Robust Feature Selection for Continuous BP Estimation in Multiple Populations: Towards Cuffless Ambulatory BP Monitoring

Author(s):
Cisnal, Ana; Li, Yanke (iD); Fuchs, Bertram; Ejtehadi, Mehdi; Riener, Robert; Paez-Granados, Diego (iD)

Robust Feature Selection for Continuous BP Estimation in Multiple Populations: Towards Cuffless Ambulatory BP Monitoring

Ana Cisnal, *Student Member, IEEE*, Yanke Li, Bertram Fuchs, Mehdi Ejtehadi,
Robert Riener, *Senior Member, IEEE*, and Diego Paez-Granados[†], *Member, IEEE*

Abstract— **Current blood pressure (BP) estimation methods have not achieved an accurate and adaptable approach for application in populations at risk of cardiovascular disease, with generally limited sample sizes. Here, we introduce an algorithm for BP estimation solely reliant on photoplethysmography (PPG) signals and demographic features. Our approach automatically obtains signal features and employs the Markov Blanket (MB) feature selection to discern informative and transmissible features, achieving a robust space adaptable to the population shift. We validated our approach with the Aurora-BP database, compromising ambulatory wearable cuffless BP measurements for over 500 individuals. By evaluating several machine-learning regression methods, Gradient Boosting emerged as the most effective. The comparative assessment encompassed both a generic model (trained on unclassified BP data) and specialized models (tailored to each distinct BP population), with the former demonstrating consistent superiority with MAE of** $10.2\text{mmHg}(0.28)$ **for systolic BP and** $6.7\text{mmHg}(0.18)$ **for diastolic BP on the whole dataset. Moreover, a comparison of in-clinic and ambulatory model performance showed a significant decrease in accuracy for the latter of** 2.85mmHg **in systolic (**$p < 0.0001$**, F-value = 32764.76) and** 2.82mmHg **for diastolic (**$p < 0.0001$**, F-value = 65675.36) estimation errors. Our work contributes to a resilient BP estimation algorithm from PPG signals, underscoring the advantages of causal feature selection and quantifying the disparities between ambulatory and in-clinic measurements.**

Index Terms— **Continuous cuffless blood pressure, photoplethysmography, pulse wave analysis.**

A. Cisnal is with the Spinal Cord Injury & Artificial Intelligence Lab, ETH Zurich, Switzerland and with Instituto de las Tecnologías Avanzadas de la Producción (ITAP) at University of Valladolid, Spain (e-mail: cisnal@ieee.org).

B. Fuchs is with the Spinal Cord Injury & Artificial Intelligence Lab, ETH Zurich, Switzerland and the School of Computation, Information and Technology at Technical University of Munich, Germany (e-mail: bfuchs@student.ethz.ch)

Y. Li and M. Ejtehadi are with the Spinal Cord Injury & Artificial Intelligence Lab, ETH Zurich, Switzerland (e-mail: yanke.li, mehdi.ejtehadi@hest.ethz.ch).

R. Riener is with the Sensory-Motor Systems Lab ETH Zurich, and the Spinal Cord Injury Center, University Hospital Balgrist, Zurich, Switzerland (e-mail: robert.riener@hest.ethz.ch)

D. Paez-Granados is the corresponding author and is with the Spinal Cord Injury & Artificial Intelligence Lab, ETH Zurich and Swiss Paraplegic Research (SPF), Switzerland (e-mail: dfpg@ieee.org)

I. INTRODUCTION

BLOOD Pressure (BP) is a widely accepted surrogate biomarker in many health conditions that aid in identifying individuals at risk of cardiovascular disease (CVD) [1] and a direct biomarker in conditions such as hypertension. Traditionally, BP measurements have been obtained in clinical settings, providing valuable insights into a person's cardiovascular health. However, the increasing recognition of the dynamic nature of BP calls for ambulatory measurements that capture BP changes during daily living to further understand the cardiovascular system.

In recent works, various cuffless wearable devices using a variety of signal modalities have been proposed for ambulatory blood pressure (ABP) monitoring, such as electrocardiogram (ECG), tonometry, bioimpedance, and photoplethysmography (PPG), with the latter the most widespread [2]. PPG is a non-invasive, non-occlusive, optical technique for measuring volumetric changes in blood flow. Models based on the pulse arrival time (PAT) [3], pulse transit time (PTT) [4], and pulse wave velocity (PWV) [5] remain as the most common ones. However, these methods require a PPG signal, combined with either an ECG signal or another PPG signal from a different peripheral site.

An alternative is developing models that rely solely on one PPG signal, in turn requiring a pulse wave analysis (PWA) involving a morphological understanding of the PPG pulse to extract features that can be used to estimate BP. This has been shown through multiple linear regression [6], regression trees [7], random forest [8], support vector machine [8] and artificial neural network [9]–[11]. Nonetheless, in the presence of noise and motion artefacts, PPG signals adopt different morphologies; according to Dawber [12], four PPG classes can be distinguished according to the diastolic phase (Fig. 1). The dicrotic notch (DN) is defined as the time point for the end of the systole and the beginning of the diastole. Despite its importance, neither its formation nor its exact location is clear [13]. Moreover, it has been suggested that DN is sometimes only present in people with healthy compliant arteries [14]. Thus previous work had reduced its scope to devices with class I and II, whereas ambulatory measurements are on class III and IV. Our work has been validated for class IV, as we aim for continuous ambulatory BP estimations.
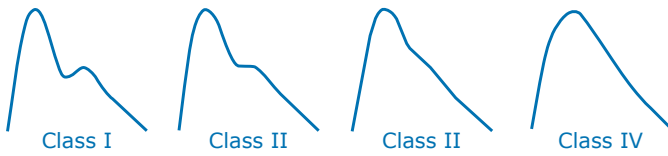
Fig. 1: Classification of PPG waveforms: I) a DN is evident; II) no evident DN but the descendent line becomes horizontal; III) no evident DN but there is a change in the angle of the descendent line; IV) no evidence of a DN is visible.

The accuracy of machine learning (ML) models is fully dependent on the dataset used and the feature space selected, and no agreement has been reached. BP estimation models to date, have considered only a few fiducial points for time-domain feature extraction with a focus on performance optimization for singular datasets, without attending to the problem of generalization across different populations. This is particularly important for populations with limited available sample data, such as pregnant women, neonates, and spinal cord injury (SCI) people. These are in turn the populations that would benefit drastically from continuous BP estimations as they require special care for the cardiovascular function.

Given these populations' small sample sizes, personal models and domain-transferred features are essential for developing accurate and reliable BP estimation models. Transfer learning, involving feature selection, can help to identify a discriminatory set of features and enhance quality learning by dropping less descriptive features. Moreover, since the BP waveform is not yet fully understood, the feature extraction and selection phase is particularly important for subsequent interpretable model development.

