2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Introduction of Boosting Algorithms in Continuous Non-Invasive Cuff-less Blood Pressure Estimation using Pulse Arrival Time

Aayushman Ghosh[1], Tamaghno Chatterjee[1], and Sayan Sarkar

Abstract— **Blood Pressure (BP) is a critical biomarker for cardiorespiratory health. Conventional non-invasive BP measurement devices are mostly built on the principle of auscultation, oscillometry, or tonometry. The strong correlation between the Pulse Arrival Time (PAT) and BP has enabled unconstrained cuff-less BP monitoring. In this paper, we exploited that relationship for estimating Systolic Blood Pressure (SBP), Diastolic Blood Pressure (DBP), and Mean Arterial pressure (MAP) values. The proposed model involves extraction of PAT values by denoising the signals using advanced filtering techniques and finally employing machine learning algorithms to estimate cuff-less BP. The results are validated against Advancement of Medical Instrumentation (AAMI) standards and British Hypertension Society (BHS) protocols. The proposed method meets the AAMI standards in the context of estimating DBP and MAP values. The model's accuracy achieved Grade A for both MAP and DBP values using the CatBoost algorithm, whereas it achieved grade A for MAP and Grade B for DBP using the XGBoost algorithm based on the BHS standards.**

I. INTRODUCTION

Cardiovascular ailments (CVDs) are considered as one of the biggest threats to human lives, which are often aggravated by high BP. World Health Organization (WHO) reports, approximately 17.9 M people succumbed to death in 2016 [1] due to heart-related diseases. Continuous noninvasive Blood pressure (cNIBP) estimation proves to be an efficient diagnostic parameter for many diseases, such as stroke, hypertension, etc. [1],[2]. BP is affected by food-intake habits, mental state, physical workouts, etc., which proves the patient-specific nature of results [2],[6]. Continuous BP assessment methodologies are broadly categorized into two segments: Invasive and Noninvasive. The gold standard for continuous BP monitoring is invasive in nature [3]. The invasive measurement method provides accurate results for BP measurement, but it requires skilled physicians and has a high chance of triggering infections in patients [2],[3]. Hence, this method is strictly limited to clinical settings. 'Noninvasive methods for discrete BP estimation' involves conventional cuff-based BP measurement methods such as auscultation and oscillometry [3]. Careful monitoring of Korotkoff sound makes auscultation time-consuming as well as induces patient discomfort. This technique also limits the patient's mobility due to the inflatable cuff [3]. Cuff-based non-invasive methods such as Volume-clamp and Applanation tonometry [1]-[3] introduces continuous BP monitoring. These methods involve discomfort to patients, and are highly susceptible to motion artifacts [1]. Cuff-less BP estimation methodologies has emerged as an attractive surrogate in the continuous BP monitoring process [1]-[7],[9]-[16]. This methodology relies on the information coded on cardiorespiratory signals, generally, electrocardiogram (ECG), photoplethysmography (PPG) signal, and Arterial Blood Pressure (ABP) signal [2]. Several Cardiovascular parameters such as pulse transit time (PTT) [2],[5], pulse arrival time (PAT) [4],[7] and pulse wave velocity (PWV) [10], have proved effective in cNIBP measurement. The time delay between ECG R-peak and the characteristic point [7] on the PPG waveform is called the pulse arrival time (PAT). It is equal to the sum of PTT and the pre-ejection period (PEP) of the cardiac cycle [5],[7]. PEP comprises of the time needed to convert the electrical signal into a mechanical pumping force and the isovolumetric contraction of the left ventricle to open the aortic valve [2]. The effect of PEP on the overall PAT decreases with the distance from the heart [5]. The other alternative way for BP estimation is exploiting the PWV value [10]. 'PWV' is calculated as the ratio of the distance between the two arterial measurement sites (D) and PTT as in [9] ($PWV = D/PTT$). Based on the above features, cNIBP can be estimated using Machine Learning (ML) algorithms and statistical models. Such developed systems require periodic cuff calibrations but can still be useful for hypertension screening based on numerous out-of-clinic measurements [6]. Recent literatures [6],[4] worked on the maximum calibration period that would not compromise accuracy, also stating the acceptable error limits in light of measurement averaging for PTT/PAT-based systems. However, patient-specific calibration procedures are not mentioned in detail in this brief. In this paper, advanced signal processing techniques are employed on ECG and the first derivative of PPG signals (dPPG) [7] to extract the PAT features. PAT features serves as an input to the ML algorithms, which ultimately estimates cNIBP.

