.com

Mohamed Khalifa
CEO
Tecmoled
Toulon, France
mohamed.khalifa@tecmoled.com

Philippe Tresson
Vascular Surgery Department
Marseille Public University Hospital
Marseille, France
philippe.tresson@ap-hm.fr

Nadège Thirion-Moreau
Université de Toulon, Aix Marseille Univ
UMR CNRS 7020, LIS
Toulon, France
thirion@univ-tln.fr

Eric Moreau
Université de Toulon, Aix Marseille Univ
UMR CNRS 7020, LIS
Toulon, France
moreau@univ-tln.fr

***Abstract*—Carotid femoral pulse wave velocity (cf-PWV) is the gold standard measurement of arterial stiffness that has been recognized as an effective biomarker of cardiovascular disease (CVD) risk. Although reliable and accurate, the reference method for measuring cf-PWV is time-consuming and requires the intervention of a qualified practitioner. Photoplethysmograpy (PPG) is a non-invasive cost effective technology that contains a multitude of information about the cardiovascular system. This paper aims to explore the potential of estimating cf-PWV through PPG pulse wave analysis for large-scale CVD risk screening. This includes a comparative analysis involving two machine learning models and various sensor positions. A set of features based on fiducial points extracted from in-silico PPG signals and their derivatives is used as an input for XGBoost and Support Vector Regression (SVR) models. These models are trained on simulated sensor positions, evaluated across different noise levels, and demonstrate comparable or superior performance compared to previous studies. The proposed method is deployable on a low-power embedded processor. Signals from the superficial temporal artery position exhibit the best performance, achieving an R^2 of 1.00 and a root-mean-square error (RMSE) of 0.13. The PPG signal combined with the proposed method shows promising potential for cf-PWV estimation particularly when using superficial temporal artery signals. These results motivate an in-vivo validation of the suggested method.**

***Index Terms*—Photoplethysmography, Machine-Learning, Vascular Age, Pulse wave velocity**

I. Introduction

Cardiovascular disease (CVD) is one of the leading causes of death worldwide. Many of the factors leading to this disease are preventable [1]. The implementation of an economical and widely applicable early detection system holds the potential to greatly improve risk stratification. Several methods have been developed to detect people at risk of CVD. One approach is to establish multivariable risk algorithms like The Framingham Heart Study [2] or QRISK [3] to estimate the probability of CVD events. These approaches, although effective, often require in-depth examinations such as blood tests. Another approach is to measure identified biomarkers that are related to vascular age to determine the risk of CVD.

Arterial stiffness tends to increase with age. While arterial stiffness is primarily attributed to the natural aging process, it can be influenced by various accelerating factors. Many studies have established that an increase in arterial stiffness is associated with an increase in the probability of CVD events [4] [5]. Consequently, the assessment of aortic arterial stiffness using the gold standard method of carotid femoral pulse wave velocity (cf-PWV) emerged as a significant indicator for predicting CVD occurrences [6]. However, measuring cf-PWV requires a qualified practitioner typically employing applanation tonometry or doppler flowmetry techniques, making it difficult to apply on a large scale.

Photoplethysmography (PPG) is an affordable and widely used optical method for detecting blood volume changes in arteries. PPG signal contains valuable information about the cardiovascular system and is a promising technology for the prevention of CVD. PPG signals can be acquired using two types of approaches: reflection or transmission. Regarded as the reference method for pulse oximeters, the transmission technique relies on the absorption of light by the tissues. It is used for sensors that are placed on the finger or ear. The reflection method captures light reflected back to the sensor, as both the light source and receiver are situated on the same surface. This method is much more versatile because the sensor can be placed on various body parts such as the wrist,

This work was supported by a CIFRE grant (number 2022/0097).

finger, forehead, or temple.