The focus and main contribution of this work are three-fold: (1) A method for defining and identifying features in PPG signal for robust transfer among small populations based on the Markov Blanket (MB); (2) the first ambulatory BP estimation model from single wrist PPG using a large dataset: the Aurora-BP [15] where we achieved a boosting model with a mean absolute error (MAE) 11 mmHg in nested cross-validation throughout four different sub-populations; (3) extensive analysis of the cross-population performance and model differences found especially on inpatient vs. ambulatory data, which highlights the gap with previous works where only inpatient data was used. Finally, we provide our method and library for extensive PPG wave signal analysis as an open-access library to motivate advancement in fair model sharing [16].

II. MATERIALS AND METHODS

In contrast to previous works, our approach focuses on extracting a comprehensive set of features through PWA and developing a feature selection process that prioritizes robustness across diverse populations, rather than solely optimizing prediction accuracy for a single dataset. This emphasis on feature robustness contributes to the generalizability and reliability of the BP estimation method. Our proposed method for PPG-based BP estimation consists of five steps (depicted in Fig. 2):

- Signal pre-processing for noise removal, signal segmentation, and data cleaning.
- Identification of the fiducial points in the signals.
- Extraction of signal and demographic features.
- Robust feature selection.
- Prediction of BP using ML models.

A. Dataset

In our work, we selected the Aurora-BP dataset [15] as the first large-scale collection of ambulatory and cuff-less BP measurements obtained over a 24-hour period using wearable technology. A data transfer agreement was signed with Microsoft, and the ETh Zurich IRB waived ethical review for using the dataset. The dataset comprises measurements obtained from both in-clinic and ambulatory setups. The ambulatory measurements include recordings of ECG, tonometry, PPG, and oscillometric measurement of reference BP. In-clinic measurements additionally involved the acquisition of reference BP using the auscultatory protocol. During the measurements, participants underwent an initial clinical visit and a return visit after the 24-hour recording, where reference cuff-based BP measurements were taken. During the ambulatory phase, the cuff-based ABP monitor was automatically triggered every 30 minutes during waking hours and every 60 minutes during the night. For each of the cuff-based BP measurements, 30-second physiological signal segments were recorded at a frequency of 500 Hz. In this study, we utilized data obtained from both ambulatory and in-clinic setups, incorporating oscillometric reference BP, to estimate BP based on the single channel 30-second PPG signals.

B. PPG Signal Pre-processing

Time-domain features are generated based on time, amplitude, and area extracted using PWA and they require the previous identification of the fiducial points e.g., systolic peak (S), dicrotic notch (DN). However, identifying these features can be challenging and most studies only consider S for extraction. Hence, these features are commonly complemented with methods such as PAT, PTT, and PWV values [17]–[20], with the inherent disadvantage of a second signal. Nonetheless, we present here a method to achieve it even with noisy ambulatory data.

The fiducial points of a PPG pulse are identified by analyzing its derivatives, which are highly distorted in the presence of noise. Hence, high-frequency noise removal is required, and low-pass filtering should keep a balance between maintaining the original features and reducing noise. Moreover, some studies employ polynomial interpolation [21] or apply filtering techniques such as 25-ms moving average [22] to smooth the signals and improve the computation of the derivatives. However, these techniques might alter the content of the signal recordings. In our work, the raw PPG signals were filtered using a 4th-order Butterworth bandpass filter with cut-off frequencies of 0.25 and 10 Hz [23].

To extract pulse wave features, it is necessary to identify individual beat-to-beat pulse waves which necessitate the detection of pulse onset and end (Fig. 3). However, the signals'
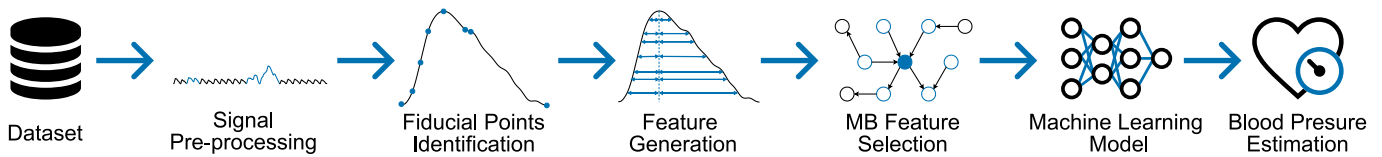
Fig. 2: Schematic diagram illustrating the methodology employed in this study for the estimation of BP using PPG.
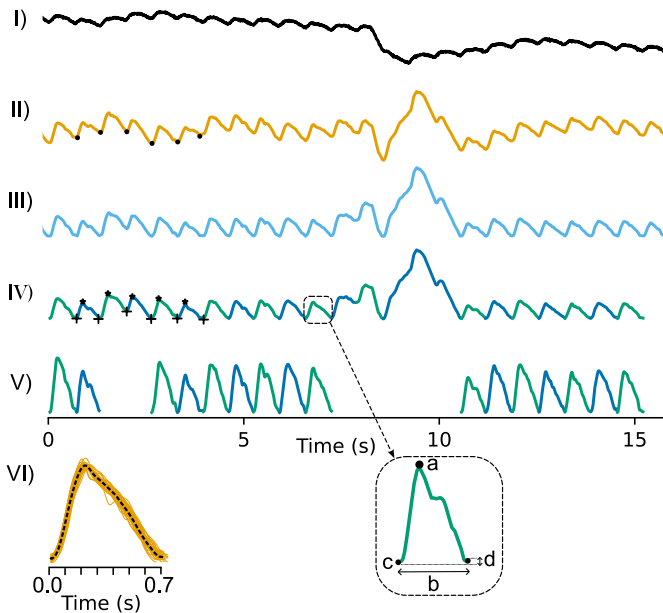

Fig. 3: PPG signal pre-processing steps: I) raw PPG signal; II) signal after Butterworth band-pass filter with DC drift, resulting in amplitude variation of the troughs (·); III) signal after baseline removal; IV) pulse segmentation by detecting peaks (∗) and troughs (+); V) remaining beat-to-beat waves after morphology analysis; VI) template matching, with the mean wave represented by (- -) and individual waves shown by (-). Pulse wave morphology analysis considers: a) peak location, b) pulse width, c) trough position, and d) trough depth difference.

troughs vary in amplitude since the DC components of the PPG signals are affected by biological characteristics such as tissue composition, respiration, vasomotor activity, and thermoregulation as well as external factors, such as light and acquisition device [23]. Hence, the baseline was removed using an adaptive iteratively reweighted penalized least squares [24].

The extraction of the beat-to-beat waves involves the detection of the peaks and troughs using an adaptive amplitude threshold [11]. In this work, a local maximum value was considered a peak if the difference between this point and the adjacent local minimum exceeded the threshold. Similarly, for a local minimum to be considered a trough, the difference between this point and the adjacent local maximum is needed to surpass the threshold. The dynamic threshold was set as 70% of the range between the median values of the identified maxima and minima within the 30-second window.

