2022 44th Annual International Conference of
the IEEE Engineering in Medicine & Biology Society (EMBC)
Scottish Event Campus, Glasgow, UK, July 11-15, 2022

2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) | 978-1-7281-2782-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/EMBC48229.2022.9871020

Using Bayesian Optimization and Wavelet Decomposition in GPU for Arterial Blood Pressure Estimation*

José A. González-Nóvoa, *Graduate Student Member, IEEE,* Laura Busto, Pablo Santana, José Fariña, *Member, IEEE*, Juan J. Rodríguez-Andina, *Senior Member, IEEE*, Pablo Juan-Salvadores, Víctor Jiménez, Andrés Íñiguez, and César Veiga.

***Abstract*— Continuous monitoring of arterial blood pressure (ABP) of patients in hospital is currently carried out in an invasive way, which could represent a risk for them. In this paper, a noninvasive methodology to optimize ABP estimators using electrocardiogram and photoplethysmography signals is proposed. For this, the XGBoost machine learning model, optimized with Bayesian techniques, is executed in a Graphics Processing Unit, which drastically reduces execution time. The methodology is evaluated using the MIMIC-III Waveform Database. Systolic and diastolic pressures are estimated with mean absolute error values of 15.85 and 11.59 mmHg, respectively, similar to those of the state of the art. The main advantage of the proposed methodology with respect to others of the current state of the art is that it allows the optimization of the estimator model to be performed automatically and more efficiently at the computational level for the data available.**

***Clinical Relevance*— This approach has the advantage of using noninvasive methods to continuously monitor patient's arterial blood pressure, reducing the risk for patients.**

I. Introduction

Cardiovascular problems are one of the leading causes of death worldwide. According to the World Health Organization, hypertension is estimated to cause 7.5 million deaths every year, that is, 12.8% of total deaths worldwide [1]. As population ages and adopts more sedentary lifestyles, cardiovascular problems are expected to grow from the current prevalence figures, already over 1 billion, to 1.5 billion people by 2025 [1]. Because the symptoms of hypertension are usually unnoticed, blood pressure values must be controlled to prevent possible damage. In cases where only routine checks are needed, e.g., once a month in aged patients, there exist well-established noninvasive methods such as the use of sphygmomanometers. In contrast, some patients in hospital, e.g., in ICUs or cath labs, require continuous measurement of blood pressure, for what an arterial catheter is used [2]. Catheterization is carried out under local anesthesia, the catheter being inserted through a small incision in the skin of the groin (femoral access) or the arm (humeral access). This invasive method puts patient health at risk, so noninvasive blood pressure measurement is of great interest [3].

There are currently numerous proposals for estimating blood pressure using machine learning techniques [4]. In these, Pulse Transit Time, Pulse Arrival Time, Pulse Wave Velocity, or photoplethysmography (PPG) signals, among others, are used as inputs. In this scenario, the need for optimizing estimator models to obtain more accurate results and to minimize execution times arises.

In this article, an XGBoost-based methodology to optimize arterial blood pressure (ABP) estimators using Bayesian techniques with electrocardiogram (ECG) and PPG signals is proposed. It allows obtaining more accurate data and achieving lower execution times than other model hyperparameters optimization methods. In addition, the methodology is suitable for execution in an NVIDIA A100 Graphics Processing Unit (GPU) [5], which drastically reduces the time required to estimate blood pressure. All these claimed contributions are supported by experimental results.

The remainder of the article is structured as follows. In Section II, the materials used, namely ICU database (MIMIC), estimation model (XGBoost), Bayesian techniques, and wavelet transform are described. Section III introduces the proposed methodology, including data preprocessing, feature extraction, and estimator set up. In Section IV, experimental results are presented and discussed. Finally, conclusions are summarized in Section V.

II. Materials

A. Data Source

The well-known MIMIC-III Waveforms Database [6] has been used. It was created by MIT and it contains data from a total of 25,328 intensive care unit stays at Beth Israel Deaconess Medical Centre. In this work, only the patients in this dataset whose entries include all the data required for the experiments in this work (namely PPG, ECG, and continuous ABP signals) have been selected. The work in [7] has been used as reference. This leads to a total of 12,000 records. Each patient has a variable length of recorded data. The sampling

*Research partially supported by Agencia Gallega de Innovación (GAIN) through "Proxectos de investigación sobre o SARS-CoV-2 e a enfermidade COVID-19 con cargo ao Fondo COVID-19" program, with Code Number IN845D-2020/29, and through "Axudas para a consolidación e estruturación de unidades de investigación competitivas e outras accións de fomento nos organismos públicos de investigación de Galicia e noutras entidades do sistema galego de I+D+i - GPC" with Code Number IN607B-2021/18.

