RESEARCH ARTICLE

Blood Pressure Estimation From Photoplethysmography by Considering Intra- and Inter-Subject Variabilities: Guidelines for a Fair Assessment

THIAGO BULHÕES DA SILVA COSTA [1], **FELIPE MENEGUITTI DIAS** [2],
DIEGO ARMANDO CARDONA CARDENAS [2], **MARCELO ARRUDA FIUZA DE TOLEDO** [2],
DANIEL MÁRIO DE LIMA [3], **JOSE EDUARDO KRIEGER** [2], **AND**
MARCO ANTONIO GUTIERREZ [2]

[1] Centro de Matemática, Computação e Cognição, Universidade Federal do ABC, São Bernardo do Campo, São Paulo 09606-045, Brazil
[2] Instituto do Coração, Hospital das Clínicas, Faculdade de Medicina, Universidade de São Paulo, São Bernardo do Campo, São Paulo 09606-070, Brazil
[3] Medical Center Information Technology, NYU Langone Health, New York, NY 10016, USA

Corresponding author: Thiago Bulhões da Silva Costa (thiagobscosta@gmail.com)

ABSTRACT Cardiovascular diseases are the leading causes of death, and blood pressure (BP) monitoring is essential for prevention, diagnosis, assessment, and treatment. Photoplethysmography (PPG) is a low-cost opto-electronic technique for BP measurement that allows the acquisition of a modulated light signal highly correlated with BP. There are several reports of methods to estimate BP from PPG with impressive results; in this study, we demonstrate that the previous results are excessively optimistic because of their train/test split configuration. To manage this limitation, we considered intra- and inter-subject data arrangements and demonstrated how they affect the results of feature-based BP estimation algorithms (i.e., XGBoost, LightGBM, and CatBoost) and signal-based algorithms (i.e., Residual U-Net, ResNet-18, and ResNet-LSTM). Inter-subject configuration performance is inferior to intra-subject configuration performance, regardless of the model. We also showed that, using only demographic attributes (i.e., age, sex, weight, and subject index number), a regression model achieved results comparable to those obtained in an intra-subject scenario. Although limited to a public clinical database, our findings suggest that algorithms that use an intra-subject setting without a calibration strategy may be learning to identify patients and not predict BP.

INDEX TERMS Blood pressure, photoplethysmography, wearables.

I. INTRODUCTION

Chronic cardiovascular diseases such as hypertension, hyperlipidemia, and atherosclerosis—combined with aging, overweight, and diabetes—are major risk factors for severe conditions such as stroke, heart failure, and myocardial infarction. Together, these conditions constitute the leading causes of human death, overtaking cancer mortality [1], [2]. Because this is a global health problem, blood pressure (BP) monitoring is essential for prevention, diagnosis, assessment,

and treatment and can predict and avoid acute events [3], [4]. Observational studies with patients aged 40 to 89, for example, indicated that the mortality caused by ischemic heart disease, stroke, and other vascular diseases grow linearly with BP, almost doubling the risk for every 20 mmHg or 10 mmHg increase in systolic and diastolic BP (SBP and DBP), respectively [5].

Reference devices for BP measurements are aneroid, mercury, and electronic sphygmomanometers with appropriate cuff and bladder sizes matched to the arm circumference [6]. Although there are automated devices, clinical and ambulatory guidelines advise that only trained healthcare

The associate editor coordinating the review of this manuscript and approving it for publication was Eyuphan Bulut [.].

professionals should measure and evaluate BP to determine and monitor cardiovascular risks [4], [7], [8]. However, these requirements render continuous BP monitoring difficult and impractical. Furthermore, sphygmomanometers (though they are adapted to gauge out-of-office BP in daily life) can only provide non-continuous measurements through repeated cuff inflation, which causes discomfort and even pain when the measurement requires a high level of air pressure. These devices make nighttime assessment complex and often inaccurate, given that nocturnal cuff inflation may disturb sleep, indirectly influencing BP [9].

Despite the difficulty of continuous monitoring, when recorded using an intra-arterial catheter, BP is a highly dynamic physiological variable, consisting of a sequence of pulse waves changing its frequency and amplitude over time windows within 24 hours: from beat to beat, minute to minute, hour to hour, and even from day to night [10]. In addition to these short-term fluctuations, population studies indicated that BP might fluctuate over days, weeks, months, and even years—revealing a complex interaction between environmental/behavioral factors and cardiovascular/physiological mechanisms [11]. Moreover, because sustained or sudden increases in BP variability may be associated with underlying pathological conditions, assessing these fluctuations can help guide clinical and prognostic decisions [12]. For example, reliable out-of-office BP measurement is essential to minimize "white coat hypertension" and to diagnose masked hypertension [13]. It is possible to perform non-invasive, continuous BP measurements using the volume-clamp method or the artery applanation tonometry [14], [15]. However, these methods require bulky apparatuses, can be uncomfortable for users and are not feasible devices for day-to-day use [16].

On the other hand, photoplethysmography (PPG) can be adapted to a wristband or a ring [17], [18] and has been considered as an approach for continuous BP monitoring outside of clinical settings to assess hypertension and other cardiovascular diseases [19], [20]. PPG is an opto-electronic method by which alterations in blood volume can be detected in the microvasculature of the subcutaneous tissue. It uses light-emitting diodes at the green, red, or infrared wavelengths and an arrangement of photodetectors to measure (through transmission or reflection) small variations in light intensity due to volumetric oscillations in perfusion of that tissue [21], [22]. In practice, this method has been successfully adopted in hospitals and clinics via fingertip pulse oximeters to obtain blood oxygen saturation and heart rate [23], [24]. Nevertheless, because PPG and continuous BP morphologies are strongly correlated ($r > 0.9$) [25], and the cardiovascular system ultimately generates both waves, there is a relative consensus in the literature that the former can also carry information about the latter. Based on this principle, signal processing and machine-learning techniques have been proposed to estimate BP from PPG without calibration [26].

Although using a single PPG for assessing hypertension is promising, the relationship between PPG and BP is not entirely elucidated. In an attempt to indirectly solve this issue, several authors ([27], [28], [29], [30], [31], [32], [33], [34], [35], [36], [37], [38], [39]) reported accuracies in line with two guidelines, one by the American Association for the Advancement of Medical Instrumentation (AAMI) and the other published by the British Hypertension Society (BHS) [40], [41]. However, as observed by Schrumpf et al. [42], there is a lack of data regarding the distribution of the datasets to ensure (1) no mixing samples of the same subject and (2) equal data size per subject. Indeed, except for three studies ([43], [44], [45]), authors who partially or wholly follow this recommendation identify errors far above the reference values ([42], [46], [47], [48], [49], [50]), suggesting that the estimation problem is not yet resolved. Therefore, the present work investigates how intra- and inter-subject variabilities in BP lead to different results of machine-learning algorithms. By considering specifically the intra-subject scenario, we compare single PPG machine-learning algorithms with a regression using age, sex, weight, and subject index number as attributes and obtain similar results, suggesting that the algorithms might be learning to identify patients and not predict BP.

II. BACKGROUND AND STATE OF THE ART
Various PPG-based methods have been studied over the past few decades to monitor cuff-less BP with reliability and feasibility. In general, these methods fall into four approaches: pulse transit time (PTT), pulse arrival time (PAT), pulse wave velocity (PWV), and pulse wave analysis (PWA) [26], [51], [52], [53]. These approaches are summarized below.

PTT is the time a pressure wave takes to move from a proximal to a distal arterial site. This interval can be recorded using two PPG sensors in the ears, fingers, and toes [21]. Using empirical regression models, calculating logarithmic, linear, and quadratic relationships from PTT to BP is relatively straightforward but requires generating subject-specific calibration curves [54]; this demands periodic recalibrations, especially for chronic BP monitoring. Moreover, the sensors are sensitive to user motion; therefore, the captured signals must pass through a signal processing stage without losing their synchronization [26], [55].

PAT is the interval between a proximal electrocardiography (ECG) wave and the corresponding pressure wave in a distal arterial site. It can be measured using an ECG and PPG sensors in the upper or lower extremities, ears, or forehead [55], [56]. Similar to PTT, despite the simplicity of using regression models [54], BP functions of PAT require personal and periodic recalibrations and necessitate wearing two sensors very sensitive to movements to maintain synchronous signal processing and monitoring [26], [56].

PWV is the speed at which a pressure wave travels through blood vessels propelled by ventricular ejection [57]. It can be estimated using two PPG sensors positioned at different locations of the same arterial branch; values are calculated using the ratio of artery length between these sensors and PTT [51], [55]. The mapping from PWV to BP is subsequently given by a relationship that each maintains with arterial vessel

elasticity [58]. Even though it is a simple formula, because PWV depends on a reliable PTT, the former carries the same limitations as the latter. Therefore, calibration and frequent recalibrations are mandatory for the success and viability of the PWV method [26], [51], [55]. Furthermore, non-invasive estimation of artery pathways is difficult to obtain accurately other than by magnetic resonance imaging, an expensive and cumbersome technique [59], [60].

Finally, PWA consists of processing and inspecting a PPG wave to provide suitable features for creating models that relate this wave to other physiological signs. It shares with the methods mentioned above the characteristics of being prone to motion artifacts but has the advantage of requiring only one sensor [26]. Teng and Zhang [61] were among the first authors to propose a BP estimation based on a single PPG. Unlike PTT, PAT, and PWV, which are based on physical principles, the relationship between PPG and BP waves is not fully understood, despite both signals being similar and correlated [23], [25]. Therefore, PWA is a valuable technique used in classical and deep learning approaches for BP estimation, although not limited to this purpose.

Feature- and signal-based techniques have been techniques of PWA for reaching this goal. Feature-based strategies extract morphological and spectral characteristics from the PPG signal—and its first and second derivatives—to generate relevant information for a machine-learning algorithm. Several regression methods have been used: multilayer perceptron [27], multiple linear regression [28], [30], [31], [36], support vector machine [28], [29], [31], decision tree [28], [29], [30], [31], adaptive boosting [29], [30], random forest [29], [30], [47], Gaussian process [31], ridge regression [38], fully connected neural network [33], [43], convolutional neural network [38], long short-term memory [34], [36] and gated recurrent unit [34], [36]. Signal-based techniques explore deep learning models to extract features and perform the estimation. Several methods, alone and in combination, have also been used: fully connected neural network [49], [50], convolutional neural network [32], [37], [48], [50], long short-term memory [32], [37], [42], [48], [49], [50], gated recurrent unit [50], AlexNet [42], [44], ResNet [42], [46], [49], WaveNet [49], U-Net [35], residual U-Net [39] and generative adversarial network [48]. These contributions are summarized in Table 7, highlighting the dataset, data split, techniques and results for each one.

The present work proposes to investigate how a single PPG signal can be used to estimate BP with a representative set of machine-learning algorithms covering feature- and signal-based methods. We emphasized the effect of the data arrangement—by considering intra- and inter-subject variabilities—in two frequently used databases: Multiparameter Intelligent Monitoring in Intensive Care II and III (MIMIC-II, MIMIC-III). Regarding the intra-subject scenario, a regression using age, sex, weight, and subject index number—therefore, without a PPG signal (or its attributes) as input—achieves excellent performance, in line with AAMI and BHS standards. Finally, we discuss possible reasons for

it and provide guidelines regarding what we believe to be essential procedures to properly evaluate BP estimation from PPG. In summary, our main contributions are as follows:

1) We compare feature- and signal-based state-of-art techniques to estimate BP from a single PPG, using the same benchmark database.
2) We reveal the huge difference between results obtained by partitioning the dataset into intra- and inter-subject cross-validation schemes, showing that the former problem is practically resolved while the latter one is far from it.
3) We investigate an intra-subject scenario in which only non-PPG features are used and obtain excellent results, indicating that algorithms might be learning to identify patients and not predict BP.
4) We provide a guideline on how to perform a fair assessment to help future works to prevent overestimating performance.