Following the identification of the pulse waves, the noisy wave units were discarded by analyzing the pulse wave morphology, utilizing four parameters with heuristically determined thresholds. The pulse width, representing the time interval between two consecutive troughs, was constrained to be within 0.3-2 seconds, corresponding to an extreme pulse rate of 30-200 bpm. The maximum value, corresponding to the systolic peak, was expected to occur in the first half part of the segment [25]. Similarly, the trough or minimum value should be located at the beginning or at the end of the segment, corresponding to the onset or valley point. Lastly, the trough depth difference between successive pulses was limited to less than 20% of the PPG segment height [25].

The remaining beat-to-beat waves were normalized to zero mean and unit variance. Given that the dataset provided 30-second PPG segments, we assumed that the PPG morphology should remain relatively stable within this time frame. Therefore, a template-matching approach was employed to further investigate the reliability of the remaining PPG waves [26]. During this procedure, the mean wave was calculated based on all the extracted waves within the time window. Subsequently, the Euclidean distance between each individual wave and the mean wave was calculated. If the distance exceeded a device-specific threshold of the expected noise, indicating a low correlation with the mean wave, the corresponding beat-to-beat wave was deemed unreliable (likely noisy) and discarded for further analysis. Lastly, the analysis progressed only if there were more than five reliable pulse waves; in the absence of this criterion, all pulse waves within the 30-second PPG segment were excluded. The aforementioned signal quality checking procedure is integrated as the initial processing stage within a Python package [16], accessible for open-access usage.

C. Fiducial Points Identification

The fiducial points were identified for each pulse wave using derivative analysis (Fig. 4). Typically, a change of sign of the first derivative identifies the exact point for class I waveforms (Fig. 1), but for PPG without an evident DN, the first derivative is always negative. On the other hand, the peak of the second derivative of a class I wave provides a good but not perfect approximation of the DN [21], [27], or as the location where a change of sign in the first derivative from negative to positive occurs [22]. Likewise, the D is not always noticeable. In such cases, it can be difficult to identify it, and different methods combining the first, second, and/or third derivatives have been proposed in the literature to do so [28], [29].

In our work, we provide algorithm 1 for PPG ambulatory data (including class IV), where the onset (O) and valley (V) points were set to the onset and end of the wave. The S was
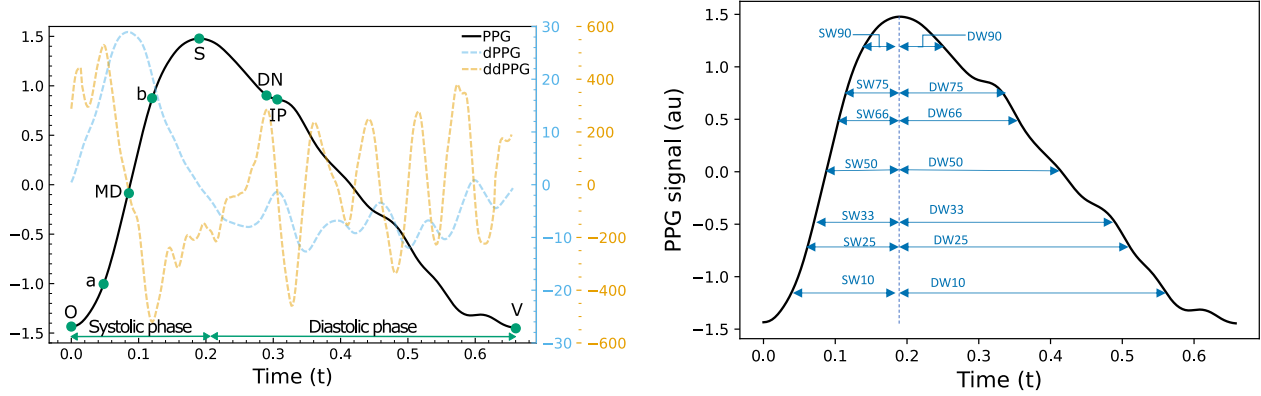
Fig. 4: Left: Identification of the fiducial points based on the analysis of PPG, dPPG (first derivative or velocity of PPG), and ddPPG (second derivative or acceleration of PPG) signals. Right: BW-based features: systolic branch width (SBWx) and diastolic branch width (DBWx).

detected as the largest peak of the pulse, and it was used to split the wave into a systolic phase and a diastolic phase. The maximum derivative (MD) point corresponds to the maximum peak of the first derivative in the systolic phase. The diastolic peak (D) was identified as the local maxima occurring within the diastolic phase, within a time interval of 80 ms to 0.6 times the duration of the diastolic phase. Then, the inflexion point (IP) was identified as the local maximum of the first derivative before D, and DN as the local minimum before IP. In some cases, there were no local maxima and only an inflection point existed. Hence, D corresponded to IP, and it was detected as the absolute maximum of the first derivative in the region of search, while DN was the local maximum in the second derivative. Several local maxima can be identified in the region of search, especially in low-quality signals. In these cases, IP was identified as the maximum peak of the first derivative, D was the local maximum right after IP, and DN was the local minimum before IP. Point a was the maximum second derivative and the b point was the point of strongest negative acceleration in the falling edge [30].

D. Feature Generation

The features of each validated pulse wave were extracted, and subsequently, the average value of each feature within the 30-second window was computed to serve as input for the models. The generated features can be divided into four main groups: time domain (TF), frequency-based (FF), statistical (SF), and demographic features (DF). A Python package that encompasses the processing and feature extraction stages was made readily accessible [16] with this publication. The package accepts a raw signal, along with parameters specifying the window size and step size. As a result, it produces a feature vector that represents the observed time interval. The package allows users to adjust various parameters, including window and step size, as well as additional tuning options to optimize denoising and feature extraction according to specific task requirements. Across multiple windowing steps, a feature table is constructed, which can subsequentially be used for feature selection. Details of each feature can be found in the online VitalPy library [16].

1) Time-domain Features (TF): The proposed time-domain features can be divided into six subcategories regarding intensity, time, area, slope, branch width, and others.

Intensity-based and time-based features include the absolute and normalized value (time or amplitude) of one fiducial point or between two points. Intensity-based features also include the intensity of the first and second derivatives of the PPG fiducial points. The slope and the area under the curve between the two points were also calculated. Additionally, the intensity ratio, time ratio, and area ratio were included. These four types of characteristics were calculated for all possible combinations of the fiducial points.

Features that only depend on the branch width (BW) at a given percentage of the pulse height have been commonly used due to their simplicity [7], [8], [10], [11], [17]. A total of 28 BW-based features were extracted (Fig. 4), including the systolic branch width (SBWx), the diastolic branch width (DBWx), the branch width (DWx = SBWx+DBWx), and the branch width ratio (BWRx=DBWx/SBWx) at x% of the pulse height (x = 10, 25, 33, 50, 66, 75, 90).