J. A. González-Nóvoa (correspondence author: jose.gonzalez@iisgaliciasur.es), L. Busto (laura.busto@iisgaliciasur.es), P. Santana (pablo.santana@iisgaliciasur.es), P. Juan-Salvadores (pablo.juan@iisgaliciasur.es) and C. Veiga (cesar.veiga@iisgaliciasur.es) are with the Cardiovascular Research Group of Galicia Sur Health Research Institute (IIS Galicia Sur), Vigo, 36213 Spain.

J. Fariña (jfarina@uvigo.es) and J. J. Rodríguez-Andina (jjrdguez@uvigo.es) are with the Department of Electronic Technology, University of Vigo, 36310 Vigo, Spain.

V. Jiménez (victor.alfonso.jimenez.diaz@sergas.es) and A. Íñiguez (andres.iniguez.romo@sergas.es) are with the Cardiology Department of Hospital Álvaro Cunqueiro (SERGAS), 36213 Vigo, Spain.

frequency of all signals is 125Hz.

B. Estimation Model: XGBoost

XGBoost [8] belongs to the category of Gradient Boosting techniques in Ensemble Learning, that is, a collection of predictors that combine multiple models in order to achieve better prediction accuracy. Boosting techniques attempt to correct the errors made by previous models in successive ones via additional weighting. Unlike other boosting algorithms where the respective weights of misclassified branches are increased, in Gradient Boosting algorithms a loss function is optimized instead. XGBoost optimizes the following objective function at each iteration t.

$$L(t) = \sum l\big(y_i, \tilde{y}_i(t-1) + f_t(x_i)\big) + \Omega(f_t) \quad (1)$$

where l is a differentiable convex loss function that must be transformed into another one in an Euclidean domain by using Taylor's Theorem, the pair (y_i, x_i) is the training set, $\tilde{y}_i$ is the final prediction, and $\Omega(f_i)$ is the regularization term used to penalize more complex models through both Lasso and Ridge regularizations and to prevent overfitting.

Once this optimization is performed the algorithm builds the next learner, which achieves the maximum possible loss reduction without exploring all tree structures, but rather by building a tree by applying the Exact Greedy Algorithm. This algorithm consists of three steps. It starts with a single root that contains all training samples. Then, it iterates over all features and values per feature, evaluating each possible split loss reduction. Finally, the stop condition is checked, stopping the branch from growing if the gain for the best split is not positive, otherwise execution continues. A more detailed explanation may be found in [8].

XGBoost stands out for its ability to obtain the best results in different benchmarks, and is one of the best-optimized algorithms for computing parallelization, which makes it one of the most used in recent biomedical works [9]–[11]. In addition, it has support for GPUs, which allows the capacity of the algorithm to be fully exploited.

Fitting XGBoost requires setting three types of parameters, namely general, booster, and learning task parameters. General parameters specify the booster used, commonly a tree or linear model. Booster parameters depend on the selected booster and define its internal configuration parameters, such as learning ratio or number of estimators. Learning task parameters decide on the learning scenario, specifying the corresponding learning objective.

C. Bayesian Optimization

Bayesian optimization is a technique used to optimize the hyperparameters of a model [12]. Although there are other techniques to do this [13], such as random search or grid search, they have some drawbacks. The first one randomly tests only a certain number of different combinations. It has the disadvantage of not following any criterion for deciding which combination of hyperparameters to test. Grid search goes through all hyperparameter combinations, hence being very computationally expensive.

The techniques based on Bayesian optimization are more efficient when performing the search. In these techniques, a probabilistic model is used that takes into account previous evaluation results, i.e., the surrogate model, which is a multivariate Gaussian stochastic model. The stages of the optimization process are:

1. The surrogate model of the objective function is built;
2. The hyperparameters that provide the best results in the surrogate model are sought;
3. These hyperparameters are applied to the real objective function;
4. The surrogate model is updated incorporating the new results;
5. The process is repeated until the maximum number of iterations defined or the predefined conditions are reached.

In this work the open-source package Hyperopt [14], which uses Bayesian optimization as search technique, has been used.

D. Wavelet Decomposition

The wavelet transform is one of the most used in the biomedical field [15]. It is characterized by working in the time – frequency spectrum, unlike others such as the Fourier transform that only works in the frequency spectrum, which limits the information that can be extracted, especially from signals that consist of sharp peaks and discontinuities.