III. METHODS

A. MIMIC-II UCI-ML AND MIMIC-III DATABASES

Initially released by the Research Resource for Complex Physiologic Signals (PhysioBank, PhysioToolkit, and PhysioNet) to stimulate studies of cardiovascular signals [62], the Multiparameter Intelligent Monitoring in Intensive Care (MIMIC) database, or simply MIMIC-I, is a collection of clinical records from 100 patients admitted to medical, surgical, and cardiac intensive care units (ICUs) at Boston's Beth Israel Hospital between 1994 and 1996. The first version consists of bedside monitor waveforms (ECG, PPG, and continuous BP), minute-by-minute hemodynamic trends (heart rate, SBP, DBP, respiratory rate, and oxygen saturation), and detailed clinical data from patient medical records and hospital medical information systems [63]. MIMIC-II is a collection of clinical records from virtually all adult patients admitted to ICUs at Boston's Beth Israel Deaconess Medical Center between 2001 and 2007—25,328 ICU stays from 22,870 hospital admissions. This second version encompasses identical hemodynamic waveforms and trends, laboratory results, and electronic clinical documentation [64], [65].

MIMIC-II represented a refinement of MIMIC-I, and MIMIC-III is better than MIMIC-II. This third version contains clinical records regarding admissions of 38,597 patients over 16 years of age to ICUs between 2001 and 2012 and 7870 neonates between 2001 and 2008 [66]. Although all records in MIMIC-III are deidentified according to Health Insurance Portability and Accountability Act standards, the MIMIC-III Matched Subset is a portion of the MIMIC-III in which patient information—age, sex, weight, and medical history—is associated with clinical records [67]; that is, each data segment received a subject unique identifier—an index number—from which that information is retrieved.

Finally, MIMIC-II UCI-ML is a clean and reduced version of bedside monitor waveforms present in the MIMIC-II, hosted in the University of California, Irvine (UCI)

Machine-Learning (ML) Repository. Using this dataset, Kachuee et al. provide ECG, PPG, and continuous BP signals already processed and validated to support and stimulate works designing cuff-less BP estimation algorithms [68], [69]. Those signals are available in four MATLAB® files, consisting of cell arrays and matrices (which facilitate their use), especially if compared to the original file structure of the MIMIC. However, MIMIC-II UCI-ML has a substantial limitation in suppressing the patient index number related to each record segment.

The MIMIC-III Matched Subset database is the benchmark through which our study is founded (henceforth, we will refer to it solely as MIMIC-III). Moreover, although MIMIC-II UCI-ML database does not report patient index numbers corresponding to each data segment, we adopt it for some analysis, given its widespread use by researchers. In both datasets, PPG and continuous BP—also labeled as arterial blood pressure (ABP)—are sampled at 125 Hz.

B. PRE-PROCESSING

Because a large volume of signals from raw MIMIC-II is distorted and corrupted, data from MIMIC-II UCI-ML are already picked, cleaned, and organized according to pre-processing steps performed by Kachuee et al. (2015): (1) the selection of files with ECG, PPG, and ABP waveforms; (2) the average filtering to smooth these signals; (3) the removal of data blocks with unacceptable BP and heart rate values and with persistent discontinuities even after the smoothing procedure; and (4) the computation of PPG autocorrelation function to identify the degree of similarity between consecutive pulses and subsequent removal segments with high degrees of alteration [68]. As a result of these steps, data blocks containing simultaneous ECG, PPG, and ABP are available [69]. Considering this dataset, we performed no further cleaning processes; we performed selection to handle only PPG and ABP signals. PPG segments were then passed through a fourth-order Chebyshev II band-pass filter from 0.5 Hz to 10 Hz, analogous to the optimal filter proposed by Liang et al. to eliminate the offset and the high-frequency noise and highlight the dicrotic notch and the systolic and diastolic phases [70]; ABP segments were preserved in their original form.

MIMIC-III is a substantial database that requires picking and cleaning procedures before being used. We followed some of the criteria suggested by Slapničar et al. [46]: (1) the selection of files containing PPG and ABP waveforms, specified as "PLETH" and "ABP" in the database; (2) the recognition of snippets with missing signals, represented as "Not a Number" entries; and (3) the identification of flat lines, which are any interval exceeding ten samples with equal values. Similar to Sun et al. [71], we set 300 mmHg and 20 mmHg as the upper and lower bounds (respectively) on the physiologic ranges of ABP and set 20 mmHg and −20 mmHg as the maximum and minimum limits (respectively) for variations between two consecutive points. Subsequently, data fragments and points from PPG and ABP were replaced by "Not a Number" entries if classified into some abnormality criteria.

All entries indicate a set of segmentation positions from which valid signals start and end; clean data are extracted at numerical intervals for PPG and ABP waveforms. From each subject, we selected the first pair of clean segments longer than 30 minutes to continue pre-processing. Inspecting PPG signals, we observed that their values are already normalized and contained in ranges between 0 and 1 [a.u.] and from 0 to 4 [a.u.]. Signals of the former type are more prevalent, and we chose them for our analysis to maintain attributes as regular as possible in the feature extraction procedure. Much like MIMIC-II UCI-ML segments, MIMIC-III PPG segments are passed through the fourth-order Chebyshev II band-pass filter, while the ABP segments are preserved. Finally, despite the large volume of the MIMIC-III, after cleaning and filtering processes, we retained a 30-minute block of simultaneous PPG and ABP for each patient with age, sex, and weight information: 633 ICU patients satisfied these criteria, totaling 316.5 hours of data. We limited the amount to 30 minutes per subject to reduce bias.

C. FEATURE EXTRACTION

Each data block was subdivided into eight-second non-overlapping windows for both pre-processed datasets. ABP, SBP, and DBP labels were obtainedred, by considering the averages of peaks and valleys along the sections. These extreme points were recognized using an algorithm adapted from Hsu et al. [33], in which systolic peaks are marked first, and diastolic valleys are detected by locating the minimum point between two consecutive peaks. From PPG windows, signal-based algorithms explore all points, whereas feature-based algorithms probe several morphological and spectral characteristics recommended in the literature.

The PPG signal is commonly formed by a sequence of pulse waves with a specific shape: a rising edge as the anacrotic phase (primarily related to systole), a falling edge as the catacrotic phase (associated with diastole and wave reflection), and a point of inflection as a dicrotic notch separating these two phases [23]. Considering this morphology, PPG features can be extracted from the original pulse and its first and second derivatives (Figure 1). Examples include systolic amplitude, pulse width, pulse area, peak-to-peak interval, pulse interval, augmentation index, and many others. These attributes fall into indices representing slopes, areas, ratios of areas, intensities, ratios of intensities, differences of intensities, periods, and ratios of periods. Lin et al., Chowdhury et al. and El-Hajj et al. provide detailed descriptions of how to extract them [31], [36], [72].

Before the calculation of these features, all PPG pulse windows and their characteristic points were identified using a procedure similar to one adopted by Hsu et al. [33]. From each pulse, every morphological attribute reported by Lin et al., Chowdhury et al. and El-Hajj et al. was then
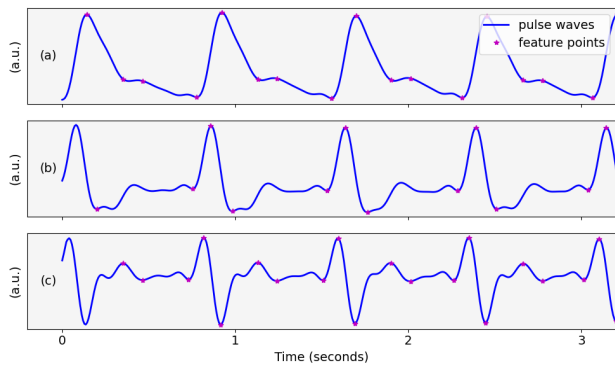
FIGURE 1. A sequence of PPG pulse waves (a) along with its first (b) and second (c) derivatives, including feature points such as peaks, valleys and dicrotic notches.
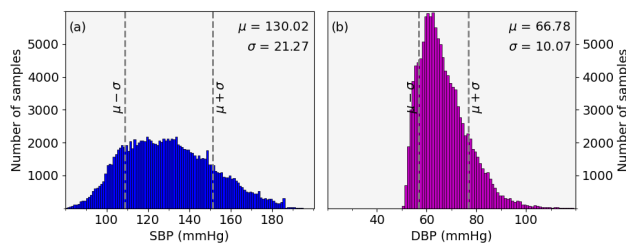

FIGURE 2. Resulting distribution of SBP and DBP samples, (a) and (b), in the MIMIC-II UCI-ML database after the corresponding PPG goes through pre-processing and feature extraction steps. DBP: diastolic blood pressure; PPG, photoplethysmography; SBP: systolic blood pressure; μ: mean; σ: standard deviation.
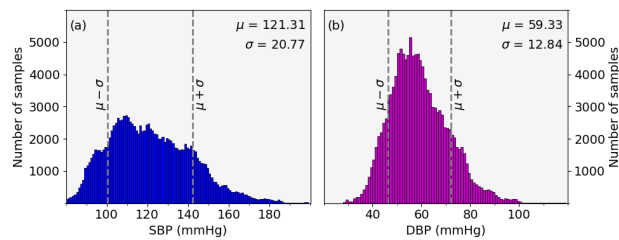

FIGURE 3. Resulting distribution of SBP and DBP samples, (a) and (b), in the MIMIC-III database after the corresponding PPG goes through pre-processing and feature extraction steps. DBP: diastolic blood pressure; PPG, photoplethysmography; SBP: systolic blood pressure; μ: mean; σ: standard deviation.

extracted. Despite these characteristics being computed from every pulse, each feature vector was arranged with the averages of the pulse attributes. In this process, whenever the calculation of the features returns inconsistent values—in general because one or more pulses are distorted—all attributes were intentionally invalidated, and the entire section was discarded. Following this criterion, about ten percent of windows are excluded from the subsequent analysis. These windows were excluded from the signal-based analysis to maintain similar conditions during comparison. Figures 2 and 3 exhibit the distributions of SBP and DBP samples for only valid windows in the MIMIC-II UCI-ML and MIMIC-III databases, respectively, indicating an acceptable BP range.

Analogous to Xin and Sun and Chowdhury et al. [31], [43], spectral features were considered for analysis and computed using the fast Fourier transform. Taking the maximum amplitude between 0.5 Hz and 3.0 Hz as a fundamental frequency, the amplitudes of the first five harmonics were extracted from the signal of each remaining window—and its first and second derivatives—and included in the feature vector. Finally, age, sex, and weight information were added.

A specific baseline analysis (as explained below) uses a feature vector consisting only of age, sex, weight, and patient index number (subject unique identifier of the MIMIC-III Matched Subset) as attributes, i.e., not including any morphological or spectral features of PPG.

D. MACHINE LEARNING ALGORITHMS
Feasible machine-learning algorithms to perform BP estimation, considering a feature-based approach, are XGBoost, LightGBM, and CatBoost. All are based on the iterative gradient descent algorithm for tree boosting, also known as gradient boosting decision tree (GBDT), initially established by Friedman [73]. The XGBoost is a scalable version of the GBDT that applies a functional space optimization adapted for sparse data and approximate tree learning using a weighted quantile sketch [74]. The LightGBM is a GBDT implementation that improves the efficiency and scalability, especially for large data sizes and high feature dimensions, by proposing gradient-based one-side sampling to exclude occurrences of data with small gradients and an exclusive feature bundling to reduce feature number [75]. Finally, the CatBoost is also a GBDT implementation that presents an ordered boosting and processing of categorical features to address the problem of prediction shift in gradient boosting [76].