Several attempts have been made to estimate arterial stiffness through cf-PWV and PPG signals, utilizing either in-vivo or in-silico data. For this specific task, no public in-vivo dataset is currently available. Consequently, initial testing and comparative analysis of methods can be performed using in-silico data. In [7] an in-vivo study was conducted using temporal features on wrist PPG and its first and second derivatives with additional demographic features on 310 subjects. Good performances have been obtained with an XGBoost regressor. Another in-vivo study was carried out in [8] on 90 subjects using features on finger PPG and its first and second derivatives and subject's height with multiple linear regression. The impact of adding parameters such as age and blood pressure was also assessed. Acceptable performance has been obtained with an improvement by considering the addition of further parameters. PPG spectrograms were used in [9] with an in-vivo dataset. Different sets of features were obtained using semi-classical signal analysis, Law's mask, and statistical calculations. Several positions and models were compared showing that spectogram derived features are suitable for estimating cf-PWV. In this study we have investigated the feasibility of using a set of features extracted from the PPG signal and its first and second derivatives to reliably estimate cf-PWV on multiple sensing position. The main goal is to develop a robust, reproducible and less complex method that can be easily implemented in an embedded processor.

II. Material and method

A. Material

Since there is no in-vivo publicly available dataset, an in-silico public dataset was used in this study [10]. This dataset contains PPG signals that were obtained at different location by 1D simulation of wave propagation in an arterial network. Several subject profiles were simulated by varying the properties of the cardiovascular system. The dataset includes a total number of 4374 virtual healthy subjects divided in six ages groups from 25 years to 75 years old with a 10-year increment.

B. Method

1) Signal Enhancement: In-silico signals are noise-free. To enhance the realism of the simulation, white noise is intentionally incorporated into the signal with a variable signal-to-noise ratio (SNR). Four noise levels were considered: Noise-free, SNR=65 dB, SNR=45 dB and SNR=30 dB. Then a 4th order Butterworth lowpass filter is applied with a cutoff frequency set at 7Hz to smooth the signal from high-frequency noise, keeping only the useful frequency range between about 1Hz and 3Hz.

2) Feature extraction: A comprehensive set of temporal features is derived from the PPG signal along with its first and second derivatives. These features based on fiducial points were widely employed in previous researches, notably for blood pressure estimation [11]. However, unlike other studies such as [12], several features were removed from this specific set. Specifically, only three fiducial points (a_2, b_2 and c_2) were used on the second derivative as opposed to the usual six fiducial points. Table I provides a representation of the features extracted from the PPG signal, its first derivative, and second derivative. Additionally, Figure 1 displays a graphical representation of these main features.

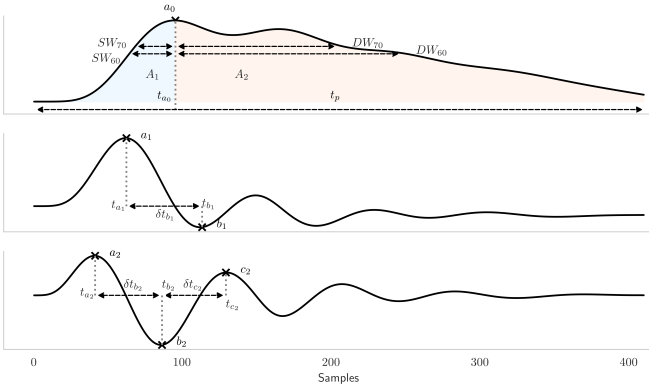


Fig. 1. Pulse signal features of the PPG signal P and its first and second derivatives (respectively P' and P'')