Some features have been proven to be specifically related to BP. Reflection index (RI) or augmentation index (AI) measures the pulse reflection, it is related to the arterial tone and it is calculated as the intensity's ratio between the intensity of the S and IP [21], [31]. Inflection point area (IPA) is defined as the ratio of the area between O, MS, S, IP, V and has proved to be an indicator of peripheral resistance [21], [31]. Crest time (CT) is the time difference between O and S, which is related to the PWV [21]. PPGK, also known as PPG characteristic value or K value [17], is related to blood viscosity and total peripheral resistance. Normalized pulse volume (mNpV) is also related to the total peripheral resistance, and it is the ratio of the peak-to-peak amplitude divided by its DC value [21]. Large Artery Stiffness Index (LASI) is an indicator of the stiffness of the arteries and it is inversely related to the time interval between S and IP [21], [31], [32]. While the aforementioned features were initially included in the analysis as intensity, time, or area-related features, the additional features of PPGK, mNpV,

---

Algorithm 1 Algorithm for fiducial points identification.

Input: wave
Output: $O, V, S, MD, D, IP, DN, a, b$
1: $d\_wave = \text{first\_derivative}(wave)$
2: $dd\_wave = \text{second\_derivative}(wave)$
3: $O = wave[\text{first point}]$
4: $V = wave[\text{last point}]$
5: $S = \text{maximum\_peak}(wave)$
6: $systolic\_phase = wave[\text{from } O \text{ to } S]$
7: $diastolic\_phase = wave[\text{from } S \text{ to } V]$
8: $MD = \text{maximum\_peak}(wave \text{ in } systolic\_phase)$
9: $search\_zone =[diastolic\_phase \text{ from 80 ms to } 0.6\times$
     $\text{duration}(diastolic\_phase)]$
10: $n = \text{number\_of\_local\_maxima}(wave \text{ in } search\_zone)$
11: **if** $n = 0$ **then**
12: 　　$D = \text{maximum\_peak}(d\_wave \text{ in } search\_zone)$
13: 　　$IP = D$
14: 　　$DN = \text{maximum\_peak}(dd\_wave \text{ right before } IP)$
15: **else if** $n = 1$ **then**
16: 　　$D = \text{maximum\_peak}(wave \text{ in } search\_zone)$
17: 　　$IP = \text{local\_maximum}(d\_wave \text{ right before } D)$
18: 　　$DN = \text{local\_minimum}(wave \text{ right before } IP)$
19: **else**
20: 　　$IP = \text{maximum\_peak}(d\_wave \text{ in } search\_zone)$
21: 　　$D = \text{local\_maximum}(wave \text{ right after } IP)$
22: 　　$DN = \text{local\_minimum}(wave \text{ right before } IP)$
23: **end if**
24: $a = \text{maximum\_peak}(dd\_wave \text{ in } systolic\_phase)$
25: $b = \text{minimum\_peak}(d\_wave \text{ in } systolic\_phase)$
26: **return** $O, V, S, MD, D, IP, DN, a, b$

---

and LASI were also incorporated into the analysis.

Finally, 17 more generic temporal features were computed based on [33]: autocorrelation, centroid, entropy calculated both using the Kernel Density Estimation (KDE) and the Gaussian function, number of minimum and maximum peaks, mean and median of differences, mean and median of absolute differences, the sum of the absolute differences between consecutive points, travelled distance, number of zero crossings of the first, second and third derivative and total energy as well as absolute energy.

2) Frequency-based Features (FT): Using the Fast Fourier Transform (FFT), the frequency and magnitude of the first, second, and third harmonics were extracted. Relative power [34] along with relative power at the first, second, and third harmonics were also included. Other 19 frequency-based features were based on [33]: spectral distance, fundamental frequency, maximum power spectrum density, maximum and median frequencies, spectral centroid, spectral decrease, spectral kurtosis, spectral skewness, spectral spread, spectral slope, spectral variation, spectral roll-off, spectral roll-on, number of maximum spectral peaks, human range energy ratio, power spectrum density bandwidth, and spectral and wavelet entropies. These features were derived from the beat-to-beat PPG pulse waveforms and PPG segments, which were constructed exclusively using validated PPG pulses (i.e., removing invalid

pulses from the original 30-second segment during the signal processing).

3) Statistical Features (SF): Previous studies used statistical features for BP estimation [8], [11], [19], [35]. In this work, for each pulse waveform, 14 statistical features were extracted: SKewness (SK), Kurtosis, Mean Absolute Value (MAV), median, Mean Absolute Deviation (MAD), Median Absolute Deviation, Root-Mean-Square (RMS), Standard Deviation (SD), Shape Factor (SF), Impulse Factor (IF), Crest Factor (CF), variance, interquartile range (IRQ) and perfusion.

4) Demographic Features (DF): Five demographic features were employed: age, weight, body mass index (BMI), rest systolic BP (SBP), and rest diastolic BP (DBP).

E. MB Feature Selection

Features that are used to train ML models have a high influence on their performance. Irrelevant or partially relevant features can negatively impact model performance. Appropriate feature selection can reduce overfitting, improve robustness, and shorten training time. Our goal here was to analyse the best transferability across sub-populations, therefore, we performed a wrapped selection over the entire dataset to provide subsequently a population feature analysis and performance comparison. As a first step in this process to reduce the computational cost, we used the minimum redundancy maximum relevance (MRMR) method, a filter-based approach that significantly reduces the number of features [36]. Then, we applied the wrapped method Predictive Permutation Feature Selection (PPFS) [37], a feature selection using the MB derived from a graphical model, which statistically determines the sufficient set of features for prediction across different subgroups. This potentially introduces a bias to the subsequent models nonetheless, we identify the best robust feature space to potentially transfer to ambulatory settings in future studies with smaller populations.