The analysis is performed using a function called base wavelet, which decomposes the target signal into multiple components at different scales. The wavelet decomposition of a function $f(t)$ at scale a, and position τ is given by:

$$W_f(\tau, a) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) \Psi^* \left(\frac{t-\tau}{a}\right) dt \quad (2)$$

where $\Psi(t)$ is the wavelet base and $1/\sqrt{a}$ is the energy normalization.

E. Hardware

In this work, two hardware platforms have been used, namely an Asus ESC4000A-E10 server with an AMD Epyc 7002 CPU consisting of 24 cores and working at 2,800 MHz and an NVIDIA A100 GPU with 40 GB of internal memory, 108 multiprocessors and a maximum number of 221,184 threads.

III. Method

In order to obtain the best parameters for a XGBoost classifier that estimates both systolic (SBP) and diastolic (DBP) blood pressure from PPG and ECG signals four stages must be carried out. The first one is data preprocessing. It includes data filtering and the extraction of SBP and DBP values to train and test the estimator models. The second stage is the extraction of features from both ECG and PPG signals. Once feature extraction has been completed, the optimization of the estimator model hyperparameters is performed by using Bayesian techniques. Finally, SBP and DBP are estimated, comparing the accuracy obtained using Bayesian optimization techniques with that obtained with default values of XGBoost hyperparameters. There are actually two different models for estimating SBP and DBP, respectively.

A. Data Preprocessing

This stage includes data filtering, signal splitting, and

golden standard calculation. Signals are split in temporal episodes (epochs), for which a 20 s window analysis is proposed. In addition, the variables used both to fit the estimator and as golden standard in the testing are calculated. In this work, SBP and DBP values are estimated as the mean of maxima (systolic) and minima (diastolic) ABPs for each epoch.

B. Feature Extraction

Feature extraction of PPG signal is carried out using the wavelet transform. Specifically, all coefficients of the transform are used as features, and statistics are computed, both from the untransformed signal and from each wavelet level. These are: median, mean, standard deviation, variation, entropy, energy, contrast, inverse different moment, and homogeneity. After performing several tests, it was decided to use Haar as wavelet base. Moreover, this in one of the most used in other works consulted [16]–[18].

In the case of ECG signals, on one hand feature extraction is carried out in the same way as in the case of PPG signals and, on the other hand, the open-source Heartpy package [19] is used, which makes it possible to extract additional features from ECG signals, namely beats per minute, inter-beat interval, standard deviation, heart rate variability, Poincaré plot measurements, and breathing rate.

C. Pressure Estimator Set Up Using XGBoost

After the preprocessing stage, each XGBoost Regressor model is fitted to estimate SBP and DBP for each epoch. Bayesian optimization is used for hyperparameter tuning. This allows the set of parameters that provides the best performance to be identified. After that, the model is trained and tested using the MIMIC dataset.

1) Hyperparameter Tuning

In order to use Bayesian optimization, three elements have to be defined, namely search space, loss function, and limit number of evaluations the model performs until it stops looking for the optimal combination of hyperparameters. In this work, the space where the Bayesian optimizer searches for the best combination of hyperparameters includes the following XGBoost hyperparameters: learning rate, maximum depth, minimum child weight, maximum delta step, subsample ratio, lambda region, alpha region, scale weight, maximum number of leaves, and number of estimators. The loss function defines the metric to identify the best set of hyperparameters. The mean absolute error (MAE) obtained after a 5-fold cross validation is used as loss function in this case. Finally, the limit number of evaluations has been set to 500.

2) ABP Estimation

Once the best set of hyperparameters is obtained, the next step is to use it for SBP and DBP estimation using the XGBoost Regression model. To evaluate the performance of the model, the cross-validation method is used, with 5 folds to avoid a lucky train – test split.

IV. Results and Discussion

Results are provided in two ways, namely the best set of hyperparameters obtained, and the quality of the ABP estimation model for ICU patients. As mentioned above, the MIMIC dataset has been used and the selected features have been extracted from 20 s epochs.

A. Hyperparameter Selection

Using Bayesian optimization and MAE as loss function, the best set of hyperparameters has been determined. Table I shows the hyperparameter search space limits and the type and best set of hyperparameters. The uniform, log uniform and q uniform search spaces return real values uniformly distributed between defined limits. Log uniform is more suitable for geometric series, whereas uniform and q uniform are more suitable for arithmetic series, with the difference that q uniform returns round values, so the selection of the search space depends on the hyperparameter type.