TABLE I
FEATURES DESCRIPTION

Variables	Description
Pulse features	
$(a_0,\, t_{a_0})$	Systolic peak
t_p	Length of the pulse
SW_x and DW_x	Systolic and diastolic pulse width at various heights (60%-70%-80%-90%)
A_1/A_2	Ratio of systolic pulse area to diastolic pulse area
First Derivative Features	
$(a_1,\, t_{a_1})$	First peak
$(b_1,\, t_{b_1})$	First onset
$\delta t_{b_1} = t_{b_1} - t_{a_1}$	Distance between first peak and first onset
Second Derivative Features	
$(a_2,\, t_{a_2})$	First peak
$(b_2,\, t_{b_2})$	First onset
$(c_2,\, t_{c_2})$	First peak after t_{b_1}
$\delta t_{b_2} = t_{b_2} - t_{a_2}$	Distance between first peak and first onset
$\delta t_{c_2} = t_{c_2} - t_{b_2}$	Distance between first onset and second onset
Feature combination	
$t_p - t_{a_0}$	Diastolic width
t_{a_0}/t_p	Ratio of rising time to pulse length
$(t_p - t_{a_0})/t_p$	Ratio of Diastolic width to pulse length
t_{a_1}/t_p	Ratio of P' first peak time to pulse length
$\delta t_{b_1}/t_p$	Ratio of P' first onset time by pulse length
b_2/a_2	Ratio of P'' first peak amplitude to P'' first onset amplitude
$P(t_{b_2})/P(t_{a_2})$	Ratio of P at t_{b_2} to P at t_{a_2}
c_2/a_2	Ratio of P'' first peak amplitude to second onset
$P(t_{c_2})/P(t_{a_2})$	Ratio of P at t_{c_2} to P at t_{a_2}
t_{a_2}/t_p	Ratio of P'' first peak time to pulse length
$\delta t_{b_2}/t_p$	Ratio of P'' first onset time to pulse length
$\delta t_{c_2}/t_p$	Ratio of P'' second onset time to pulse length

3) Model training: The dataset is divided into a training set and a validation set with a 80%/20% split respectively.

Training is performed using features extracted from noise-free pulses. Three sensing body positions are tested: digital, radial, and superficial temporal. Two models are trained for each body position: Xtreme Gradient Boosting (XGBoost) [13] and Support Vector Machine Regressor (SVR) [14]. Hyperparameter tuning is performed for each model using Tree-structured Parzen Estimator (TPE) approach [15]. Then, these trained models undergo testing across multiple noise levels. This approach reveals whether noise variations can have an impact on the model performance or not.

4) Model deployment: Finally, a deployment is done on a custom wearable forehead PPG sensor printed circuit board developed by Tecmoled company. Feature extraction and trained model are embedded on a low-power cortex M33 processor running on Zephyr real-time operating system. Memory footprint and execution time are monitored to benchmark performance on a real device Figure 2. The system consists of a battery and an electronic acquisition board. The processor contains two cores: the first is the network core, managing the Bluetooth Low Energy (BLE) communication stack, while the second is the application core, handling user programs. The sensor has a compact form factor with a width of 28mm, a length of 53mm, and a thickness of 6mm.

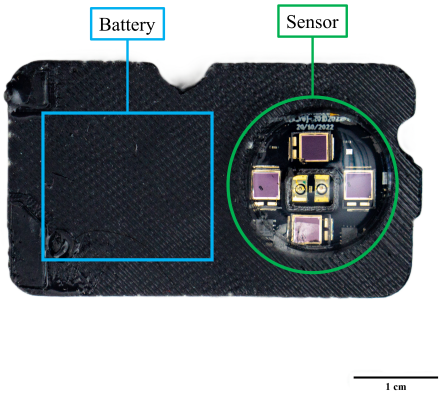


Fig. 2. Custom forehead PPG sensor used for performance benchmarking

III. Results

The deployment of the trained XGBoost model utilizing the superficial temporal position necessitates 255 kilobytes of flash memory and 16 kilobytes of RAM on an M33 cortex CPU running with Zephyr real-time operating system. The algorithm's execution time for each PPG pulse is recorded as 251 ms.