1) MB Discovery: MB Discovery refers to finding the set of variables that are sufficient for predicting the target variable based on a graphical model and d-separation. Under the assumption of causal sufficiency [38], MB consists of the parents, children, and spouses of the target node in the directed acyclic graph (as illustrated in Fig. 5b). We make use of the concept of graphical models where nodes and edges represent variables and statistical relations among them. Due to the compact statistical relations encoded and a clear structural representation, this framework has been developed and applied in various fields such as health science [39], computational biology [40], and earth science [41], facilitating a deeper understanding of complex systems. Fig. 5a shows a graphical model with 10 variables and its factorized distribution can be derived as below:

$$P(\mathbf{X}) = \prod_{i=1}^{10} p(X_i | \mathbf{X}_{pa(i)}) \qquad (1)$$

where $\mathbf{X}_{pa(i)}$ stands for the vector that includes all parental nodes of X_i in the graph. The factorization encodes sparsity in the structure via conditional Independence that can be further utilized for feature selection and robust inference.
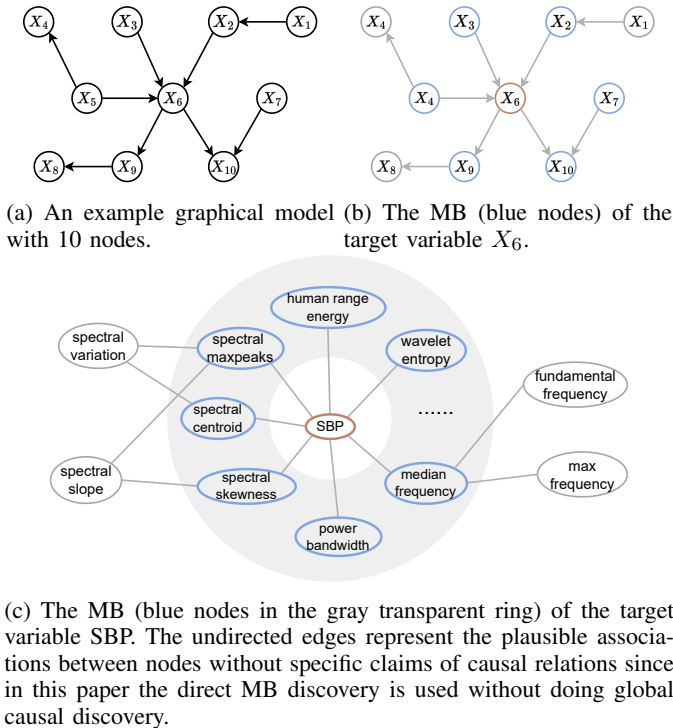
(a) An example graphical model with 10 nodes.

(b) The MB (blue nodes) of the target variable X_6.

(c) The MB (blue nodes in the gray transparent ring) of the target variable SBP. The undirected edges represent the plausible associations between nodes without specific claims of causal relations since in this paper the direct MB discovery is used without doing global causal discovery.

Fig. 5: Illustration of (a) graphical models; (b) Markov blanket (MB); (c) the potential application of using MB for SBP estimation.

According to d-separation [38], all nodes outside the MB will be conditionally independent of the target node given the MB such that those variables apart from the MB and the target would be redundant in predicting the target variable. By using the MB discovered, as shown in Fig. 5c where sparse structure and essential features are identified towards predicting SBP from generated PPG feature space, predictive models trained on the selected features are expected to learn sparse and essential relations from the optimal features to increase robustness against spurious correlations and covariate shifts in the generated feature space across different populations [42]. In this paper, we focus on approaches that can handle mixed-type features without distribution assumptions using non-parametric methods, leading to the adoption of the PPFS approach [37].

F. ML Models and Evaluation Metrics

Based on previous studies some of the best-performing ML models have been implemented to estimate BP from the selected features: Ridge Regression (RR), Gradient Boosting (GB), Decision Tree (DT), K-Neighbors (KNN), Linear Support Vector Regressor (L-SVR) and AdaBoost. Also, different scalers on each feature space were used MinMaxScaler, QuantileTransformer, Normalizer, StandarScaler and RobustScaler, details can be found in the open library VitalPy [16].

The performance of the ML models was evaluated using a nested cross-validation approach since it provides less optimistic and less biased estimation compared to the traditional train/test split method. Furthermore, it overcomes the

problem of over-fitting the training dataset. A 10-fold inner loop optimizes the model hyper-parameters. The outer loop used a leave-one-participant-out validation and provided the estimation error. The estimation error was calculated as the average of the individual model scores of every participant.

The assessment of the results was conducted using the evaluation metrics and visualizations recommended by the IEEE Std 1708-2014 standard [43], as well as ANSI/AAMI SP 10-1987 standard [44].

The IEEE standard recommends the use of scatter plots, such as the Bland-Altman plot, and BP change histograms to visualize the differences between the reference measurements and the measurements to be validated. In addition, the IEEE standard introduces the MAE as an alternative to the mean $\pm$ standard deviation (MD $\pm$ SD) proposed by the ANSI/AAMI SP and the cumulative percentage errors suggested by the British Hypertension Society (BHS) protocol [45].

The IEEE standard also introduces a new grading system based on the MAE accuracy level, which can be compared with the grading system proposed by the ANSI/AAMI SP10 (pass if MAE is less than 5 ± 8 mmHg, fail otherwise) and the BHS evaluation system. It assigns grade A for MAE less than 5 mmHg, grade B for MAE in the range of 5-6 mmHg, grade C for MAE between 6-7 mmHg, and grade D (fail) for MAE greater than 7 mmHg. Accordingly, in this paper, the performance charts include three lines at 5, 6, and 7 mmHg to delineate these grades.

Since the BP data were collected from different populations, we are particularly interested in finding how large the distance between each pair of populations can be. There exist extensive works inspecting distribution distances, e.g. KL-divergence and H-divergence, and how to utilize these metrics for further applications. In our work, we use a distance metric called Wasserstein distance to measure the distance of two empirical distributions [46]. Its formal mathematical definition is given with a hyperparameter p. The p-Wasserstein distance between probability measures μ and ν on $\mathbb{R}^d$ is defined as

$$W_p(\mu, \nu) = \inf_{X \sim \mu, Y \sim \nu} (\mathbb{E}||X - Y||^p)^{\frac{1}{p}}, \quad p \geq 1. \quad (2)$$

The distance metric from the optimal transport perspective is the minimum effort it would take to move mass points from one distribution to the other. This can be approximated using a numeric method called Sinkhorn iterations [47]. For a better focus of the paper, we skip the details of this algorithm and directly apply the method to compute the distance of each pair of sub-populations.

III. RESULTS

The performance of different ML algorithms for BP estimation with the selected set of features by the MB algorithm was firstly compared, with the aim proposing the most accurate one for subsequent evaluations. Considering training and testing BP estimation models on specific sub-populations can help improve accuracy and reliability, the cross-population stability of the selected feature space was also evaluated by training and testing several models specialized for different groups and a general model for all data. Additionally, the performance of
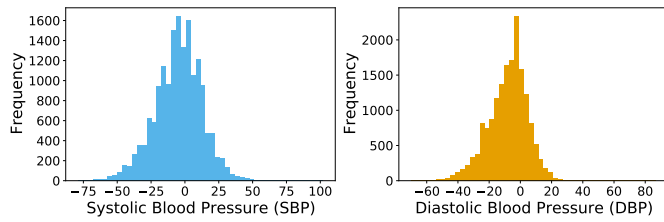
Fig. 6: Blood pressure distribution in the dataset per measurement from the individual resting point (mmHg)

(a) Features for SBP estimation   (b) Features for DBP estimation.