TABLE I. Hyperparameters Search Space

Hyperparameter	Search space type	Limits		Best values	
		Min	***Max***	***DBP***	***SBP***
Learning rate	loguniform	-8	0	0.018	0.025
Max depth	quniform	1	15	14	6
Min child weight	quniform	0	10	1	6
Max delta step	quniform	0	10	10	3
Subsample ratio	uniform	0.1	1	0.691	0.865
Lambda region	uniform	0.1	1	0.104	0.744
Alpha region	uniform	0.1	1	0.490	0.189
Scale weight	uniform	0.1	1	0.459	0.359
Max number of leaves	quniform	0	10	2	1
Number of estimators	quniform	1	1000	745	951

B. Estimation Model Quality

Once the best set of hyperparameters is determined, the next step is the evaluation of the quality of the estimation model. A 5-fold cross-validation method has been used to evaluate estimator performance. Table II shows the results obtained as well as the corresponding GPU and CPU execution times. MAE is 15.85 mmHg for SBP and 11.59 mmHg for DBP. These values are within the current state of the art range [7], [20]. As expected, the results obtained using Bayesian optimization techniques are more accurate than those obtained using the default values of the estimator model hyperparameters. The use of a GPU dramatically reduces execution times by 97%, as shown Table II.

TABLE II. Estimation Model Performance

ABP	MAE		Execution time (Optimal hyperparameters)	
	Default hyperparameters	***Optimal hyperparameters***	***GPU***	***CPU***
SBP	17.14 mmHg	15.85 mmHg	25.90 s	13.34 min
DBP	13.01 mmHg	11.59 mmHg	22.57 s	12.05 min

V. Conclusion

This article presents a methodology to optimize ABP

estimators using Bayesian techniques with ECG and PPG signals. The stages of the methodology are data preprocessing, feature extraction, model optimization, and automatic pressure estimation.

Unlike other works of the current state of the art, this methodology allows the optimization of the estimator model to be performed automatically and efficiently at the computational level for the data available, instead of using the model in a generic way. The optimization technique used, Bayesian optimization, stands out for its calculation efficiency with respect to others widely used, such as random search or grid search. Random search has the disadvantage that it does not follow any criteria when searching for hyperparameters since it does so randomly. In addition, it does not run through the entire search space, so it does not guarantee that the combination of hyperparameters obtained is optimal. Grid search, which runs through the entire search space, is very expensive at the computational level, needing high computing times, which is a serious limitation when it comes to use it in the day to day of a hospital. The adaptation of the methodology to its use in GPU is also very relevant, further improving its computational efficiency.

Experimental results clearly confirm the validity and usefulness of the proposed method. It allows identifying the best set of parameters for a XGBoost estimator that predicts SBP and DBP with features extracted from PPG and ECG signals using wavelet decomposition in 20 s epochs. The estimator defined with those parameters and trained and tested with the MIMIC dataset provides very good MAE results, within the current state of the art range. In addition, GPU execution provides nearly 30x acceleration with regard to CPU execution.

As future work the proposed methodology will be improved in several ways, aiming at enhancing the performance of the estimator model. The Bayesian optimization methodology will be adapted for execution in multiple GPUs. We will also delve into the extraction of the characteristics of the PPG and ECG signals, analyzing the wavelet transform in more detail, and adding features from other sources. In addition, model hyperparameter selection will be improved and the option of using models others than XGBoost, such as Deep Learning techniques, and datasets others than MIMIC will be considered.

REFERENCES

[1] World Health Organization, "Raised blood pressure (SBP>=140 OR DBP>=90) (age-standardized estimate)," 2017. https://www.who.int/data/gho/data/indicators/indicator-details/GHO/raised-blood-pressure-(sbp-=140-or-dbp-=90)-(age-standardized-estimate) (accessed Dec. 10, 2021).

[2] B. Saugel, K. Kouz, A. S. Meidert, L. Schulte-Uentrop, and S. Romagnoli, "How to measure blood pressure using an arterial catheter: a systematic 5-step approach," *Crit Care*, vol. 24, no. 1, Apr. 2020, doi: 10.1186/S13054-020-02859-W.

[3] X. Quan *et al.*, "Advances in Non-Invasive Blood Pressure Monitoring," *Sensors*, vol. 21, no. 13, p. 4273, Jun. 2021, doi: 10.3390/s21134273.