Table II presents a comprehensive overview of the performance across three examined positions (digital, radial, and superficial temporal) under varying noise levels, accompanied by model comparisons. The performance metrics used are the correlation coefficient R^2, the mean absolute error (MAE) and the root mean square error (RMSE). In noise-free conditions, optimal outcomes for the digital position are achieved using the SVR model, attaining R^2=0.99, MAE=0.14, and RMSE=0.20. The radial position exhibits superior performance with the XGBoost model with R^2=0.99, MAE=0.15, and RMSE=0.22. The superficial temporal position, likewise, has optimal results with the XGBoost model with R^2=1.00, MAE=0.08, and RMSE=0.13. Notably, the SVR model on the radial position records the least favorable performance, presenting R^2=0.97, MAE=0.26, and RMSE=0.37. Figure 4 and 5 show regression and Bland-Altman plots for the superficial temporal position utilizing the XGBoost model. With an SNR of 65dB, minimal alterations are observed in the results, R^2, MAE, and RMSE metrics remaining consistent across all models and positions.

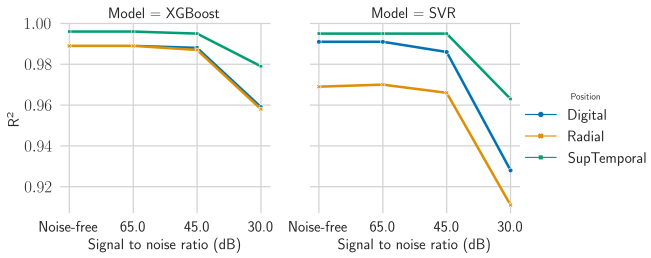
When increasing noise with an SNR of 45dB, performance remains stable with a slight increase in RMSE compared to the noise-free scenario. With a SNR of 30dB, performance begins to deteriorate significantly for the SVR model. Particularly, the most significant deterioration occurs in the radial position, where the SVR model exhibits R^2 = 0.91, MAE=0.45, and RMSE=0.63. The XGBoost model appears to sustain satisfactory performance across all positions notwithstanding a slight decline at a SNR of 30dB. The most favorable outcome is observed in the superficial temporal position with R^2=0.98, MAE=0.18, and RMSE=0.31. The evolution of RMSE and R^2 metrics over SNR is depicted in Figure 3.

TABLE II
CF-PWV ESTIMATION RESULTS FOR DIFFERENT SENSOR POSITIONS

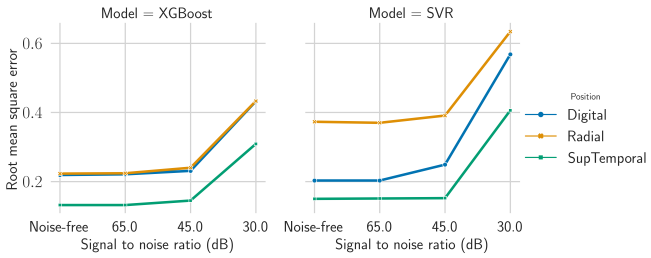
Position	Model	SNR (dB)	R2	MAE	RMSE
Digital	XGBoost	Noise-free	0.99	0.15	0.22
		65.0	0.99	0.15	0.22
		45.0	0.99	0.16	0.23
		30.0	0.96	0.29	0.43
	SVR	Noise-free	0.99	0.14	0.20
		65.0	0.99	0.14	0.20
		45.0	0.99	0.17	0.25
		30.0	0.93	0.41	0.57
Radial	XGBoost	Noise-free	0.99	0.15	0.22
		65.0	0.99	0.15	0.22
		45.0	0.99	0.16	0.24
		30.0	0.96	0.32	0.43
	SVR	Noise-free	0.97	0.26	0.37
		65.0	0.97	0.26	0.37
		45.0	0.97	0.27	0.39
		30.0	0.91	0.45	0.63
SupTemporal	XGBoost	Noise-free	1.00	0.08	0.13
		65.0	1.00	0.08	0.13
		45.0	0.99	0.09	0.14
		30.0	0.98	0.18	0.31
	SVR	Noise-free	0.99	0.10	0.15
		65.0	0.99	0.10	0.15
		45.0	0.99	0.10	0.15
		30.0	0.96	0.26	0.41