Fig. 7: Set of selected features categorized into the four main categories: time domain (TF), frequency-based (FF), statistical (SF), and demographic (DF). The time domain category is further divided into subgroups, including intensity (I), time (T), area (A), slope (SLP), branch width (BW), and others (OT).

A. Dataset

From the 548 participants of the database, only 534 have available in-clinic data regarding the BP in the rest position, especially sitting with the arm down, which is used as the reference BP. Each participant has a mean (SD) of 55.70 (4.11) measurements, split into 14.15 (0.68) and 41.55 (4.07) in-clinic and ambulatory measurements, respectively. After data pre-processing, only 523 participants remained with a mean of 23.92(11.36) measurements. The mean of the mean DBP and SBP at rest position are 86.21 (10.88) and 131.15 (15.53) mmHg respectively. The delta DBP and SBP is 40.72 (11.87) and 54.03 (16.25) mmHg, respectively.

To assess the performance of the BP estimation algorithm, it is essential to have a wide distribution of BP data relative to the resting position. According to the IEEE standard [43], the BP values from the calibration point should differ by at least 10 mmHg for DBP and 15 mmHg for SBP. The distribution of BP changes from rest position for every measurement is shown in Fig. 6.

B. Feature Selection and Machine Learning Models

The feature selection process resulted in a subset of 68 and 76 features for SBP and DBP estimation (Fig. 7), with 37 features at the intersection of the subsets. While no SFs were included in the selected features, four DFs were found to be significant for both DBP and SBP estimation. In fact, they ranked among the top 11 in terms of importance: age (5/68 for SBP; 26/76 for DBP), weight (8/68 for SBP; 11/76 for DBP), SBP at rest (7/68 for SBP; 9/76 DBP) and DBP at rest (6/68 for SBP; 10/76 DBP). Additionally, the common features for SBP and DBP estimation encompassed all six subgroups of TF and thirteen FF. An evaluation of the feature selection process robustness through MB compared to recursive feature elimination can be detailed in the appendix I.

Following the feature selection process, two datasets were created. The systolic BP dataset consists of 69 features and a reference BP value, and the DBP dataset comprises 77 features and a reference BP value. Various scalers were tested, and the MinMaxScaler yielded the highest accuracy. Table I presents the results for DBP estimation using each algorithm while applying the MinMaxScaler. Among the algorithms, GB achieved the highest accuracy exhibiting an MAE of 10.29 and 6.74 mmHg and an MD (SD) of 0.18 (13.82) and -0.22 (8.71)
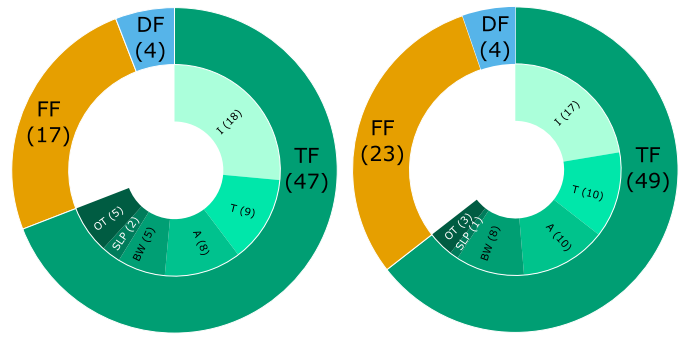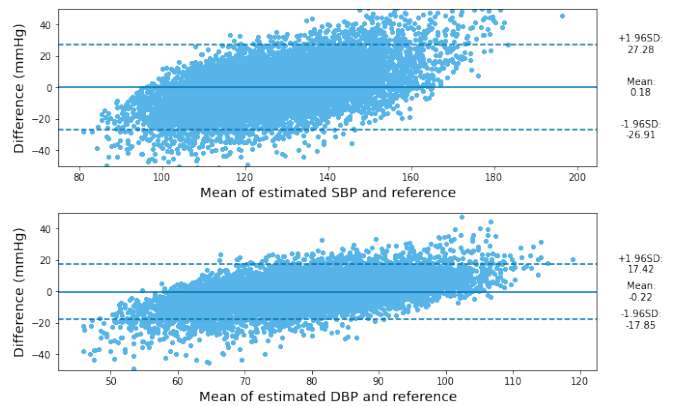

Fig. 8: Bland-Altman plot for SPB and DBP estimation showing the MD and SD of the reference BP measurements and the estimated BP values. The ANSI/AAMI SP 10 requires to have MD±SD less 5±8 mmHg.

for SBP and DBP, respectively (Fig. 8). From this point, only the performance of the GB model is reported.

C. Results for Sub-populations

According to the ACC/AHA guidelines [48], participants can be classified into four BP categories based on the rest DBP/SBP measurements (Fig. 9a). These categories included normal, elevated, hypertension stage 1 (HTN S1), and hypertension stage 2 (HTN S2).

One generic model was developed using the information coming from all the participants, while various specialized models are trained and tested with each sub-population based on BP category. The MAE of the generic model along with the MAE of the specialized models based on BP are presented in Table II. The MAE of the generic model is subdivided according to each subgroup, and the MAE of all four individual models is combined. Additionally, the number and percentage of available measurements for each sub-population, including the mean and standard deviation for each participant is also

TABLE I: Performance of best machine learning algorithms for SBP and DBP regression.

	SBP			DBP		
	MD (SD)	MAE	MAPE	MD (SD)	MAE	MAPE
RR	-0.24 (14.14)	10.79	8.72	0.00 (9.48)	7.21	9.79
GB	0.18 (13.82)	10.29	8.44	-0.22 (8.71)	6.74	9.39
AdaBoost	-0.17 (13.97)	10.57	8.51	0.14 (9.22)	6.96	9.42
DT	0.45 (15.54)	11.68	8.72	-0.14 (10.44)	7.87	10.65
KNN	0.88 (15.35)	11.75	8.72	0.13 (9.85)	7.52	10.17
LSVR	0.22 (14.28)	10.81	8.72	-0.35 (9.69)	7.32	10.02

TABLE II: Performance, expressed as MAE (SD), of one general model and specialized models based on BP category.