II. MEASUREMENT METHODOLOGY

A. Database Description

We used Multi-parameter Intelligent Monitoring in Intensive Care (MIMIC) II waveform database, provided by Physionet [8]. Each data file contained a set of three signals: standard bipolar lead II ECG signal, fingertip PPG signal, and invasive ABP signal. Based on the Physionet website, all the signals in the MIMIC-II database were sampled at a sampling rate of 125 Hz [8],[9]. The signals span was reduced by extracting the signal between the first and last 10 seconds, based on observing the clinical setup in an ICU and also on consultation with the physicians. The first few seconds of the readings are affected by the patient's movement,

[1]signifies equal contribution in this paper.

The Authors are with Wecare Medservice LLP, Kolkata, India. wecaremedservicellp@gmail.com

introduced by the discomfort induced due to various sensors and catheters. From the extracted signals, a 60-second long window is selected randomly for further extraction. The final dataset contained 90 records corresponding to 90 different patients. Each file contained three recorded signals (ECG, PPG, ABP). In creating the supervised learning task dataset, the ECG and dPPG signals are used to derive PAT_d values. ABP signal gives the target SBP and DBP values.

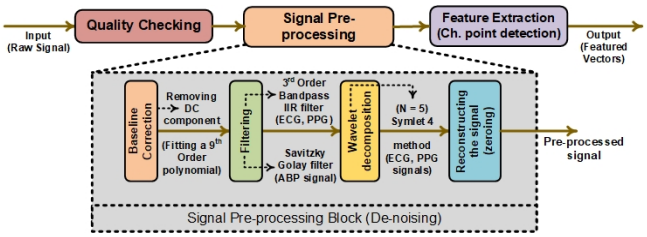

Fig. 1. Feature extraction pipeline for Continuous Cuff-less BP estimation

B. Pre-processing procedures

The raw signals contain undesired noise content. Hence, a preprocessing block was necessary to de-noise and enhance the signal's quality. In the preprocessing block, the input ECG and PPG signals are passed through a 3rd order Butterworth band-pass IIR filter. The filter has lower (LCF) and upper (UCF) cut-off frequencies for PPG signals (1.5-20 Hz), ECG signals (0.05-40 Hz). The cut-off frequencies were determined after a thorough inspection of the spectrogram of both the signals. The filter type and order were chosen as per [15]. The algorithm involves zero-phase filtering (applying the filter in both directions) to eliminate the time delay caused by IIR filters [15]. The filtered ECG and PPG signals are then decomposed employing undecimated wavelet transform with Symlet 4 mother wavelet to achieve 5 levels of decomposition [17]. The components corresponding to the very low-frequency range of 0–0.25 Hz (baseline wandering) and high frequencies between 250-500 Hz (power-line harmonics and the muscular activity-related artifacts) are eliminated by zeroing decomposition coefficients associated with the transformed signals [9]. The reconstructed signals showed prominent peaks and valleys (R peaks in ECG and systolic and diastolic peaks in PPG), and improved signal to noise ratio [17]. The amplitude of ECG and PPG signals are then normalized. A fourth-order, 19 frame length Savitzky-Golay filter was applied to the ABP waveforms to smooth it and preserve the systolic and diastolic peaks. The whole preprocessing pipeline is given in Fig. 1.

C. Extraction of PAT_d, SBP and DBP values

PAT is usually defined as the time delay between ECG R peak (proximal signal) and a characteristic point on PPG (distal signal) [5]. Four different points on the PPG waveform are usually chosen as the characteristic point for initial evaluation [7], namely (i) PPG peak, (ii) PPG foot, (iii) PPG d-peak, (iv) PPG dd-peak. Here 'd' stands for numerical 'derivative' of the PPG signal. However, among these four characteristic points, PPG d and PPG dd-peaks are easier to detect as PPG derivatives usually form sharp peaks [4],[7]. PPG derivative peaks as characteristic points for PAT calculation provide the lowest variation in calculated PAT values [7]. Time delay calculated between the ECG R peaks and the PPG d-peaks are generally termed as PAT_d [7], as shown in Fig. 2. Calculated Peak to Peak intervals [4] using PPG d-peak showed a better correlation with ECG RR intervals than other characteristic points. A total of 4624 PAT_d data points were obtained from 90 datasets. SBP and DBP values were extracted from the ABP signal by detecting the systolic and diastolic peaks as shown in Fig. 2.