Convenient algorithms that accomplish the same task in a signal-based approach using deep learning models include ResNet18, ResNet-LSTM, and residual U-Net. The residual network (ResNet) is a popular deep learning model in which shortcut connections between layers (i.e., skipping one or more layers performing identity mapping) are implemented to overcome the gradient degradation problem, an unexpected increase in training error as the network becomes deeper. In this sense, ResNet18 is merely an 18-layer ResNet [77]. ResNet-LSTM combines a ResNet with a long short-term memory (LSTM) architecture—a learning model that stores information through recurrent back-propagation. Although recurrent neural networks may cause the gradient to blow up or vanish, LSTM partially solves the latter problem by using a self-connected unit to enforce a constant error backflow, which allows long-term features to be learned [78]. Finally, U-Net is a neural network model initially designed to perform image segmentation, consisting of a contracting path (encoder) similar to a common convolution network and an expansion path (decoder) in which features from the first path are concatenated with the second one—resulting in an almost symmetrical U-like shape [79], [80]. Inspired by ResNet,

residual U-Net only includes shortcut connections to facilitate information propagation between layers [81]. To calculate the BP using the residual U-Net, after obtaining the reconstructed ABP signal output [82], we calculated the mean value of the peaks and valleys to represent SBP and DBP, respectively.

It is possible to use machine-learning algorithms other than the ones previously presented. However, because we aimed to investigate intra- and inter-subject variabilities, we selected a set of algorithms in line with the state-of-the-art instead of exhausting them all.

E. EXPERIMENTAL DESIGN

Two partitioning strategies were adopted to denote BP variability explicitly throughout the MIMIC-III data. The first consists of a ten-fold cross-validation in which signal windows were randomly selected to compose each folder. In this case, there was no concern about where samples of the same patient index number were allocated; they were distributed in all folders with equal chance. We called this arrangement an intra-subject scenario because samples of the same patient can appear in the training and testing phases, causing data leakage. The second consists of a ten-fold cross-validation in which the patient now groups samples to compose each folder. In this case, there was concern about not allowing samples of the same patient to be located in more than one folder to guarantee a testing phase without data leakage—simulating a more realistic and challenging application. We called this arrangement an inter-subject scenario because only cross-subject BP variability was available to those algorithms during the testing phase.

Unfortunately, regarding MIMIC-II UCI-ML, the subject indices were not reported, impacting our original two partitioning strategies. Therefore, an intra-subject scenario was implemented using a ten-fold cross-validation; however, there was no way to limit the data per subject, which probably led to bias. Moreover, a ten-fold inter-subject scenario is completely infeasible without these indices. To overcome this difficulty, we adopted a four-fold cross-validation in which the four MATLAB® files of the database represent each folder [83], expecting that there would be no mix of patients in these files. Even if there were, such an arrangement would at least reduce data leakage to a minimum.

Considering these cross-validation schemes, two machine-learning experiments were performed. The first design investigated a PPG feature-based approach with age, sex, and weight as attributes, in which XGBoost, LightGBM, and CatBoost were used to perform BP estimation. The second design explored a PPG signal-based approach in which ResNet18, ResNet-LSTM, and residual U-Net were tested. These machine-learning experiments were implemented using Python libraries (xgboost 1.5.2, CatBoost 1.0.5, and scikit-learn 0.24.1), and deep learning models were implemented using a TensorFlow library (tensorflow-gpu 2.3.0). The primary settings of these algorithms are summarized in Table 1 and Figure 4. To select parameters

for GBDT models, we performed brute-force searches with 90% subsets of three random training folds and validated with the remaining 10% in each scenario (intra- and inter-subject variabilities) and each database (MIMIC-II UCI-ML and MIMIC-III). The values are displayed in Table 1.

Considering the intra-subject scenario of the MIMIC-III exclusively, one final design examines a feature vector with age, sex, weight, and patient index number as attributes (therefore, without PPG samples) feeding into an XGBoost-based estimator to provide a baseline for comparison with the preceding scenarios.

F. METRICS

The AAMI and BHS standards are significant guidelines for evaluating a device while taking BP measurements. The accuracy criteria of the former states that SBP and DBP must have a mean difference (MD) between reference and test measures of ± 5 mmHg with a standard deviation (STD) less than or equal to 8 mmHg in a study population of at least 85 subjects [41]. By contrast, the accuracy criterion of the latter states that SBP and DBP must have cumulative percentages of the absolute differences (between reference and test measures) within ≤ 5 mmHg, ≤ 10 mmHgred, and ≤ 15 mmHg intervals, in accordance with the values in Table 2. If all three cumulative percentages are equal or greater than the tabulated values in each row, the test device receives grades A, B, C, or D, in case it is worse than C [40].

We also present our results in terms of mean absolute error (MAE), calculated as follows:

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|, \qquad (1)$$

where y_i and $\hat{y}_i$ are the target and predicted values, respectively.

In addition to these metrics, essential assessment tools include the scattering of the target versus predictions, Bland-Altman plots, and error histograms. The scattering of the target versus prediction is the most common visual resource for estimation problems due to its relationship with the correlation coefficient displayed next to the chart. Nevertheless, because a high correlation coefficient does not necessarily denote agreement, the Bland-Altman plot exhibits the 95% limits of agreement between two measurements, highlighting their differences relative to their means. In this plot, we expect the mean and standard deviation of the differences to be constant over the entire range and such differences to be normally distributed. The error histogram is essential to verify the latter condition because a skewed histogram leads the Bland-Altman plot to be misinterpreted [84]. We explore all three graphics in our analysis.

IV. RESULTS AND DISCUSSIONS

Table 3 summarizes the performance of the BP algorithms in terms of MD, STD, and MAE in the MIMIC-III dataset considering the intra- and inter-subject data arrangements.

TABLE 1. Parameter values of the boosting algorithms.

| Dataset	Algorithm	Intra-subject parameters						
		max_depth	learning_rate	n_estimators	colsample_bylevel	l2_leaf_reg	min_data_in_leaf	num_leaves
	XGBoost	10	0.06	500	0.7	-	-	-
	LightGBM	10	0.10	-	-	-	512	100
	CatBoost	10	0.10	500	-	5	-	-
MIMIC-III	Algorithm	Inter-subject parameters						
		max_depth	learning_rate	n_estimators	colsample_bylevel	l2_leaf_reg	min_data_in_leaf	num_leaves
	XGboost	4	0.03	100	0.5	-	-	-
	LightGBM	4	0.03	-	-	-	32	100
	CatBoost	7	0.03	100	-	9	-	-
	Algorithm	Intra-subject parameters						
		max_depth	learning_rate	n_estimators	colsample_bylevel	l2_leaf_reg	min_data_in_leaf	num_leaves
	XGBoost	10	0.06	500	0.5	-	-	-
	LightGBM	10	0.10	-	-	-	512	100
	CatBoost	10	0.10	500	-	5	-	-
MIMIC-II UC-ML	Algorithm	Inter-subject parameters						
		max_depth	learning_rate	n_estimators	colsample_bylevel	l2_leaf_reg	min_data_in_leaf	num_leaves
	XGBoost	4	0.03	100	0.9	-	-	-
	LightGBM	4	0.03	-	-	-	32	300
	CatBoost	4	0.03	100	-	9	-	-
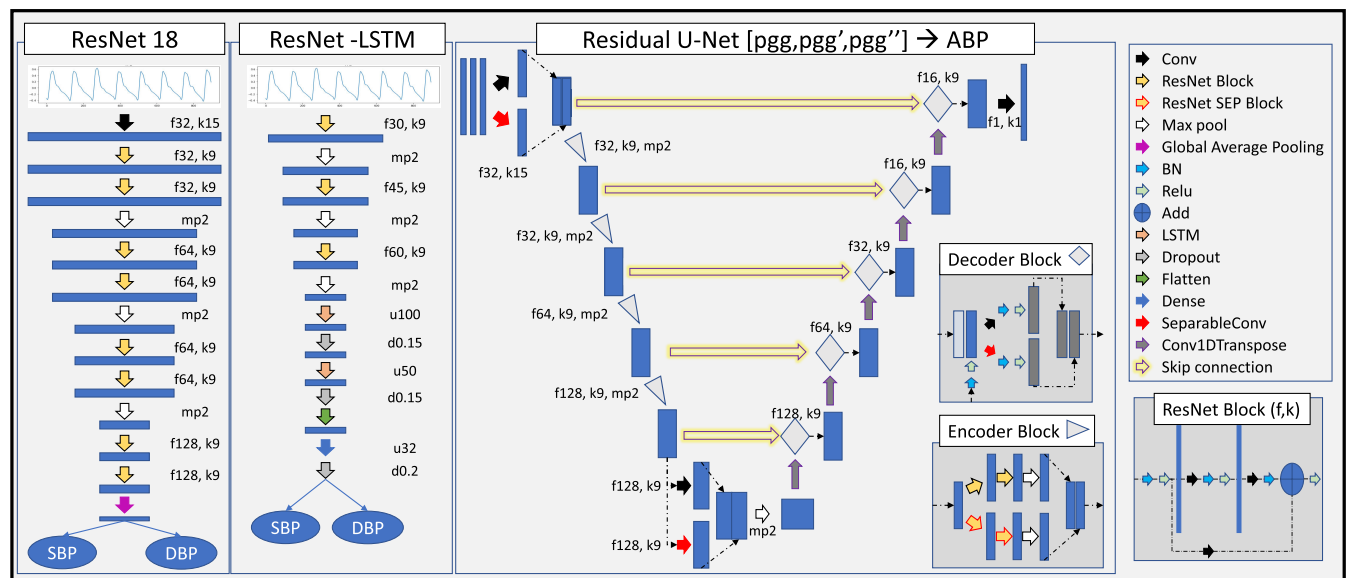
FIGURE 4. Deep learning architectures for BP estimation. In particular, to calculate the BP using the residual U-Net, we get the ABP signal reconstructed by the model, and calculate the mean value of the peaks and valleys in order to represent SBP and DBP, respectively. ABP: arterial blood pressure; DBP: diastolic blood pressure; SBP: systolic blood pressure; Conv: convolutional 1-dimension operation; SeparableConv: depth wise separable 1-dimension convolution; LSTM: long short-term memory; ResNet: residual network; ResNet SEP Block: ResNet Block but changing the convolutional 1-dimension operation by a depth wise separable 1-dimension convolution; BN: batch normalization operation; Relu: rectified linear unit activation function; f: number of output filters in the convolutions; k: length of the 1-dimension convolution window; mp: size of the max pooling window; u: number of unities in LSTM layer; d: fraction of the input units to dropout.

TABLE 2. Grading criteria for BHS [40].

Grade	Cumulative error (%)		
	≤ 5 mmHg	≤ 10 mmHg	≤ 15 mmHg
A	60.00	85.00	95.00
B	50.00	75.00	90.00
C	40.00	65.00	85.00
D		Worse than C	

BHS: British Hypertension Society.

In the intra-subject scenario, SBP and DBP values estimated with XGBoost, LightGBM, CatBoost, residual U-Net, and

ResNet18 met the AAMI standard, suggesting that both feature- and signal-based methods are equally capable of solving the problem. ResNet-LSTM did not reach the metrics for SBP and, for DBP, attains it by a narrow margin, suggesting that intra-subject variability can be learned only by specific techniques. XGBoost achieved an MD and an STD of 0.04 ± 4.31 for SBP and 0.01 ± 2.28 for DBP. Residual U-Net achieved 0.36 ± 4.64 and -0.08 ± 2.49, respectively. Both results are comparable to the state-of-the-art methods that do not explicitly inform a separation of individuals during cross-validation. Nevertheless, the same algorithms show

higher error values in estimating BP in the latter scenario. Indeed, because they exhibit an STD close to 20.77 mmHg and 12.84 mmHg for SBP and DBP, respectively, these methods seemingly pursue no other tendency than the standard deviations relative to SBP and DBP averages of the dataset itself, as shown in Figure 3, revealing their complete inability to extract information from the inter-subject variability.