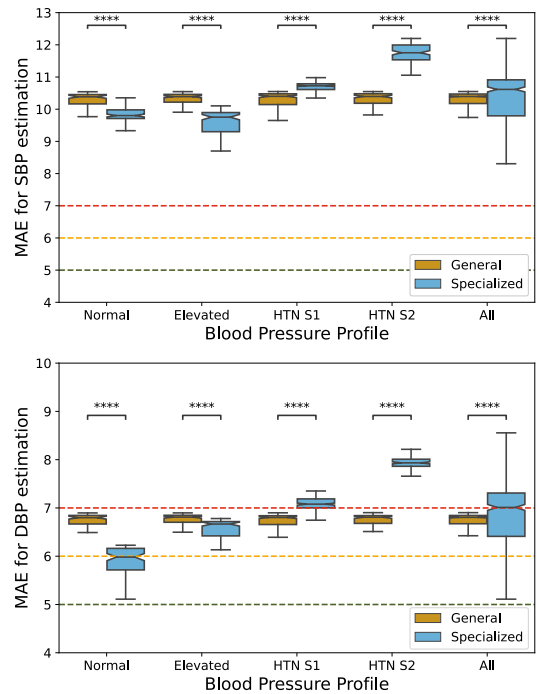|	Available measurements			MAE (SD) - SBP		MAE (SD) - DBP	
	Total	Percentage	Mean (SD)	Individual	Generic	Individual	Generic
Normal	2057	16,4%	23,38 (11.55)	9.82 (0.29)	$10.27 (0.29)^2$	5.91 (0.34)	$6.71 (0.22)^2$
Elevated	3082	24,6%	24,66 (11.45)	9.63 (0.41)	$10.32 (0.24)^2$	6.58 (0.19)	$6.76 (0.14)^2$
HTN S1	4359	34,8%	23,19 (11.42)	10.58 (0.50)	$10.28 (0.32)^2$	7.01 (0.29)	$6.73 (0.18)^2$
HTN S2	3012	24,1%	24,69 (12.22)	11.77 (0.47)	$10.31 (0.24)^2$	7.95 (0.19)	$6.74 (0.17)^2$
All	12510	100,0%	23,92 (11.36)	$10.50 (0.91)^1$	10.29 (0.28)	$6.94 (0.72)^1$	6.74 (0.18)

^{1} Indicates combined results from individual models. ^{2} Indicates split results from the generic model by BP group.
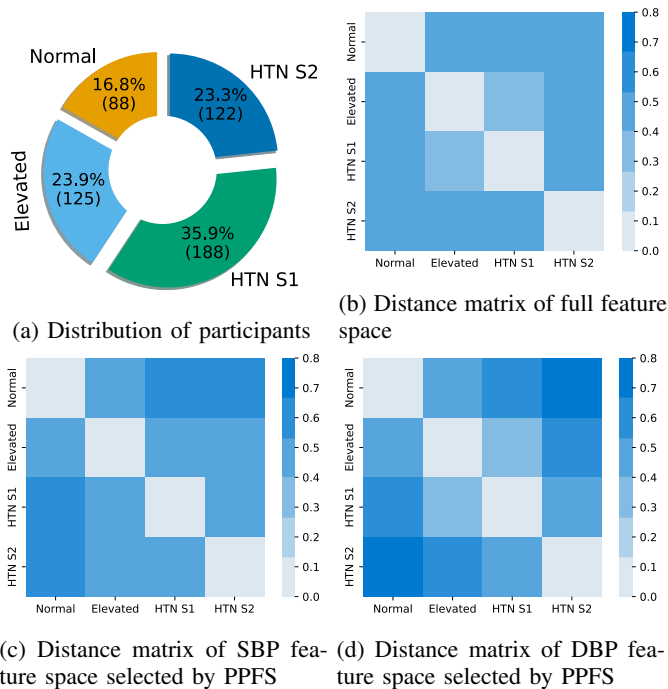
Fig. 9: (a) Distribution of participants' BP profile. (b) Normalized distance matrix for the full feature space (after standardization) across different sub-populations by BP profile measured by the corresponding sub-population pair using the Sinkhorn algorithm [47] normalized to the number of dimensions. (c) Normalized distance matrix of selected SBP feature space by PPFS. (d) Normalized distance matrix of selected DBP feature space by PPFS.

(a) Distribution of participants

(b) Distance matrix of full feature space

(c) Distance matrix of SBP feature space selected by PPFS

(d) Distance matrix of DBP feature space selected by PPFS

Fig. 10: Performance evaluation of the general model and specialized models for BP estimation across sub-populations (normal, elevated, HTN S1, and HTN S2) as well as all data. ∗ denotes a significant difference by a paired t-test at the level $p < 0.0001$.

provided in Table II.

The overall BP estimation accuracy of the generic and BP-based individual models are compared in Fig. 10. The generic model had a lower MAE and SD when compared to the combined BP estimation accuracy of the group-specific models. Additionally, the p-value at the level of 0.0001 from a paired t-test showed that there were significant differences between the generic and individualized models for all four

BP subgroups (normal, elevated, HTN S1 and HTN S2) and all participants. While the specialized models provided better results for the normal and elevated subgroups, the results were worse for HTN S1 and S2 subgroups.

Additionally, the results show that the MAE of the generic model remains relatively consistent MAE across all sub-populations in contrast to specialized models (Fig. 11). Specifically, the ANOVA revealed that the dependent variable (BP profile) had no significant impact on the results of the generic model for estimating both SBP ($p - level > 0.1$, F value =
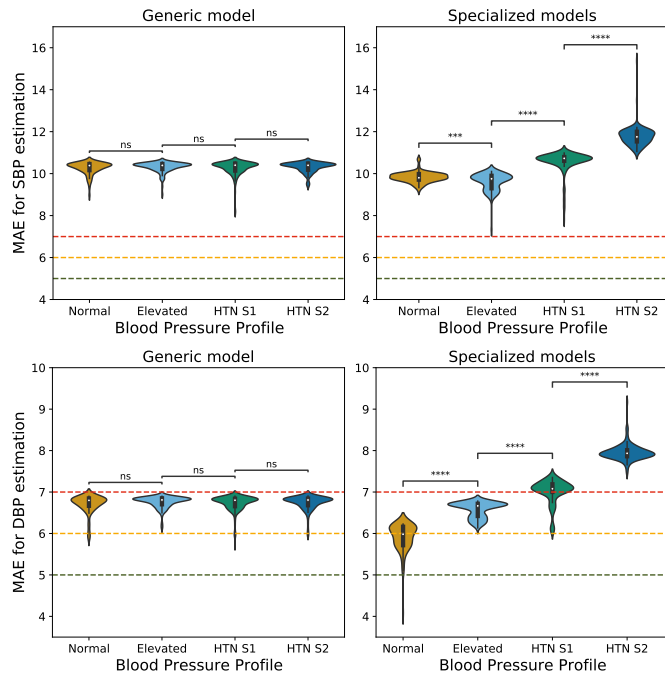
Fig. 11: Performance evaluation of the general model and specialized models for BP estimation evaluated across different BP categories, including normal, elevated, HTN S1 and HTN S2. $*$ denotes significant difference by a paired t-test at the level $p < 0.0001$.
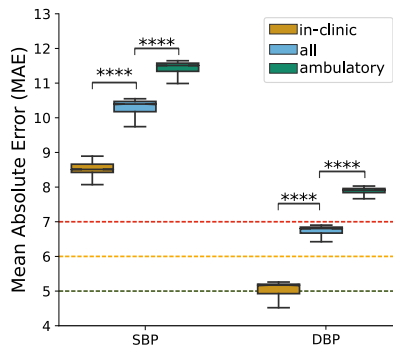

Fig. 12: Performance of the BP estimation algorithm using in-clinic, ambulatory, and all measurements. $*$ denotes a significant difference by a paired t-test at the level $p < 0.0001$.