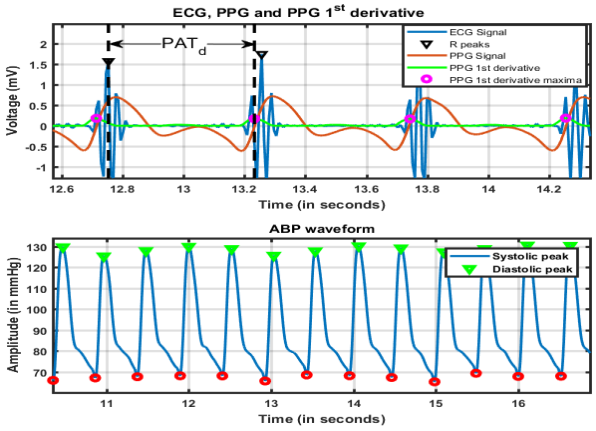

Fig. 2. Processed ECG, PPG (red) and first derivative of PPG signal (green) with characteristic point (top); ABP signal with systolic and diastolic peak

D. Machine Learning model Architectures

During our study, we selected 3 highly cited regression algorithms (SVM, AdaBoost, Random Forest) [9]-[11], [13],[14] and introduced 2 new boosting based regression algorithms – XGBoost [12], and CatBoost [18],[19]) in the domain of cuff-less BP estimation. Python's 'Scikit-learn library' is used for the first three algorithms [9]. We divided the entire dataset into two parts: a) training part (70%) and b) testing part (30%). Our models were trained using only one feature vector (PAT_d) to reduce computational complexity [6]. Predicted values were compared with the test dataset's target values. Hyper-parameters associated with the models were optimized using a random search algorithm with 5 fold cross-validation. A brief description of the algorithms used here are as follows -

Support Vector Machine (SVM): A supervised learning algorithm performs significantly well on datasets with a non-linear feature-target relationship. This algorithm maps feature vectors to a higher dimensional space where the feature-target relationship is linear and then fits a hyper-plane to the data [9]. We constructed a 'Support Vector Regression model' using 'ε-insensitive loss function' and used 'Radial Basis' Function kernel [9].

Random Forest Regression (RFR): This ensemble learning method constructs multiple weak learners (decision trees) [9], [13] during training to solve regression problems. The final prediction is based on the mean of all predictions.

Adaptive Boosting (AdaBoost): It is a boosting-based ensemble learning meta-algorithm and works by constructing multiple weak learners. The final outcome is a weighted sum of each learner's outcomes [9],[13]. AdaBoost model is more prone to overfitting compared to other models. So, there is a chance of the model's failure to track the original trend of the data, hence poor performance on test data in terms of prediction. The 'AdaBoost regressor model' used in this work is based on the 'AdaBoost R_2' algorithm.

XGBoost: XGBoost or Extreme Gradient Boosting algorithm is an integrated highly efficient gradient boosting algorithm that combines the base function and weight with boosting idea. It provides a parallel tree boosting structure to solve classification-regression problems in a fast and efficient manner [12]. In literature [12], this algorithm is used for 20 subjects in dual PPG-based BP estimation, relying on the time-domain features. As the distance between two measurement points is small, the reported error is also less in this paper. However, this algorithm has not been used in the ECG-PPG combination yet, as per the author's best knowledge. The distance between measurement sites is larger in the ECG-PPG combination case, so the chances of error is high. Still, our model using XGBoost achieved significant accuracy, as evident from the result. The generalization ability of XGBoost is strong, and it can also handle training and prediction with a large number of missing value cases.