Table 4 exhibits the performance of those algorithms according to the BHS standard. The values met the same general conclusion. In the case of intra-subject variability, XGBoost, LightGBM, CatBoost, and residual U-Net are classified as Grade A for SBP and DBP—a grading similar to the state-of-the-art. ResNet18 is classified as Grade B for SBP and Grade A for DBP, whereas ResNet-LSTM is classified as Grades D and B. Once more, XGBoost presented the best indices, with 84.93%, 96.70% and 98.86% of the differences between actual and expected values for SBP within the ≤5 mmHg, ≤10 mmHg and ≤15 mmHg criteria, respectively, and 96.92%, 99.37% and 99.76%, respectively for DBP. Residual U-Net showed 85.59%, 96.66% and 98.58% for SBP and 97.22%, 99.25% and 99.65% for DBP. However, in the case of inter-subject variability, all methods notably fell into Grade D for SBP and DBP. XGBoost, for example, despite being one of the best in the preceding scenario, reached low cumulative percentages (19.39%, 37.75%, and 53.93% for SBP and 37.63%, 69.25%, and 86.52% for DBP), suggesting that these algorithms are incapable of solving the problem in the most general case: when the samples of each subject are allocated in the training or testing set (but not in both) and when these samples are arranged to be the same size.

Figures 5 and 6 display the scattering of the target versus prediction (items (a) and (b)), the Bland-Altman plot (items (c) and (d)), and the histogram of the errors (items (e) and (f)) for SBP and DBP estimation using XGBoost. These findings help to understand how machine-learning algorithms behave. On the one hand, Figure 5 depicts the intra-subject scenario, in which the target and prediction have correlation coefficients (ρ) of 0.98 for SBP and DBP, respectively. The distribution of samples within the 95% limits of agreement in the Bland-Altman plot is flattened along the entire range, and the histogram of the errors is normally distributed around zero. On the other hand, Figure 6 depicts the inter-subject scenario, in which the correlation coefficients drop to 0.24 for SBP and 0.43 for DBP. The sample distribution is biased when moving away from the midpoint of averages, and the histogram of the errors is more widespread and no longer zero-centered.

Figures 7 and 8 for residual U-Net have similar descriptions. In conclusion, the contrast between the two scenarios is evident and reinforces the point that BP estimation from PPG alone is not resolved in the most challenging case.

Table 5 and Table 6 exhibit, for the MIMIC-II UCI-ML database, the performance of the BP algorithms concerning AAMI and BHS standards, respectively. Although the general appointments are essentially the same, regarding intra-subject
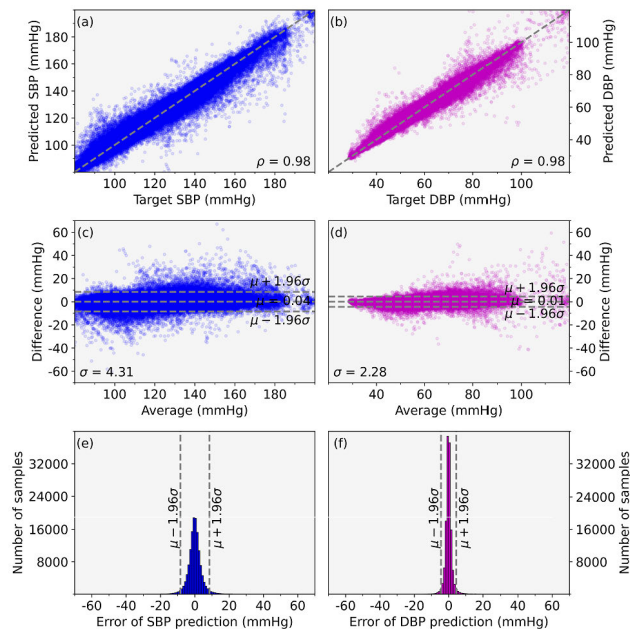

FIGURE 5. Scattering of the target versus prediction, (a) and (b), Bland-Altman plot, (c) and (d), and histogram of the errors, (e) and (f), for SBP and DBP estimation using XGBoost in the MIMIC-III database, by considering the intra-subject variability. DBP: diastolic blood pressure; SBP: systolic blood pressure; μ: mean; σ: standard deviation; ρ: correlation coefficient.

FIGURE 6. Scattering of the target versus prediction, (a) and (b), Bland-Altman plot, (c) and (d), and histogram of the errors, (e) and (f), for SBP and DBP estimation using XGBoost in the MIMIC-III database, by considering the inter-subject variability. DBP: diastolic blood pressure; SBP: systolic blood pressure; μ: mean; σ: standard deviation; ρ: correlation coefficient.
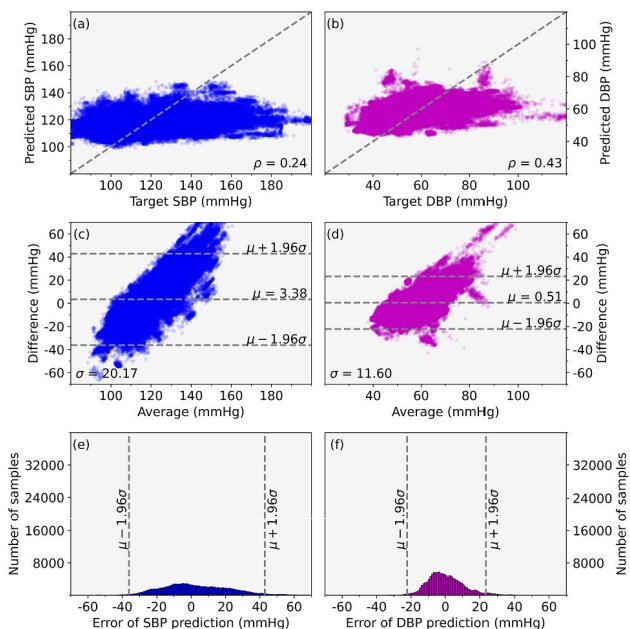
variability, only XGBoost and residual U-Net attain the STD criterion and Grade A for SBP and DBP. LightGBM, CatBoost, and ResNet18 exceed the reference limit for SBP,

TABLE 3. Evaluation of BP estimation in the MIMIC-III dataset with regard to AAMI standard and MAE.

Algorithm		Intra-subject metrics (mmHg)			Inter-subject metrics (mmHg)		
		MD	STD	MAE	MD	STD	MAE
XGBoost	SBP	0.04	4.31	2.85	3.38	20.77	16.18
	DBP	0.01	2.28	1.32	0.51	11.60	8.45
LightGBM	SBP	0.00	6.33	4.43	0.02	20.23	16.28
	DBP	0.00	3.30	2.10	−0.03	11.79	8.59
CatBoost	SBP	0.01	4.92	3.40	0.06	20.16	16.32
	DBP	0.01	2.71	1.73	0.07	11.68	8.60
Residual U-Net	SBP	0.36	4.64	2.87	0.00	23.32	18.60
	DBP	−0.08	2.49	1.31	−0.82	14.35	10.94
ResNet18	SBP	−0.06	7.12	5.23	0.91	22.74	18.16
	DBP	−0.03	4.73	3.40	0.67	13.70	10.38
ResNet-LSTM	SBP	0.69	11.54	8.80	1.08	21.86	17.56
	DBP	0.42	7.73	5.72	1.64	13.53	10.47

AAMI: Association for the Advancement of Medical Instrumentation; BP: blood pressure; DBP: diastolic blood pressure; MAE: mean absolute error MD: mean difference; SBP: systolic blood pressure, STD: standard deviation.

TABLE 4. Evaluation of BP estimation in the MIMIC-III dataset with regard to BHS standard.

Algorithm		Intra-subject cumulative error (%)			Inter-subject cumulative error (%)		
		≤ 5 mmHg	≤ 10 mmHg	≤ 15 mmHg	≤ 5 mmHg	≤ 10 mmHg	≤ 15 mmHg
XGBoost	SBP	84.93	96.70	98.86	19.39	37.75	53.93
	DBP	96.92	99.37	99.76	37.63	69.25	86.52
LightGBM	SBP	68.67	90.63	96.66	17.89	35.50	52.79
	DBP	91.70	98.46	99.49	37.44	67.80	85.96
CatBoost	SBP	78.90	95.38	98.50	17.75	35.54	52.25
	DBP	94.93	99.12	99.69	36.37	67.34	85.29
Residual U-Net	SBP	85.59	96.66	98.58	17.24	33.14	47.62
	DBP	97.22	99.25	99.65	29.83	54.87	73.94
ResNet18	SBP	59.60	87.26	95.73	17.29	33.54	48.74
	DBP	78.32	96.06	98.98	31.76	57.69	76.23
ResNet-LSTM	SBP	37.85	65.80	82.69	18.11	34.93	49.73
	DBP	54.68	84.41	95.09	29.50	56.03	76.53

BHS: British Hypertension Society; BP: blood pressure; DBP: diastolic blood pressure; SBP: systolic blood pressure.

satisfy it for DBP, and reach no more than Grade B for SBP while achieving Grade A for DBP. Finally, ResNet-LSTM achieves reasonable performance only for DBP. Nevertheless, regarding inter-subject variability, similar to MIMIC-III, all methods decisively fail because they are skewed toward the standard deviations of SBP and DBP averages of the dataset—21.27 and 10.07 (Figure 2)—and are at most ranked as Grade C, once again revealing their inability to solve the general problem.

Comparing our results with the literature (Table 7), we inferred that Wang et al., Khalid et al., Mousavi et al., Chowdhury et al., Panwar et al., Hasanzadeh et al., Hsu et al., El-Hajj and Kyriacou, Athaya and Choi, Rong and Li, Wang et al. and Kim et al. ([27], [28], [29], [30], [31], [32], [33], [34], [35], [36], [37], [38], [39]) probably yielded an overlap between training and testing subjects because their results are similar to our intra-subject scenario. Hasanzadeh et al. even tried to prevent it by retaining the original order of the dataset during ten-fold cross-validation. However, there is no guarantee that such a strategy has worked (although it has possibly reduced it) because the patient index numbers are not reported in their datasets. Slapničar et al., Xing et al., Schrumpf et al., Brophy et al., Paviglianiti et al. and Leitner et al. ([42], [46], [47], [48], [49], [50]) explicitly ensured training, validation, and testing

sets without mixing subjects in an attempt at a calibration-free approach. Their results are similar to our inter-subject scenario, in which the performances are far from assured. Ultimately, only Xing and Mingshan, Schlesinger et al. and Mahmud et al. ([43], [44], [45]) did not mix individuals, and they achieved impressive results; nevertheless, we failed to reproduce them.

Finally, Figure 9 exhibits the scattering of the target versus prediction, the Bland-Altman plot, and the histogram of the errors for a BP estimation using XGBoost (in this case, trained and tested with age, sex, weight, and patient index number as attributes instead of PPG features). Curiously, the target and prediction have a high correlation coefficient, the distribution of samples within the Bland-Altman limits is flat along the entire range, and the error histogram has a narrow and normal shape around zero. This result is very similar to the one obtained with PPG feature-based XGBoost. Concerning BHS and AAMI, this approach achieves an MD and an STD of 0.00 ± 5.67 for SBP, 0.00 ± 2.76 for DBP, and Grade A for both (74.03%, 92.85% and 97.54%; 94.52%, 99.16% and 99.69%; respectively). Such a finding is not surprising and occurs for a simple reason: although BP is the training label throughout the learning process, it is a feature for recognizing individuals. Consequently, any powerful technique can learn the average BP values per subject and, only with

TABLE 5. Evaluation of BP estimation in the MIMIC-II UCI-ML dataset with regard to AAMI standard and MAE.