IV. Discussions

Comparing with previous work in the noise-free scenario on the same dataset [9], similar results are obtained for the digital position. For the radial positions, slightly better results are obtained with the proposed method (RMSE=0.22 vs RMSE=0.25). The best result overall is achieved by our



(a) Evolution of R^2 over SNR



(b) Evolution of RMSE over SNR

Fig. 3. Evolution of R^2 and RMSE over SNR for XGBoost and SVR model for digital, radial and superficial temporal positions

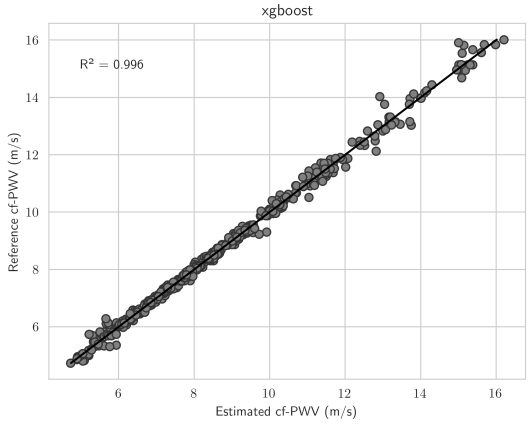


Fig. 4. XGBoost results on superficial temporal artery in the noise free case

proposed method on the superficial temporal position, outperforming the best result obtained on other positions (R^2=1.00, RMSE=0.13 vs R^2=0.99 RMSE=0.19). However, previous studies did not evaluate the performance of the superficial temporal position so no conclusion can be drawn whether this position is particularly suitable for cf-PWV estimation or if the method that we suggest performs specifically better in this configuration. Furthermore, the proposed method appears to be simpler to implement in an embedded processor with limited computing power. Compared to previous work using in-vivo datasets, the proposed method does not rely on additional patient data apart from the PPG signal, making it easier to implement.

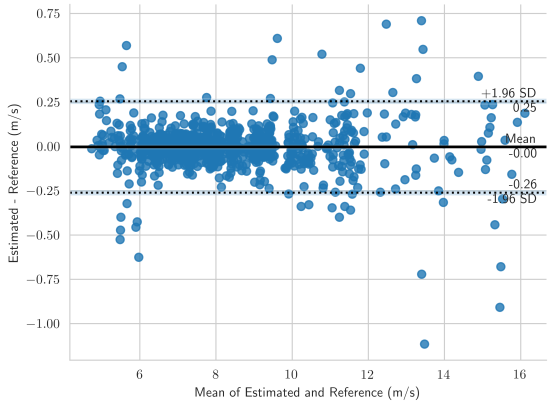


Fig. 5. Blan-Altman plot for the XGBoost model on superficial temporal artery in the noise free case

Although using fiducial points is known to be noise-sensitive because of timing shift that can be induced on temporal features [16], the combination of the chosen feature set with the XGBoost model yields relatively robust results even with an increase in SNR. Notably, the SVR model exhibits a more severe deterioration compared to XGBoost.

The best overall results were obtained on the superficial temporal artery with the XGBoost model, which to our knowledge had not been evaluated until now. The PPG signal on this body position seems to contain the best information for estimating cf-PWV. Furthermore, performance even with noise remains very good, demonstrating the robustness of the method on this body position.

The excellent results obtained in this study are encouraging for further investigation of cf-PWV estimation using the PPG signal. However, an important limitation lies in the the use of in-silico data which may differ from reality. The custom PPG sensor that was used to benchmark performance could be utilized in the future to collect real-world data.