Category Boost (CatBoost): Highly efficient machine learning algorithm (developed by 'Yandex' researchers [18]) based on gradient boosting principle using decision trees. 'Catboost' is used in recent literature [19] for 'Cardiovascular risk prediction'. It approaches a problem (classification/regression) [18] using a permutation-driven method (ordered boosting), a newer strategy compared to classical algorithms.

III. RESULTS

A model's accuracy is defined by MAE (Mean Absolute Error) and RMSE (Root Mean Square Error). Occasionally, SD (Standard Deviation) value is reported in place of RMSE [9]. The results obtained for different highly used algorithms are presented and compared with other existing works in Table I. Table II compares the best result obtained by others using a specific algorithm, with algorithms used in this study.

A. Performance of XGBoost and CatBoost Algorithm

The accuracy obtained by XGBoost algorithm in estimating cuff-less BP is mentioned in Table II. The Pearson's correlation coefficient [9] between actual and predicted BP values in three BP categories (SBP, DBP, MAP) are found to be: 0.41, 0.59, 0.62, respectively. In this brief, CatBoost is introduced as a novel method for estimating cuff-less BP. The accuracy achieved by this algorithm in estimating cuff-less BP is mentioned in Table II. This algorithm achieved better accuracy (MAE) than highly cited existing algorithms [9]-[14]. The Pearson's correlation coefficient value for the Catboost algorithm between target BP values and predicted BP values for the three BP categories (SBP, DBP, MAP) is found to be 0.37, 0.55, and 0.61. The Bland-Altman plots in Fig. 3 depicts the Mean Error (M) and SD for both the models, along with the percentage of data points lying within the $M \pm 1.96 \times SD$ range (limits of acceptance) as shown in red.

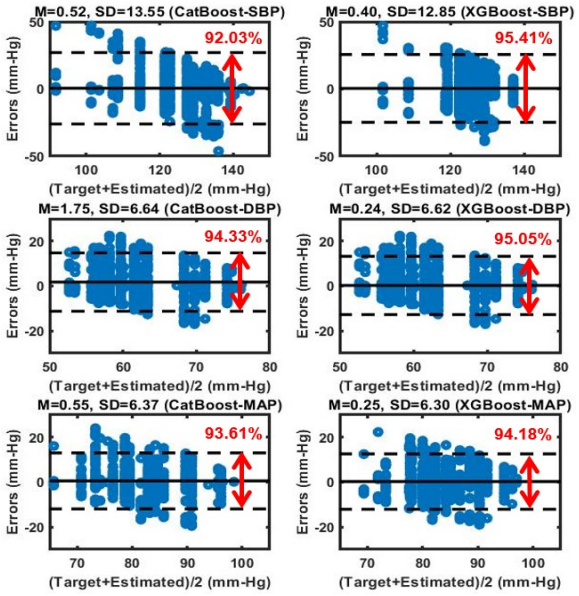

Fig. 3. Bland–Altman plots of SBP, DBP, MAP. The middle line represents mean error (M) and dashed lines represent $M \pm 1.96 \times SD$ range $(N = 90)$.

B. Evaluation with BHS and AAMI standards

The results obtained using XGBoost and CatBoost algorithm are validated against both the BHS and the AAMI standards. BHS grades BP measurement's accuracy based on their cumulative error percentage under the thresholds of 5 mmHg, 10 mmHg, and 15 mmHg [20]. Table III compares our models using XGBoost and CatBoost algorithm with BHS standard. CatBoost algorithm achieved Grade-A for both DBP and MAP measurement according to the BHS standard. XGBoost algorithm achieved Grade-A for MAP measurement and Grade-B for DBP measurement. AAMI grades BP measurement methodology based on their Mean error (M) and standard deviation (SD) of errors. According to the AAMI standard for cuff-less BP measurement devices, $(M \pm SD)$, has to be $(\leq (5 \pm 8)$ mmHg), and the number of subjects (N) should be ≥ 85. As evident from Fig. 3, our models achieve M values within the acceptable limit. Regarding the SD criterion, the DBP and MAP values are within the maximum acceptable margin, but the SD of SBP is out of the maximum acceptable limit.