Algorithm		Intra-subject metrics (mmHg)			Inter-subject metrics (mmHg)		
		MD	STD	MAE	MD	STD	MAE
XGBoost	SBP	0.07	7.48	4.86	4.09	20.49	16.67
	DBP	0.02	3.89	2.29	0.87	9.60	7.14
LightGBM	SBP	0.01	11.00	8.00	0.26	20.45	16.49
	DBP	0.00	5.65	3.77	0.10	9.59	7.21
CatBoost	SBP	0.01	8.58	5.94	0.42	20.29	16.43
	DBP	0.01	4.59	2.94	0.12	9.77	7.38
Residual U-Net	SBP	0.41	5.22	3.17	−0.65	24.08	19.10
	DBP	0.07	2.89	1.54	−0.20	11.09	8.18
ResNet18	SBP	0.30	9.94	7.23	2.04	24.72	19.85
	DBP	−0.05	5.75	4.05	−0.34	11.25	8.42
ResNet-LSTM	SBP	0.75	14.76	11.19	0.15	22.30	17.71
	DBP	0.02	7.94	5.74	−0.76	10.26	7.96

AAMI: Association for the Advancement of Medical Instrumentation; BP: blood pressure; DBP: diastolic blood pressure; MAE: mean absolute error MD: mean difference; SBP: systolic blood pressure, STD: standard deviation.

TABLE 6. Evaluation of BP estimation in the MIMIC-II UCI-ML dataset with regard to BHS standard.

Algorithm		Intra-subject cumulative error (%)			Inter-subject cumulative error (%)		
		≤ 5 mmHg	≤ 10 mmHg	≤ 15 mmHg	≤ 5 mmHg	≤ 10 mmHg	≤ 15 mmHg
XGBoost	SBP	68.17	87.59	94.29	17.97	36.55	53.24
	DBP	89.37	97.65	99.17	43.73	76.72	92.23
LightGBM	SBP	44.64	71.61	85.29	18.50	36.11	52.00
	DBP	75.16	94.15	98.01	42.42	75.81	92.60
CatBoost	SBP	58.14	82.60	92.04	18.42	36.04	51.74
	DBP	83.72	96.46	98.85	40.80	74.66	92.38
Residual U-Net	SBP	82.88	95.21	97.87	17.48	32.84	46.93
	DBP	95.73	98.93	99.57	40.84	70.17	86.65
ResNet18	SBP	47.34	75.63	88.76	16.57	31.49	44.90
	DBP	71.26	93.24	98.15	38.91	68.56	85.38
ResNet-LSTM	SBP	31.33	56.12	73.05	17.52	34.70	50.60
	DBP	55.07	85.17	95.16	37.85	69.79	89.11

BHS: British Hypertension Society; BP: blood pressure; DBP: diastolic blood pressure; SBP: systolic blood pressure.

this information, achieve a reasonable or even excellent performance, which makes ambiguous how much of the PPG-based estimation is owing to hemodynamic outcomes [85].

Although the inter-subject configuration performs worse than the intra-subject one, and the former represents a more realistic and challenging application, several works have suggested that a feasible solution is to include a calibration step—a manner to provide intra-subject information to the machine learning algorithm. After all, everyone is different, which makes inferring an individual's vital signs from other individual's training sets an extremely difficult task. On the other hand, techniques for predicting BP from PPG is very convenient, especially for continuous long-term BP monitoring out of hospital, even if the training and testing data are from the same individual. Nevertheless, a dedicated algorithm, trained on a person's historical BP and PPG data, has practical value.

Because of the previous analyses, we suggest the following guidelines regarding what we believe to be the best practices for a fair assessment of a BP estimation from PPG only:

1) Do not mix subjects during the k-fold cross-validation to avoid data leakage and deliver new samples during the test.
2) Select an equal data size per subject rather than different proportions to avoid bias.

3) Do not use databases not reporting the subject index number to meet previous recommendations.
4) In addition to the AAMI and BHS standards, provide plots for the scattering of the target versus prediction, the Pearson correlation coefficient, the Bland-Altman plot, and the error histogram to complement the analysis and provide a unified point of view.
5) Finally, compare the results with the average BP values per subject to confirm that the proposal is better than directly selecting these values.

We are convinced that any algorithms or methods genuinely concerned with the problem of estimating BP from a single PPG will be considered only if they achieve their results in line with all preceding recommendations. However, there is a balance between convenience and precision of a new technology, i.e., the AAMI and BHS standards are rules for medical devices and possibly too strict for wearable devices that seek to predict BP based on PPG. Wearable applications are developed and positioned to be used anywhere, anytime for home health care (rather than medical or hospital use), where even a less precise estimation may help to avoid acute cardiovascular events.

Finally, our work has limitations to some extent. First, MIMIC-II and MIMIC-III are collections of records from patients admitted to ICUs, with the most varied clinical

TABLE 7. Summary of papers in literature that proposed BP estimation methods.

Article	Data and Split	Citation	ML Techniques	Result (mmHg)
A Novel Neural Network Model for Blood Pressure Estimation Using Photoplethesmography without Electrocardiogram [27]	MIMIC II - UCI It probably mixes patients.	*"In total, there are 58,795 valid intervals of PPG signal (subject number is 72) and corresponding BP values for different people and different time instances. In order to avoid overfitting, we use 70% of them for network training, 15% of them for validation, and 15% of them for testing."*	**a) Artificial neural networks (ANN)** b) Linear regression c) Regression support vector machine (R-SVM)	(ME ± std) SBP: −0.0217±4.8950 DBP: 0.0975 ± 2.9160
Blood Pressure Estimation Using Photoplethysmography Only Comparison between Different Machine Learning Approaches [28]	Queensland Vital Signs It probably mixes patients.	*"In total, 8131 × 3 good quality PPG signal features and reference BPs were used to train and test the above three machine learning algorithms with 10-fold cross-validation. In each iteration, 9 folds were used to train an algorithm, and the remaining fold was used to test that algorithm. The process continued until 10 iterations were completed."*	**a) Regression tree** b) Multiple Linear Regression (MLR) c) Linear SVM	(ME ± std) SBP: −0.1 ± 6.5 DBP: −0.6 ± 5.2
Blood pressure estimation from appropriate and inappropriate PPG signals using A whole-based method [29]	MIMIC II - UCI It probably mixes patients.	*"training and testing data with the 10 fold cross validation algorithm."*	**a) Adaptive Boosting Regression (Adaboost)** b) Decision tree regression c) Support Vector Regression d) Random Forest Regression	(ME ± std) SBP: 0.187 ± 4.173 DBP: −0.05 ± 8.901
Blood pressure estimation using photoplethysmogram signal and its morphological features [30]	MIMIC II - UCI It probably mixes patients.	*"We used 10-fold cross validation method to divide the data into training and testing sets. It is worth mentioning that, in the UCI dataset, the PPG recordings belonging to each subject are placed consecutively in the dataset but do not have a common identification number. As a result, in order to prevent the overlapping of training and testing subjects, no shuffling was applied and the order of samples in the dataset was retained"*	**a) Adaboost** b) Linear regression c) Decision tree d) Random forest	(ME ± std) SBP: 0.09 ± 10.38 DBP: 0.23 ± 4.22
Estimating blood pressure from the photoplethysmogram signal and demographic features using machine learning techniques [31]	Liang et al. [70] It probably mixes patients.	*"19 algorithms were trained using 10-fold cross validation"*	a) Gaussian Process Regression b) Ensemble trees	(MAE ± std) SBP: 3.02 ± 9.29 DBP: 1.74 ± 5.54
PP-Net: A Deep Learning Framework for PPG-Based Blood Pressure and Heart Rate Estimation [32]	MIMIC II - UCI It probably mixes patients.	*"These pre-processing steps result in reduction of total subjects from 12000 to 1557. Now, final data of approximately 1557 subjects were included for evaluation"*	Convolutional neural network (CNN) + long short-term memory (LSTM)	(MAE ± std) SBP: 3.97 ± 0.064 DBP: 2.3 ± 0.196 (ME ±std) SBP: 1.55 ± 5.41 DBP: −1.25 ± 5.65
Generalized Deep Neural Network Model for Cuffless Blood Pressure Estimation with Photoplethysmogram Signal Only [33]	MIMIC II - UCI It probably mixes patients.	*"The selected feature set (η 32 × 2,176,188) is split into three parts, and each part contains 70%, 20% and 10% of the data, which serve as training, testing and validation datasets, respectively."*	Fully Connected Neural Network	(MAE ± std) SBP: 3.21 ± 3.35 DBP: 2.23 ± 2.44
Deep learning models for cuffless blood pressure monitoring from PPG signals using attention mechanism [34]	MIMIC II It probably mixes patients.	*"Both datasets were partitioned into 60 % train, 20 % validation and 20 % test set"*	**a) Bidirectional Gated Recurrent Units (Bi-GRU) + attention** b) Bidirectional LSTM c) LSTM	(MAE ± std) SBP: 2.58 ± 3.35 DBP: 1.26 ± 1.63
An estimation method of continuous non-invasive arterial blood pressure waveform using photoplethysmography: A u-net architecture-based approach [35]	MIMIC II MIMIC III It probably mixes patients.	*"We used 70% of the total data for training our model, 15% for validation, and the remaining 15% for testing. The training, validation, and test datasets were completely separated from each other."*	U-NET	(MAE ± std) SBP: 3.68 ± 4.42 DBP: 1.97 ± 2.92

TABLE 7. (Continued.) Summary of papers in literature that proposed BP estimation methods.