[4] C. El-Hajj and P. A. Kyriacou, "A review of machine learning techniques in photoplethysmography for the non-invasive cuff-less measurement of blood pressure," *Biomedical Signal Processing and Control*, vol. 58, p. 101870, Apr. 2020, doi: 10.1016/j.bspc.2020.101870.

[5] J. Choquette, W. Gandhi, O. Giroux, N. Stam, and R. Krashinsky, "NVIDIA A100 Tensor Core GPU: Performance and Innovation," *IEEE Micro*, vol. 41, no. 2, pp. 29–35, Mar. 2021, doi: 10.1109/MM.2021.3061394.

[6] M. Saeed *et al.*, "Multiparameter Intelligent Monitoring in Intensive Care II: A public-access intensive care unit database*," *Critical Care Medicine*, vol. 39, no. 5, pp. 952–960, May 2011, doi: 10.1097/CCM.0b013e31820a92c6.

[7] M. Kachuee, M. M. Kiani, H. Mohammadzade, and M. Shabany, "Cuff-less high-accuracy calibration-free blood pressure estimation using pulse transit time," in *2015 IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2015, vol. 2015-July, pp. 1006–1009. doi: 10.1109/ISCAS.2015.7168806.

[8] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, vol. 13-17-Augu, pp. 785–794, Mar. 2016, doi: 10.1145/2939672.2939785.

[9] Z. S. H. Abad, D. M. Maslove, and J. Lee, "Predicting Discharge Destination of Critically Ill Patients Using Machine Learning," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 3, pp. 827–837, Mar. 2021, doi: 10.1109/JBHI.2020.2995836.

[10] D. Bertsimas, L. Mingardi, and B. Stellato, "Machine Learning for Real-Time Heart Disease Prediction," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 9, pp. 3627–3637, Sep. 2021, doi: 10.1109/JBHI.2021.3066347.

[11] J. A. González-Nóvoa *et al.*, "Using explainable machine learning to improve intensive care unit alarm systems," *Sensors*, vol. 21, no. 21, p. 7125, Oct. 2021, doi: 10.3390/s21217125.

[12] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. de Freitas, "Taking the Human Out of the Loop: A Review of Bayesian Optimization," *Proceedings of the IEEE*, vol. 104, no. 1, pp. 148–175, Jan. 2016, doi: 10.1109/JPROC.2015.2494218.

[13] P. Liashchynskyi and P. Liashchynskyi, "Grid Search, Random Search, Genetic Algorithm: A Big Comparison for NAS," Dec. 2019, Accessed: Dec. 10, 2021. [Online]. Available: https://arxiv.org/abs/1912.06059v1

[14] J. Bergstra, D. Yamins, and D. D. Cox, "Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures," in *30th International Conference on Machine Learning, ICML 2013*, 2013, vol. 28, no. PART 1, pp. 115–123.

[15] S. A. P. Haddad and W. A. Serdijn, *Ultra Low-Power Biomedical Signal Processing*. Dordrecht: Springer Netherlands, 2009. doi: 10.1007/978-1-4020-9073-8.

[16] C. Yang, C. Veiga, J. J. Rodríguez-Andina, J. Fariña, A. Iñiguez, and S. Yin, "Using PPG Signals and Wearable Devices for Atrial Fibrillation Screening," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 11, pp. 8832–8842, 2019, doi: 10.1109/TIE.2018.2889614.

[17] H. B. Seidel, M. M. A. da Rosa, G. Paim, E. A. C. da Costa, S. J. M. Almeida, and S. Bampi, "Approximate Pruned and Truncated Haar Discrete Wavelet Transform VLSI Hardware for Energy-Efficient ECG Signal Processing," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 68, no. 5, pp. 1814–1826, May 2021, doi: 10.1109/TCSI.2021.3057584.

[18] M. Singla, S. Azeemuddin, and P. Sistla, "Accurate Fiducial Point Detection Using Haar Wavelet for Beat-by-Beat Blood Pressure Estimation," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 8, pp. 1–11, 2020, doi: 10.1109/JTEHM.2020.3000327.

[19] P. van Gent, H. Farah, N. van Nes, and B. van Arem, "HeartPy: A novel heart rate algorithm for the analysis of noisy signals," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 66, pp. 368–378, Oct. 2019, doi: 10.1016/j.trf.2019.09.015.

[20] A. M. Zakrzewski and B. W. Anthony, "Noninvasive blood pressure estimation using ultrasound and simple finite element models," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 9, pp. 2011–2022, Sep. 2018, doi: 10.1109/TBME.2017.2714666.