IV. CONCLUSION

The implementation of PAT-based cuff-less blood pressure estimation architecture accurately predicts blood pressure values non-invasively. Here, we presented a method for continuous BP monitoring, helpful for non-clinical environments. The proposed method involves the processing of the ECG and PPG signals to obtain the PAT_d values and employing them as input to the regression models. This study introduced two novel algorithms, namely CatBoost and

TABLE I

COMPARATIVE ERROR ANALYSIS OF BP USING CONVENTIONAL ALGORITHMS

ML algorithms	Kachuee [9],[10]			Thambiraj [13]			Kurylyak [11]			Zhang [14]			This Work		
Errors (mmHg)	SBP	DBP	MAP	SBP	DBP	MAP	SBP	DBP	MAP	SBP	DBP	MAP	SBP	DBP	MAP
AdaBoost															
MAE	11.17	5.35	-	14.20	11.99	10.95	-	-	-	-	-	-	10.48	5.41	4.97
RMSE / (SD)	(10.09)	(6.14)	-	17.17	14.32	13.07	-	-	-	-	-	-	12.81	6.62	6.34
SVM															
MAE	12.26	5.91	7.52	14.85	10.98	10.70	13.60	7.70	-	11.64	7.61	-	10.17	5.05	4.78
RMSE / (SD)	(10.32)	(5.78)	(9.54)	18.92	14.41	13.76	13.60	7.90	-	(8.20)	(6.78)	-	13.29	6.79	6.35
RFR															
MAE	11.80	5.83	-	10.83	8.43	8.35	-	-	-	-	-	-	10.48	5.30	4.93
RMSE / (SD)	(9.87)	(5.71)	-	14.96	11.62	11.05	-	-	-	-	-	-	12.80	6.55	6.29

TABLE II

COMPREHENSIVE ERROR ANALYSIS OF OLD AND NEWLY INTRODUCED ALGORITHM

Machine Learning Algorithms	Comparison with state-of-the-art from Table I	MAP		DBP		SBP	
		MAE	RMSE	MAE	RMSE	MAE	RMSE
Adaptive Boosting Algorithm (AdaBoost)	best from TABLE I	10.95	13.07	5.35	14.32	11.17	17.17
	This work	5.02	6.34	5.51	6.58	10.48	12.81
Support Vector Machine (SVM)	best from TABLE I	7.52	13.76	5.91	7.90	11.64	13.60
	This work	4.82	6.54	5.05	6.79	10.17	13.29
Random Forest Regression (RFR)	best from TABLE I	8.35	11.05	5.83	11.62	10.83	14.96
	This work	4.93	6.29	5.30	6.55	10.48	12.80
Extreme Gradient Boosting Algorithm (XGBoost)	Che [12]	-	-	3.11	3.78	1.56	3.39
	This work	4.90	6.32	5.34	6.62	10.41	12.83
Category Boosting Algorithm (CatBoost)	This work	4.71	6.38	5.02	6.86	10.07	13.51

TABLE III

EVALUATION AGAINST BHS STANDARD [20]

ML Algorithms		Cumulative Error Percentage (Rounded Off)		
		≤ 5 mmHg	≤ 10 mmHg	≤ 15 mmHg
CatBoost/ (XGBoost)	DBP	60% / (55%)	85% / (86%)	96% / (98%)
	MAP	63% / (60%)	88% / (89%)	97% / (98%)
	SBP	35% / (30%)	61% / (55%)	78% / (75%)

XGBoost, in BP estimation, which achieves better results than other conventional algorithms. Results obtained are in accordance with both AAMI and BHS standards.

REFERENCES

[1] A. Stojanova *et al.*, "Continuous Blood Pressure Monitoring as a Basis for Ambient Assisted Living (AAL) - Review of Methodologies and Devices," in *J. Med Syst.* vol. 43, no. 2, Jan. 2019.

[2] T. Le *et al.*, "Continuous Non-Invasive Blood Pressure Monitoring: A Methodological Review on Measurement Techniques," in *IEEE Access*, vol. 8, pp. 212478-212498, 2020.

[3] A. Argha *et al.*, "Artificial Intelligence Based Blood Pressure Estimation From Auscultatory and Oscillometric Waveforms: A Methodological Review," in *IEEE Reviews in Biomedical Engineering*, 2020.