Cuffless blood pressure estimation from ppg signals and its derivatives using deep learning models [36]	MIMIC II - UCI It probably mixes patients.	"The datasets were divided into 70% train, 15% validation and 15% test sets. The test set was reserved for the final evaluation of the optimised model and remained completely disjoint from the training data"	**a) Bidirectional LSTM + LSTM.** b) MLR. c) GRU	(MAE ± std) SBP: 4.51 ± 7.81 DBP: 2.6 ± 4.41
A multi-type features fusion neural network for blood pressure prediction based on photoplethysmography [37]	MIMIC II - UCI It mixes patients.	"In this experiment, 60% of the data is used as a training set, 20% as a test set, and 20% as a validation set. The three branch networks were trained separately and the results were recorded. We repeated the training for each model several times, each time using a different random seed to ensure the randomness of the data."	CNN + BLSTM	(ME ± std) SBP: −1.13 ± 7.25 DBP: 0.14 ± 4.48
Cuff-less blood pressure estimation from photoplethysmography via visibility graph and transfer learning [38]	MIMIC II - UCI It probably mixes patients.	"For every combination of these settings, we used 85% of the available PPG windows for training, and 15% for testing"	CNN + regression **a) AlexNet** b) Inception v3 c) VGG-19	(ME ± std) SBP: −0.00 ± 8.46 DBP: 0.04 ± 5.36
Deepcnap: A deep learning approach for continuous noninvasive arterial blood pressure monitoring using photoplethysmography [39]	MIMIC II - UCI It mixes patients.	"The dataset was randomly split into 80% for training, 10% for validation, and 10% for test. Finally, a 10-fold cross-validation method was conducted to evaluate the data generalization of the models"	Residual neural network with U-Net	(MAE ± std) SBP: 3.5 ± 4.21 DBP: 1.81 ± 2.30
Assessment of non-invasive blood pressure prediction from ppg and rppg signals using deep learning [42]	Exp: 1) MIMIC II - UCI. It mixes patients to decide window size. 2.a) MIMIC III It do not mix patients. 2.b) MIMIC III It mixes patients.	"To determine the optimal window length we employed a twofold strategy. 2.a) The datasets were split into training, validation and test sets on a subject-basis to prevent contamination of the validation and test set by training data. We used 3750 subjects for training (1M samples) and 625 (250k samples) subjects for validation and testing. 2.b) In a second experiment, we randomly selected 750 subjects from our sample pool to create the dataset. Each of these subjects contributed 2000 samples. In contrast to the first experiment, the dataset was split randomly into training, validation and test set. The goal was to evaluate the difference in performance between mixed and non-mixed datasets"	**a) ResNet** b) AlexNet c) Slapničar et al. [46] d) bidirectional LSTM	(MAE) Mixed SBP: 16.4 DBP: 8.5 Non Mixed SBP: 7.7 DBP: 4.4
Optical blood pressure estimation with photoplethysmography and fft-based neural networks [43]	a) MIMIC II It probably mixes patients. b) External test on healthy volunteers.	"A Levenberg-Marquardt algorithm was used to train the ANN. 70% of data were used for training, 15% for validation and 15% for test. To identify outliers, we firstly trained individual ANN (MLP) for each patient separately. In the end, we trained the general ANN with 69 patient data, which includes 175,477 waveforms"	ANN-MLP	(RMSE ± std) SBP: 0.06 ± 7.08 DBP: 0.01 ± 4.66
Blood pressure estimation from ppg signals using convolutional neural networks and siamese network [44]	MIMIC II It do not mix patients. Tested with and without calibration.	"We randomly divided the clean dataset into 60% training set, 20% validation set and 20% test set. Unlike many other works in this field, great attention was paid to separating patients (and not windows) across the three sets."	a) No calibration: CNN b) With calibration: Siamese	(MAE ± std) No calibration SBP: 7.34 ± 8.65 DBP: 3.91 ± 4.48 With calibration SBP: 5.95 ± 6.69 DBP: 3.41 ± 3.97
A shallow u-net architecture for reliably predicting blood pressure (bp) from photoplethysmogram (ppg) and electrocardiogram (ecg) signals [45]	a) MIMIC II - UCI b) Ballistocardiogram (BCG) Apparently, it do not mix patients.	"The UCI dataset (12,000 instances from 942 subjects) was originally divided into four equal 'parts'. The first three parts of the UCI dataset were combined to make the train set (75% of the dataset) and the fourth part was taken as an independent test set (25% of the dataset). These four parts being independent in terms of subjects (i.e.,	Autoencoders (U-NET) + Regressor	(ME ± std) SBP: −0.018 ± 2.876 DBP: 0.09 ± 0.94

TABLE 7. *(Continued.)* **Summary of papers in literature that proposed BP estimation methods.**

		no overlap of subject data across these parts). The external BCG dataset was investigated using two different methods. (...) Secondly, (...) the model was trained using the BCG dataset through 5-Fold Cross-Validation (Method 2)... the BCG dataset was divided into train-test fold (80:20) and validated using a five-fold cross-validation approach"		
Blood pressure estimation from photoplethysmogram using a spectro-temporal deep neural network [46]	MIMIC III It do not mix patients.	*"a leave-one-subject-out (LOSO) experiment was ran"*	CNN - Custom ResNet	(MAE) SBP: 9.43 DBP: 6.88
An unobtrusive and calibration-free blood pressure estimation method using photoplethysmography and biometrics [47]	Own data (n=661) It do not mix patients.	*"leave-one-out procedure all the measurement data from this particular subject were excluded from the training set to avoid contamination"*	Random forest	(MAE) Older SBP: -0.68 ± 14.1 DBP: -0.20 ± 9.0 Young SBP: 0.45 ± 11.3 DBP: 0.31 ± 8.55
Estimation of continuous blood pressure from ppg via a federated learning approach [48]	a) MIMIC II - UCI b) Queensland Vital Signs It do not mix patients.	*"Using a completely independent test dataset from the training dataset grants us the freedom to implement a leave-one-out strategy and see how well our model generalises to other ABP-PPG datasets."*	time-series-to-time-series generative adversarial networks (GAN)	(MAP) MAP-BP: -4.02 ± 22.6

MAE: Mean Absolute Error; ME: Mean Error; MAP: Mean Arterial Pressure; RMSE: Root Mean Square Error;
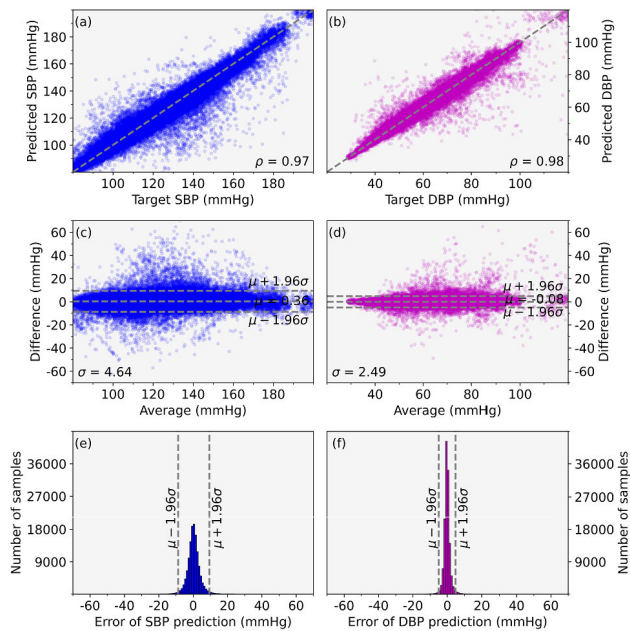

FIGURE 7. Scattering of the target versus prediction, (a) and (b), Bland-Altman plot, (c) and (d), and histogram of the errors, (e) and (f), for SBP and DBP estimation using residual U-Net in the MIMIC-III database, by considering the intra-subject variability. DBP: diastolic blood pressure; SBP: systolic blood pressure; μ: mean; σ: standard deviation; ρ: correlation coefficient.

FIGURE 8. Scattering of the target versus prediction, (a) and (b), Bland-Altman plot, (c) and (d), and histogram of the errors, (e) and (f), for SBP and DBP estimation using residual U-Net in the MIMIC-III database, by considering the inter-subject variability. DBP: diastolic blood pressure; SBP: systolic blood pressure; μ: mean; σ: standard deviation; ρ: correlation coefficient.

conditions and under the effects of medication. Therefore, PPG and ABP waveforms from these patients are not necessarily representative samples of a broader healthy population.
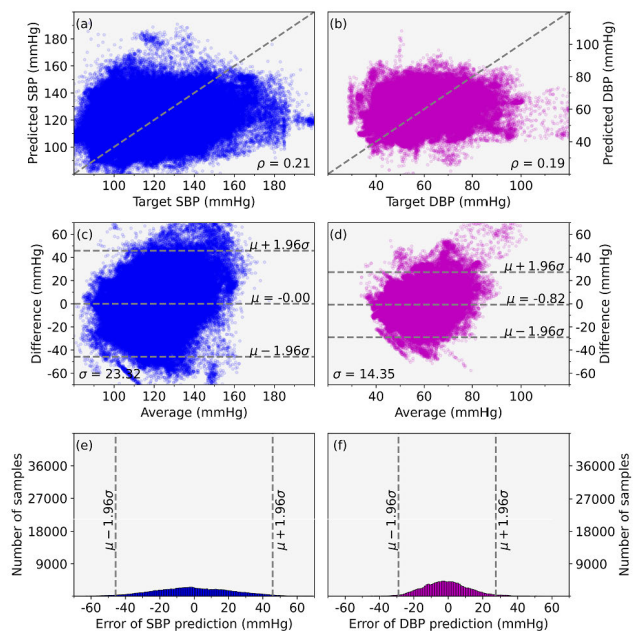
Second, our results are limited to a single PPG analysis. Because it was out of our scope, we did not investigate PPG combination with other signals such as ECG for predicting
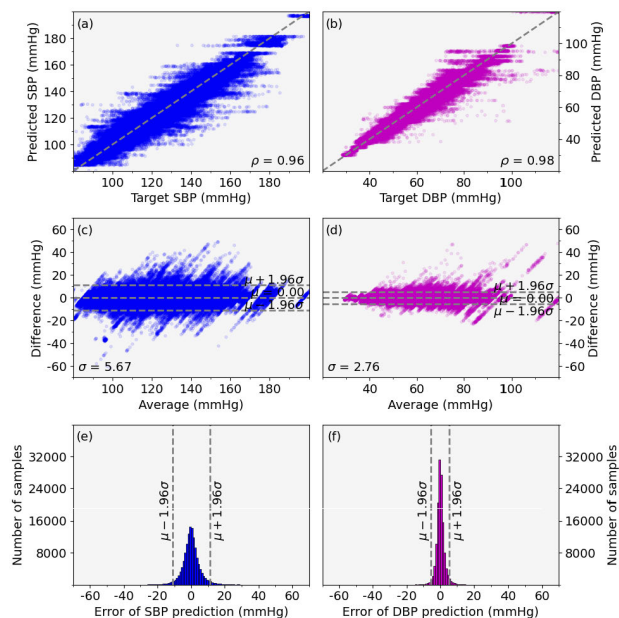
FIGURE 9. Scattering of the target versus prediction, (a) and (b), Bland-Altman plot, (c) and (d), and histogram of the errors, (e) and (f), for SBP and DBP estimation using XGBoost in the MIMIC-III database, by considering only age, sex, weight and patient index number as attributes. DBP: diastolic blood pressure; SBP: systolic blood pressure; μ: mean; σ: standard deviation; ρ: correlation coefficient.

BP. Third, our results cannot be extended to works using calibration strategies. We only use a continuous 30-minute segment of each patient, which does not have enough variance to study how long a calibration fit lasts.

V. CONCLUSION

The importance of BP monitoring for diagnosing, assessing, and treating cardiovascular diseases resides in the fact that this simple procedure (with clinical follow-up) can prevent diseases from becoming severe. However, reference devices to perform BP measurements remain aneroid, mercury, and electronic sphygmomanometers, providing discontinuous values through repeated cuff inflation—uncomfortable and painful methods. On the other hand, PPG monitors BP and screens for hypertension because it can be adapted to a wristband or a ring. PPG and ABP are strongly correlated, and ultimately both are generated by the cardiovascular system. The former can carry information about the latter.

Guided by this principle, many works have sought signal processing and machine-learning techniques to estimate BP from a single PPG, especially those for which the calibration step can be waived. These works have reported satisfactory results. As a matter of principle, we then investigate how intra- and inter-subject variabilities in BP influence the training of a representative set of machine-learning algorithms—by considering two different cross-validation schemes. They succeed in the first scheme and fail decisively in the second scheme. We also found that, regarding the intra-subject scenario, these machine-learning algorithms—covering both

PPG feature-based and PPG signal-based methods—achieve similar performance with an XGBoost-based regression using age, sex, weight, and patient index number as attributes. The latter finding suggests that, although BP is the training label during the learning process, it is a feature for recognizing individuals. Hence, any powerful technique can learn the average BP values per subject and reach a performance meeting the evaluation metrics.

However, considering the importance and benefits of continuous long-term BP monitoring in non-clinical settings, it is imperative to further investigate the constraints of PPG-based methods for predicting BP, before evaluating their practical feasibility. While striving for a universally applicable algorithm, the potential suitability of specialized algorithms cannot be dismissed. Furthermore, the criteria set forth by AAMI and BHS standards seem to be excessively rigorous for wearable devices intended mainly for home health-care use rather than hospital applications. Ultimately, even a less precise BP estimation may offer valuable insights for preventing cardiovascular diseases.

REFERENCES

[1] R. Lozano, R. Lozano, M. Naghavi, K. Foreman, S. Lim, K. Shibuya, V. Aboyans, J. Abraham, T. Adair, R. Aggarwal, and S. Y. Ahn, "Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: A systematic analysis for the global burden of disease study 2010," *Lancet*, vol. 380, no. 9859, pp. 2095–2128, 2012.