The addition of white noise provides a more realistic simulation of the measurement noise present in a real system. However, this method does not address additional sources of noise, such as motion, which can significantly influence the shape of the PPG signal. In practical settings, integrating a pulse selection algorithm becomes essential. Considering that cf-PWV variation occurs on a timescale significantly exceeding the measurement time, it is possible to be highly selective about the pulses that are used. It may be necessary to request the subject to remain still during data collection. Sensor pressure can also play a role in the shape of PPG pulses [17]. The implementation of these considerations will be crucial for the success of the in-vivo study.

V. Conclusion

With the growing number of cardiovascular diseases in the world, large-scale prevention and screening methods are required for at-risk populations. cf-PWV is the gold standard

for measuring arterial stiffness, which is strongly associated with CVD mortality. We have investigated the possibility of using temporal features on the PPG signal and its first and second derivatives to estimate cf-PWV. The proposed method shows very good results on simulated data. The simplicity of the suggested method makes it easily implementable on embedded processors, which could make CVD prevention more accessible in clinical practice. Positioning the PPG sensor on the superficial temporal artery seems very promising for reliable estimation of cf-PWV even in the presence of noise. An evaluation using in-vivo data collected with our custom forehead PPG sensor will be necessary in the future to determine the reproducibility of the obtained results in real acquisition conditions.

ACKNOWLEDGMENT

Thanks to Ikrame Fathllah for translating the python signal processing algorithm into embedded C language. This work was partly financed by the ANRT via a CIFRE PhD contract and the Tecmoled company.

REFERENCES

[1] G. A. Roth, G. A. Mensah, C. O. Johnson, G. Addolorato, E. Ammirati, L. M. Baddour *et al.*, "Global Burden of Cardiovascular Diseases and Risk Factors, 1990–2019: Update From the GBD 2019 Study," *Journal of the American College of Cardiology*, vol. 76, no. 25, pp. 2982–3021, Dec. 2020.
[Online]. Available: https://www.sciencedirect.com/science/article/pii/S0735109720377755

[2] R. B. D'Agostino, Sr, R. S. Vasan, M. J. Pencina, P. A. Wolf, M. Cobain, J. M. Massaro *et al.*, "General cardiovascular risk profile for use in primary care: the Framingham Heart Study," *Circulation*, vol. 117, no. 6, pp. 743–753, Jan. 2008, place: United States.
[Online]. Available: http://dx.doi.org/10.1161/CIRCULATIONAHA.107.699579

[3] J. Hippisley-Cox, C. Coupland, and P. Brindle, "Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study," *BMJ*, vol. 357, p. j2099, May 2017.
[Online]. Available: http://www.bmj.com/content/357/bmj.j2099.abstract

[4] E. D. Kim, S. H. Ballew, H. Tanaka, G. Heiss, J. Coresh, and K. Matsushita, "Short-Term Prognostic Impact of Arterial Stiffness in Older Adults Without Prevalent Cardiovascular Disease," *Hypertension*, vol. 74, no. 6, pp. 1373–1382, Dec. 2019, publisher: American Heart Association.
[Online]. Available: https://doi.org/10.1161/HYPERTENSIONAHA.119.13496

[5] S. Laurent, P. Boutouyrie, R. Asmar, I. Gautier, B. Laloux, L. Guize *et al.*, "Aortic Stiffness Is an Independent Predictor of All-Cause and Cardiovascular Mortality in Hypertensive Patients," *Hypertension*, vol. 37, no. 5, pp. 1236–1241, May 2001, publisher: American Heart Association.
[Online]. Available: https://doi.org/10.1161/01.HYP.37.5.1236

[6] Q. Zhong, M.-J. Hu, Y.-J. Cui, L. Liang, M.-M. Zhou, Y.-W. Yang *et al.*, "Carotid–Femoral Pulse Wave Velocity in the Prediction of Cardiovascular Events and Mortality: An Updated Systematic Review and Meta-Analysis," *Angiology*, vol. 69, no. 7, pp. 617–629, Aug. 2018, publisher: SAGE Publications Inc.
[Online]. Available: https://doi.org/10.1177/0003319717742544