[4] F. S. Cattivelli and H. Garudadri, "Noninvasive Cuff-less Estimation of Blood Pressure from Pulse Arrival Time and Heart Rate with Adaptive Calibration," *2009 Sixth International Workshop on Wearable and Implantable Body Sensor Networks*, Berkeley, CA, USA, pp. 114-119.

[5] R. Mukkamala *et al.*, "Toward Ubiquitous Blood Pressure Monitoring via Pulse Transit Time: Theory and Practice," in *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 8, pp. 1879-1901, Aug. 2015.

[6] R. Mukkamala and J. -O. Hahn, "Toward Ubiquitous Blood Pressure Monitoring via Pulse Transit Time: Predictions on Maximum Calibration Period and Acceptable Error Limits," in *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 6, pp. 1410-1420, June 2018.

[7] S. Rajala, T. Ahmaniemi, H. Lindholm and T. Taipalus, "Pulse arrival time (PAT) measurement based on arm ECG and finger PPG signals - comparison of PPG feature detection methods for PAT calculation," *Annual Int Conf IEEE Eng Med Biol Soc (EMBC)*, 2017, pp. 250-253.

[8] A. L. Goldberger *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet : Components of a new research resource for complex physiologic signals," in *Circulation*, vol. 101, no. 23, pp. 215–220, 2000.

[9] M. Kachuee *et al.*, "Cuffless Blood Pressure Estimation Algorithms for Continuous Health-Care Monitoring," in *IEEE Transactions on Biomedical Engineering,* vol. 64, no. 4, pp. 859-869, 2017.

[10] M. Kachuee *et al.*, "Cuff-less high-accuracy calibration-free blood pressure estimation using pulse transit time," *2015 IEEE International Symposium on Circuits and Systems*, Lisbon, Portugal, pp. 1006-1009.

[11] Y. Kurylyak, F. Lamonaca and D. Grimaldi, "A Neural Network-based method for continuous blood pressure estimation from a PPG signal," *2013 IEEE International Instrumentation and Measurement Technology Conference (I^2MTC)*, 2013, pp. 280-283.

[12] X. Che *et al.*, "Continuous Blood Pressure Estimation from Two-Channel PPG Parameters by XGBoost," *IEEE International Conference on Robotics and Biomimetics*, Dali, China, 2019, pp. 2707-2712.

[13] G. Thambiraj *et al.*, "Investigation on the effect of Womersley number, ECG and PPG features for cuff- less blood pressure estimation using machine learning," in *Biomed. Signal Process. Control.*, vol. 60, 2020.

[14] Y. Zhang and Z. Feng, "A SVM method for continuous blood pressure estimation from a PPG signal," *9th International Conference on Machine Learning and Computing*, pp. 128–32, Singapore, Feb. 2017.

[15] A. Esmaili, M. Kachuee and M. Shabany, "Nonlinear Cuffless Blood Pressure Estimation of Healthy Subjects Using Pulse Transit Time and Arrival Time," in *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 12, pp. 3299-3308, Dec. 2017.

[16] S. Sarkar, A. Ghosh and S. S. Ghosh, "Study of Cardiorespiratory and Sweat Monitoring Wearable Architecture for Coal Mine Workers," *IEEE REGION 10 CONFERENCE (TENCON)*, 2020, pp. 355-360.

[17] S. Li and J. Lin, "Optimal De-noising Algorithm for ECG Using Stationary Wavelet Transform," *World Congress on Computer Science and Information Engineering*, Los Angeles, USA, 2009, pp. 469-473.

[18] Anna Veronika Dorogush *et al.*, "CatBoost: gradient boosting with categorical features support," *arXiv abs/1810.11363*, 2018.

[19] H. R. H. Al-Absi *et al.*, "Risk Factors and Comorbidities Associated to Cardiovascular Disease in Qatar: A Machine Learning based Case-Control Study," in *IEEE Access*, vol. 9, pp. 29929-29941, 2021.

[20] E. O. Brien *et al.*, "The british hypertension society protocol for the evaluation of automated and semi-automated blood pressure measuring devices with special reference to ambulatory systems." in *J. Hypertens.*,vol. 8, no. 7, pp. 607–619, 1990.