[2] F. H. Messerli and B. W. E. Ritz, "Essential hypertension," *Lancet*, vol. 370, no. 9587, pp. 591–603, 2007.

[3] P. E. Drawz, N. M. Pajewski, J. T. Bates, N. A. Bello, W. C. Cushman, J. P. Dwyer, L. J. Fine, D. C. Goff Jr., W. E. Haley, and M. Krousel-Wood, "Effect of intensive versus standard clinic-based hypertension management on ambulatory blood pressure: results from the sprint (systolic blood pressure intervention trial) ambulatory blood pressure study," *Hypertension*, vol. 69, no. 1, pp. 42–50, 2017.

[4] D. M. Rabi, K. A. McBrien, R. Sapir-Pichhadze, M. Nakhla, S. B. Ahmed, S. M. Dumanski, and S. Butalia, "Hypertension Canada's 2020 comprehensive guidelines for the prevention, diagnosis, risk assessment, and treatment of hypertension in adults and children," *Can. J. Cardiol.*, vol. 36, no. 5, pp. 596–624, 2020.

[5] S. Lewington, R. Clarke, N. Qizilbash, and R. Peto, "Age-specific relevance of usual blood pressure to vascular mortality: A meta-analysis of individual data for one million adults in 61 prospective studies," *Lancet*, vol. 360, no. 9349, pp. 1903–1913, 2002.

[6] G. Ogedegbe and T. Pickering, "Principles and techniques of blood pressure measurement," *Cardiol. Clinics*, vol. 28, no. 4, pp. 571–586, Nov. 2010.

[7] G. Parati et al., "European society of hypertension practice guidelines for ambulatory blood pressure monitoring," *J. Hypertension*, vol. 32, no. 7, pp. 1359–1366, 2014.

[8] P. K. Whelton et al., "2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ ASH/ASPC/NMA/PCNA guideline for the prevention, detection, evaluation, and management of high blood pressure in adults: A report of the American college of cardiology/American heart association task force on clinical practice guidelines," *Hypertension*, vol. 71, no. 6, pp. 1–10, Jun. 2018.

[9] G. Parati, "Clinical relevance of continuous and cuffless blood pressure monitoring," in *The Handbook of Cuffless Blood Pressure Monitoring*. Cham, Switzerland: Springer, 2019, pp. 9–13.

[10] G. Mancia, A. Ferrari, L. Gregorini, G. Parati, G. Pomidossi, G. Bertinieri, G. Grassi, M. di Rienzo, A. Pedotti, and A. Zanchetti, "Blood pressure and heart rate variabilities in normotensive and hypertensive human beings," *Circulat. Res.*, vol. 53, no. 1, pp. 96–104, Jul. 1983.

[11] G. Parati, J. E. Ochoa, C. Lombardi, and G. Bilo, "Assessment and management of blood-pressure variability," *Nature Rev. Cardiol.*, vol. 10, no. 3, pp. 143–155, Mar. 2013.

[12] S. L. Stevens, S. Wood, C. Koshiaris, K. Law, P. Glasziou, R. J. Stevens, and J. R. McManus, "Blood pressure variability and cardiovascular disease: Systematic review and meta-analysis," *BMJ*, vol. 354, pp. 1–7, Jan. 2016.

[13] K. Kario, L. Thijs, and J. A. Staessen, "Blood pressure measurement and treatment decisions: Masked and white-coat hypertension," *Circulat. Res.*, vol. 124, no. 7, pp. 990–1008, Mar. 2019.

[14] B. Saugel, R. Dueck, and J. Wagner, "Measurement of blood pressure," *Best Pract. Res. Clin. Anaesthesiol.*, vol. 28, no. 4, pp. 309–322, 2014.

[15] T. Le, F. Ellington, T. Lee, K. Vo, M. Khine, S. K. Krishnan, N. Dutt, and H. Cao, "Continuous non-invasive blood pressure monitoring: A methodological review on measurement techniques," *IEEE Access*, vol. 8, pp. 212478–212498, 2020.

[16] M. S. Dhillon and M. J. Banet, "Pulse arrival time techniques," in *The Handbook of Cuffless Blood Pressure Monitoring*, 2019, pp. 43–59.

[17] S. Rhee, B.-H. Yang, and H. H. Asada, "Artifact-resistant power-efficient design of finger-ring plethysmographic sensors," *IEEE Trans. Biomed. Eng.*, vol. 48, no. 7, pp. 795–805, Jul. 2001.

[18] T. Tamura, Y. Maeda, M. Sekine, and M. Yoshida, "Wearable photoplethysmographic sensors—Past and present," *Electronics*, vol. 3, no. 2, pp. 282–302, Apr. 2014.

[19] Y. Liang, Z. Chen, R. Ward, and M. Elgendi, "Hypertension assessment using photoplethysmography: A risk stratification approach," *J. Clin. Med.*, vol. 8, no. 1, pp. 1–15, 2018.

[20] M. Elgendi, R. Fletcher, Y. Liang, N. Howard, N. H. Lovell, D. Abbott, K. Lim, and R. Ward, "The use of photoplethysmography for assessing hypertension," *NPJ Digit. Med.*, vol. 2, no. 1, pp. 1–11, Jun. 2019.

[21] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiol. Meas.*, vol. 28, no. 3, pp. 1–39, Feb. 2007.

[22] A. A. Alian and H. K. Shelley, "Photoplethysmography," *Best Pract. Res. Clin. Anaesthesiol.*, vol. 28, no. 4, pp. 395–406, 2014.

[23] M. Elgendi, "On the analysis of fingertip photoplethysmogram signals," *Current Cardiol. Rev.*, vol. 8, no. 1, pp. 14–25, Jun. 2012.

[24] M. Nitzan, A. Romem, and R. Koppel, "Pulse oximetry: Fundamentals and technology update," *Med. Devices*, vol. 7, pp. 231–239, Jul. 2014.

[25] G. Martínez, N. Howard, D. Abbott, K. Lim, R. Ward, and M. Elgendi, "Can photoplethysmography replace arterial blood pressure in the assessment of blood pressure?" *J. Clin. Med.*, vol. 7, no. 10, pp. 2–13, 2018.

[26] 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," *Biomed. Signal Process. Control*, vol. 58, pp. 1–14, Jan. 2020.

[27] L. Wang, W. Zhou, Y. Xing, and X. Zhou, "A novel neural network model for blood pressure estimation using photoplethesmography without electrocardiogram," *J. Healthcare Eng.*, vol. 2018, pp. 1–9, Jan. 2018.

[28] S. G. Khalid, J. Zhang, F. Chen, and D. Zheng, "Blood pressure estimation using photoplethysmography only: Comparison between different machine learning approaches," *J. Healthcare Eng.*, vol. 2018, pp. 1–13, Oct. 2018.

[29] S. S. Mousavi, M. Firouzmand, M. Charmi, M. Hemmati, M. Moghadam, and Y. Ghorbani, "Blood pressure estimation from appropriate and inappropriate PPG signals using a whole-based method," *Biomed. Signal Process. Control*, vol. 47, pp. 196–206, Jan. 2019.

[30] N. Hasanzadeh, M. M. Ahmadi, and H. Mohammadzade, "Blood pressure estimation using photoplethysmogram signal and its morphological features," *IEEE Sensors J.*, vol. 20, no. 8, pp. 4300–4310, Apr. 2020.

[31] M. H. Chowdhury, M. N. I. Shuzan, M. E. H. Chowdhury, Z. B. Mahbub, M. M. Uddin, A. Khandakar, and M. B. I. Reaz, "Estimating blood pressure from the photoplethysmogram signal and demographic features using machine learning techniques," *Sensors*, vol. 20, no. 11, pp. 1–24, 2020.

[32] M. Panwar, A. Gautam, D. Biswas, and A. Acharyya, "PP-Net: A deep learning framework for PPG-based blood pressure and heart rate estimation," *IEEE Sensors J.*, vol. 20, no. 17, pp. 10000–10011, Sep. 2020.

[33] Y.-C. Hsu, Y.-H. Li, C.-C. Chang, and L. N. Harfiya, "Generalized deep neural network model for cuffless blood pressure estimation with photoplethysmogram signal only," *Sensors*, vol. 20, no. 19, pp. 1–19, 2020.

[34] C. El-Hajj and P. A. Kyriacou, "Deep learning models for cuffless blood pressure monitoring from PPG signals using attention mechanism," *Biomed. Signal Process. Control*, vol. 65, pp. 1–15, Jan. 2021.

[35] T. Athaya and S. Choi, "An estimation method of continuous non-invasive arterial blood pressure waveform using photoplethysmography: A u-net architecture-based approach," *Sensors*, vol. 21, no. 5, pp. 1–17, 2021.

[36] C. El-Hajj and P. Kyriacou, "Cuffless blood pressure estimation from PPG signals and its derivatives using deep learning models," *Biomed. Signal Process. Control*, vol. 70, pp. 1–12, Jan. 2021.

[37] M. Rong and K. Li, "A multi-type features fusion neural network for blood pressure prediction based on photoplethysmography," *Biomed. Signal Process. Control*, vol. 68, pp. 1–8, Oct. 2021.

[38] W. Wang, P. Mohseni, K. L. Kilgore, and L. Najafizadeh, "Cuff-less blood pressure estimation from photoplethysmography via visibility graph and transfer learning," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 5, pp. 2075–2085, May 2022.

[39] D. Kim, Y. Kim, H. Kim, and D. Kim, "DeepCNAP: A deep learning approach for continuous noninvasive arterial blood pressure monitoring using photoplethysmography," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 8, pp. 3697–3707, Aug. 2022.

[40] E. O'Brien, J. Petrie, W. A. Littler, M. De Swiet, P. L. Padfield, D. Altman, M. Bland, A. Coats, and N. Atkins, "The British hypertension society protocol for the evaluation of blood pressure measuring devices," *J. Hypertension*, vol. 11, no. 2, pp. S43–S62, 1993.

[41] W. B. White, A. S. Berson, C. Robbins, M. J. Jamieson, L. M. Prisant, E. Roccella, and S. G. Sheps, "National standard for measurement of resting and ambulatory blood pressures with automated sphygmomanometers," *Hypertension*, vol. 21, no. 4, pp. 504–509, Apr. 1993.

[42] F. Schrumpf, P. Frenzel, C. Aust, G. Osterhoff, and M. Fuchs, "Assessment of non-invasive blood pressure prediction from PPG and RPPG signals using deep learning," *Sensors*, vol. 21, no. 18, pp. 1–27, 2021.

[43] X. Xing and M. Sun, "Optical blood pressure estimation with photoplethysmography and FFT-based neural networks," *Biomed. Opt. Exp.*, vol. 7, pp. 3007–3020, Aug. 2016.

[44] O. Schlesinger, N. Vigderhouse, D. Eytan, and Y. Moshe, "Blood pressure estimation from PPG signals using convolutional neural networks and Siamese network," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 1135–1139.

[45] S. Mahmud, N. Ibtehaz, A. Khandakar, A. M. Tahir, T. Rahman, K. R. Islam, and M. S. Hossain, "A shallow U-Net architecture for reliably predicting blood pressure (BP) from photoplethysmogram (PPG) and electrocardiogram (ECG) signals," *Sensors*, vol. 22, no. 3, pp. 1–23, 2022.

[46] G. Slapničar, N. Mlakar, and M. Luštrek, "Blood pressure estimation from photoplethysmogram using a spectro-temporal deep neural network," *Sensors*, vol. 19, no. 15, pp. 1–17, 2019.