[7] Y. Li, Y. Xu, Z. Ma, Y. Ye, L. Gao, and Y. Sun, "An XGBoost-based model for assessment of aortic stiffness from wrist photoplethysmogram," *Computer Methods and Programs in Biomedicine*, vol. 226, p. 107128, Nov. 2022.
[Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169260722005090

[8] A. Gentilin, C. Tarperi, A. Cevese, A. V. Mattioli, and F. Schena, "Estimation of carotid-femoral pulse wave velocity from finger photoplethysmography signal," *Physiological Measurement*, vol. 43, no. 7, p. 075011, Jul. 2022, publisher: IOP Publishing.
[Online]. Available: https://dx.doi.org/10.1088/1361-6579/ac7a8e

[9] J. M. Vargas, M. A. Bahloul, and T.-M. Laleg-Kirati, "A learning-based image processing approach for pulse wave velocity estimation using spectrogram from peripheral pulse wave signals: An in silico study," *Frontiers in Physiology*, vol. 14, 2023.
[Online]. Available: https://www.frontiersin.org/articles/10.3389/fphys.2023.1100570

[10] P. H. Charlton, J. Mariscal Harana, S. Vennin, Y. Li, P. Chowienczyk, and J. Alastruey, "Modeling arterial pulse waves in healthy aging: a database for in silico evaluation of hemodynamics and pulse wave indexes," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 317, no. 5, pp. H1062–H1085, Nov. 2019, publisher: American Physiological Society.
[Online]. Available: https://doi.org/10.1152/ajpheart.00218.2019

[11] C. El-Hajj and P. Kyriacou, "Cuffless blood pressure estimation from PPG signals and its derivatives using deep learning models," *Biomedical Signal Processing and Control*, vol. 70, p. 102984, Sep. 2021.
[Online]. Available: https://www.sciencedirect.com/science/article/pii/S1746809421005814

[12] M. A. Bahloul, A. Chahid, and T. -M. Laleg-Kirati, "A Multilayer Perceptron-based Carotid-to-Femoral Pulse Wave Velocity Estimation using PPG Signal," in *2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI)*, Jul. 2021, pp. 1–6, journal Abbreviation: 2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI).

[13] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '16. New York, NY, USA: Association for Computing Machinery, 2016, pp. 785–794, event-place: San Francisco, California, USA.
[Online]. Available: https://doi.org/10.1145/2939672.2939785

[14] V. N. Vapnik, *The nature of statistical learning theory*. Springer-Verlag New York, Inc., 1995.

[15] J. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl, "Algorithms for Hyper-Parameter Optimization," in *Advances in Neural Information Processing Systems*, J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K. Q. Weinberger, Eds., vol. 24. Curran Associates, Inc., 2011.
[Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2011/file/86e8f7ab32cfd12577bc2619bc635690-Paper.pdf

[16] H. Liu, J. Allen, S. G. Khalid, F. Chen, and D. Zheng, "Filtering-induced time shifts in photoplethysmography pulse features measured at different body sites: the importance of filter definition and standardization," *Physiological Measurement*, vol. 42, no. 7, p. 074001, Jul. 2021, publisher: IOP Publishing.
[Online]. Available: https://dx.doi.org/10.1088/1361-6579/ac0a34

[17] C. W. Tsao, A. Lyass, M. G. Larson, D. Levy, N. M. Hamburg, J. A. Vita *et al.*, "Relation of Central Arterial Stiffness to Incident Heart Failure in the Community," *Journal of the American Heart Association*, vol. 4, no. 11, p. e002189, publisher: Wiley.
[Online]. Available: https://doi.org/10.1161/JAHA.115.002189