[47] X. Xing, Z. Ma, M. Zhang, Y. Zhou, W. Dong, and M. Song, "An unobtrusive and calibration-free blood pressure estimation method using photoplethysmography and biometrics," *Sci. Rep.*, vol. 9, no. 1, pp. 1–8, Jun. 2019.

[48] E. Brophy, M. De Vos, G. Boylan, and T. Ward, "Estimation of continuous blood pressure from PPG via a federated learning approach," *Sensors*, vol. 21, no. 18, pp. 1–12, 2021.

[49] A. Paviglianiti, V. Randazzo, S. Villata, G. Cirrincione, and E. Pasero, "A comparison of deep learning techniques for arterial blood pressure prediction," *Cognit. Comput.*, vol. 2021, pp. 1–22, Jan. 2021.

[50] J. Leitner, P. Chiang, and S. Dey, "Personalized blood pressure estimation using photoplethysmography: A transfer learning approach," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 1, pp. 218–228, Jan. 2022.

[51] M. Kachuee, M. M. Kiani, H. Mohammadzade, and M. Shabany, "Cuffless blood pressure estimation algorithms for continuous health-care monitoring," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 4, pp. 859–869, Apr. 2017.

[52] G. Wang, M. Atef, and Y. Lian, "Towards a continuous non-invasive cuffless blood pressure monitoring system using PPG: Systems and circuits review," *IEEE Circuits Syst. Mag.*, vol. 18, no. 3, pp. 6–26, 3rd Quart., 2018.

[53] E. Martinez-Ríos, L. Montesinos, M. Alfaro-Ponce, and L. Pecchia, "A review of machine learning in hypertension detection and blood pressure estimation based on clinical and physiological data," *Biomed. Signal Process. Control*, vol. 68, pp. 1–14, Jan. 2021.

[54] M. Sharma, K. Barbosa, V. Ho, D. Griggs, T. Ghirmai, S. K. Krishnan, T. K. Hsiai, J.-C. Chiao, and H. Cao, "Cuff-less and continuous blood pressure monitoring: A methodological review," *Technology*, vol. 5, no. 2, pp. 1–21, 2017.

[55] R. Mukkamala, J. Hahn, O. T. Inan, L. K. Mestha, C. Kim, H. Töreyin, and S. Kyal, "Toward ubiquitous blood pressure monitoring via pulse transit time: Theory and practice," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 8, pp. 1879–1901, Aug. 2015.

[56] Y. Yoon, J. M. Kang, Y. Kwon, S. Park, S. Noh, Y. Kim, J. Park, and S. W. Hwang, "Cuff-less blood pressure estimation using pulse waveform analysis and pulse arrival time," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 4, pp. 1068–1074, Jul. 2018.

[57] C. Vlachopoulos, M. O'Rourke, and W. W. Nichols, *McDonald's Blood Flow in Arteries: Theoretical, Experimental and Clinical Principles.* Boca Raton, FL, USA: CRC Press, 2011.

[58] D. B. McCombie, A. T. Reisner, and H. H. Asada, "Adaptive blood pressure estimation from wearable PPG sensors using peripheral artery pulse wave velocity measurements and multi-channel blind identification of local arterial dynamics," in *Proc. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2006, pp. 3521–3524.

[59] P. Segers, J. Kips, B. Trachet, A. Swillens, S. Vermeersch, D. Mahieu, E. Rietzschel, M. De Buyzere, and L. Van Bortel, "Limitations and pitfalls of non-invasive measurement of arterial pressure wave reflections and pulse wave velocity," *Artery Res.*, vol. 3, no. 2, pp. 79–88, 2009.

[60] S. A. M. Huybrechts, D. G. Devos, S. J. Vermeersch, D. Mahieu, E. Achten, T. L. M. de Backer, P. Segers, and L. M. van Bortel, "Carotid to femoral pulse wave velocity: A comparison of real travelled aortic path lengths determined by MRI and superficial measurements," *J. Hypertension*, vol. 29, no. 8, pp. 1577–1582, 2011.

[61] X. F. Teng and Y. T. Zhang, "Continuous and noninvasive estimation of arterial blood pressure using a photoplethysmographic approach," in *Proc. 25th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Apr. 2003, pp. 3153–3156.

[62] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, Jun. 2000.

[63] G. Moody and R. Mark, "A database to support development and evaluation of intelligent intensive care monitoring," in *Proc. Comput. Cardiol.*, Sep. 1996, pp. 657–660.

[64] M. Saeed, C. Lieu, G. Raber, and R. Mark, "Mimic II: A massive temporal ICU patient database to support research in intelligent patient monitoring," in *Proc. Comput. Cardiol.*, 2002, pp. 641–644.

[65] M. Saeed, M. Villarroel, A. T. Reisner, G. Clifford, L.-W. Lehman, G. Moody, T. Heldt, T. H. Kyaw, B. Moody, and R. G. Mark, "Multiparameter intelligent monitoring in intensive care ii (MIMIC-II): A public-access intensive care unit database," *Crit. Care Med.*, vol. 39, no. 5, pp. 1–18, 2011.

[66] A. E. W. Johnson, T. J. Pollard, L. Shen, L.-W.-H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. Anthony Celi, and R. G. Mark, "MIMIC-III, a freely accessible critical care database," *Sci. Data*, vol. 3, no. 1, pp. 1–9, May 2016.

[67] B. Moody, G. Moody, M. Villarroel, G. Clifford, and I. Silva. (2020). *MIMIC-III Waveform Database Matched Subset (Version 1.0).* [Online]. Available: https://physionet.org/content/mimic3wdb-matched/1.0/

[68] M. Kachuee, M. M. Kiani, H. Mohammadzade, and M. Shabany, "Cuffless high-accuracy calibration-free blood pressure estimation using pulse transit time," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2015, pp. 1006–1009.

[69] M. Kachuee, M. M. Kiani, H. Mohammadzade, and M. Shabany. (2015). *UCI Machine Learning Repository: Cuff-Less Blood Pressure Estimation Data Set.* [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Cuff-Less+Blood+Pressure+Estimation

[70] Y. Liang, M. Elgendi, Z. Chen, and R. Ward, "An optimal filter for short photoplethysmogram signals," *Sci. Data*, vol. 5, no. 1, pp. 1–12, May 2018.

[71] J. X. Sun, A. T. Reisner, and R. G. Mark, "A signal abnormality index for arterial blood pressure waveforms," in *Proc. Comput. Cardiol.*, 2006, pp. 13–16.

[72] W.-H. Lin, X. Li, Y. Li, G. Li, and F. Chen, "Investigating the physiological mechanisms of the photoplethysmogram features for blood pressure estimation," *Physiolog. Meas.*, vol. 41, no. 4, pp. 1–13, 2020.

[73] J. H. Friedman, "Greedy function approximation: A gradient boosting machine.," *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001.

[74] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 785–794.

[75] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "LightGBM: A highly efficient gradient boosting decision tree," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–15.

[76] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: unbiased boosting with categorical features," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 31, 2018, pp. 1–22.

[77] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2016, pp. 770–778.

[78] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[79] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2015, pp. 234–241.

[80] N. Siddique, S. Paheding, C. P. Elkin, and V. Devabhaktuni, "U-Net and its variants for medical image segmentation: A review of theory and applications," *IEEE Access*, vol. 9, pp. 82031–82057, 2021.

[81] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual U-Net," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 749–753, May 2018.

[82] N. Ibtehaz, S. Mahmud, M. E. H. Chowdhury, A. Khandakar, M. A. Ayari, A. Tahir, and M. S. Rahman, "PPG2ABP: Translating photoplethysmogram (PPG) signals to arterial blood pressure (ABP) waveforms using fully convolutional neural networks," 2020, *arXiv:2005.01669*.

[83] F. M. Dias, T. B. S. Costa, D. A. C. Cardenas, M. A. F. Toledo, J. E. Krieger, and A. M Gutierrez, "A machine learning approach to predict arterial blood pressure from photoplethysmography signal," in *Proc. Comput. Cardiol. (CinC)*, vol. 498, 2022, pp. 1–4.

[84] W. J. Verberk, "Design of clinical trials to validate cuffless blood pressure monitors," in *The Handbook of Cuffless Blood Pressure Monitoring*, 2019, pp. 203–224.

[85] R. Mukkamala, M. Yavarimanesh, K. Natarajan, J.-O. Hahn, K. G. Kyriakoulis, A. P. Avolio, and G. S. Stergiou, "Evaluation of the accuracy of cuffless blood pressure measurement devices: Challenges and proposals," *Hypertension*, vol. 78, no. 5, pp. 1161–1167, Nov. 2021.

THIAGO BULHÕES DA SILVA COSTA received the Graduate, master's, and Ph.D. degrees in electrical engineering from the University of Campinas, Brazil, in 2013, 2015, and 2020, respectively. Currently, he is a Visiting Professor with the Federal University of ABC, Brazil. His research interests include biomedical signal processing, machine learning, brain–computer interfaces, and neuroscience.

FELIPE MENEGUITTI DIAS received the B.Sc. and M.Sc. degrees from the Federal University of Juiz de Fora (UFJF), in 2017 and 2020, respectively. He is currently pursuing the Ph.D. degree in biomedical engineering with the University of São Paulo (USP), working with machine learning applications in electrocardiogram and photoplethysmogram biomedical signals. Furthermore, he is also a Researcher with the Heart Institute (Incor-HCFMUSP). His research interests include biomedical signal processing, machine learning, and compressive sensing.

DIEGO ARMANDO CARDONA CARDENAS received the degree in bioengineering from the University of Antioquia, Colombia, and the M.Sc. and Ph.D. degrees in electrical engineering from the University of São Paulo, in 2018. He is currently a Researcher with the Laboratory of Biomedical Informatics, Heart Institute, Clinics Hospital, University of São Paulo Medical School. His research interests include medical image processing, pattern recognition, ultrasound tomography, biological signal processing, and machine learning.

MARCELO ARRUDA FIUZA DE TOLEDO received the B.Sc., Lic., M.Sc., and Ph.D. degrees in biology from the University of São Paulo (USP), in 2009, 2010, 2013, and 2018, respectively. He is currently a Researcher with the Laboratory of Biomedical Informatics, Heart Institute, Clinics Hospital, University of São Paulo Medical School. His research interests include animal physiology, neuroscience, animal behavior, computational biology, and biological applications of artificial intelligence and statistics, especially in machine learning applied to medical image analysis and signals. For more information visit the link (http://lattes.cnpq.br/6595953925093891).

JOSE EDUARDO KRIEGER received the M.D. and Ph.D. degrees. He is currently a Professor in genetics and molecular medicine with the University of São Paulo Medical School and the Director of the Laboratory of Genetics & Molecular Cardiology, Heart Institute (InCor). His research interests include the genetic determinants of cardiovascular diseases to improve health management algorithms and to the development of novel therapeutics.

DANIEL MÁRIO DE LIMA is Graduated in technologist information systems from IFPI - Piaui Federal Institute, Brazil, in 2008, and the M.Sc. degree in computer science and computational mathematics from the ICMC, University of São Paulo, Brazil, in 2013, where he is currently pursuing the Ph.D. degree in computer science and computational mathematics. He is a Research Software Developer with NYU Langone Health. His research interests include databases, information visualization and deep learning, especially applications in automation, energy, and healthcare.

MARCO ANTONIO GUTIERREZ received the B.Eng. and D.Sc. degrees in electrical engineering from the University of São Paulo, Brazil, in 1985 and 1996, respectively. He has been with the Heart Institute, University of São Paulo, since 1986, where he is currently the Head of the Biomedical Informatics Laboratory and the Informatics Division. He has been an Assistant Professor with the Polytechnic School, since 1997, the School of Medicine, since 2004, and the University of São Paulo, Brazil. His research interests include biomedical image and signal processing and health information systems.

• • •