0.77) and DBP ($p - level > 0.1$, F value = 1.76). Conversely, the BP profile exerted a significant influence on the outcomes of their corresponding individual models for SBP ($p - level < 0.0001$, F value = 585.57) and DBP estimation ($p - level < 0.0001$, F value = 1188.12).

D. Ambulatory vs In-clinic Data

The database was separated into two subsets: in-clinic and ambulatory measurements. Two GB models were trained and tested with each dataset separately after the feature selection process. The number of selected features for SBP estimation is 61 and 49 for ambulatory and in-clinic, while for DBP estimation is 62 and 46, respectively. The MAE (SD) of SBP

and DBP are 8.49 (0.25) mmHg and 5.03 (0.25) mmHg for in-clinic measurements and 11.35 (0.44) mmHg and 7.85 (0.25) mmHg for ambulatory measurements (Fig. 12). Statistically significant differences were found between ambulatory and in-clinic data for the performance on both SBP ($p - level < 0.0001$, F-value = 32764.76) and DBP ($p - level < 0.0001$, F-value 65675.36) and estimation algorithms.

IV. DISCUSSION AND CONCLUSIONS

A. Feature Selection

The results coming from the feature selection process indicate that more than half of the selected features belong to the TF category, including all six subcategories, similar to [8], [11] which also found TF of great importance for BP estimation. The results also indicate the importance of DF features, in line with findings from other studies [11], [49], although BMI was not selected.

Noteworthy features were identified within the FF category as well. Specifically, the wavelet entropy was selected for both SBP and DBP estimation. These findings contradict the results in [19], where a set of 35 features spanning TF, FF, and SF was examined, and although the approximate entropy was ranked among the top 15 features, while no features in the wavelet domain were identified as influential.

Additionally, we found that the SF offer relatively less informative value compared to other feature types and hence, none of them were selected. These findings align with previous works [8], [11].

Prior studies have focused on selecting the optimal features for BP estimation, yet many of them proposed a limited range of feature types. For instance, [9], [17], [50] proposed feature sets confined to TF, with 24, 42 and 65 features. These restricted set overlooked valuable information encoded in other feature domains. Similarly, [11] proposed 74 features encompassing TF, SF and DF features, while disregarding FT. Likewise, [19] proposed 34 features in the TF, FF and SF domains, omitting DF. In contrast, our work offers a comprehensive set of features belonging to four main categories (TF, FF, ST and DF), ensuring a more robust feature selection process through the Markov Blanket.

B. Domain Transfer

The study conducted by Miao et al. [51] emphasized the substantial variation in feature importance across individuals. Their study demonstrated the subject-specific nature of feature selection, as the optimal feature set differed among subjects, highlighting the importance of individualized approaches for accurate BP estimation. However, in practical applications, subject-specific feature selection may not be feasible due to the need for extensive individual data for model pre-training. In this work, we propose the adoption of an MB-based feature selection method to obtain a reliable and consistent feature set across individuals, enabling robust BP estimation without the requirement for subject-specific customization.

As shown in Fig. 9b-9d, compared with the normalized distance matrix of full feature space, the normalized distances of selected feature spaces for estimation of SBP and DBP

appear more salient, which means using PPFS can identify the set of features that are more informative and differentiated across sub-populations so as to enable the consequent generic modeling to capture those individualized distinctions.

The findings of this study reflected that the generic algorithm exhibited significantly superior estimation accuracy compared to the overall estimation accuracy of the specialized models for both SBP and DBP. Moreover, the BP profile did not have a significant impact on the accuracy of DBP/SBP estimation when using the generic model.

In contrast, Khalid et al. [7] proposed a two-step algorithm that involved the classification of BP into three categories (hypotensive, normotensive, and hypertensive), followed by the application of a specific BP estimation algorithm corresponding to the classified BP category. The two-step algorithm was found to outperform the generic algorithm, and statistically significant differences were observed for SBP and DBP estimation. Additionally, the BP profile significantly influenced the DBP and SBP estimation accuracy using either two models. Similarly, [52] found that the estimation errors were significantly higher in HTN subjects compared to normotensive individuals.

The use of feature selection based on the MB allowed for the selection of a stable set of features applicable across all populations, enabling the use of a single model for the entire population with improved performance. Moreover, the feature space selected by MB should have better transferability in terms of prediction accuracy and robustness. Indeed, the outcomes of prediction (Fig. 13) demonstrated the superiority of the PPFS method based on the MB criteria over the RFE, indicating a more effective generalization of the approach. This has significant implications for BP estimation in small samples of populations, such as pregnant women, neonates and SCI individuals, where there may be insufficient data to train large deep learning models.

C. Ambulatory vs In-clinic data

Performance validation for ambulatory measurement showed that the estimation deviated more from the reference BP compared to in-clinic, with the MAE exceeding 8 mmHg for DBP and SBP. Previous research has pointed out the inadequacy of those methods to follow the dynamic changes in BP elicited by the cardiovascular autonomic nervous activities [52]. However, it is important to remark that in-clinic data included not only rest measurements (supine, sitting arm down/lap/up) but also dynamic BP changes induced by various cardiac maneuvers (walking, cooling down, and running).

It can be therefore concluded that BP estimation accuracy from in-clinic measurements is superior to the 24-hour recording of ambulatory measurements as ambulatory measurements are more prone to noise. BP estimation models are highly influenced not only by biological characteristics (respiration, vasomotor activity, thermoregulation) but also by external factors, such as motion or light [14]. Therefore, while ambulatory measurements could provide a comprehensive picture of a person's BP levels, in-clinic measurements are far more accurate and reliable.

D. Conclusion

Our work is the first in its kind using both ambulatory and inpatient data and aiming to identify relevant transnational features for robust estimation of BP across populations. Therefore, comparing it to previous work is not straight forward due to difference in the number of subjects [53], data and population.

The only work using Aurora-BP dataset is the one from Mieloszyk et al. [15], where a BP estimation algorithm using different physiological signals, included ECG, PPG, and tonometry. Their results from PPG signal achieved an SBP and DBP estimation accuracy of 0.42 (8.98) and 0.54 (5.95) mmHg, respectively. In contrast, our work showed lower accuracy with values of 0.48 (13.93) and -0.22 (8.71) mmHg for SBP and DBP estimation, respectively. Nonetheless, their study only used an average of 19.3 (10.1) measurement of a total of 227 participants, without providing specific details regarding the resulting BP distribution. In contrast, our study incorporated data from 523 participants, with a mean of 23.92 (11.36) measurements per participant and the BP distribution in Fig. 9. These discrepancies highlight the influence of data processing and cleaning techniques on the accuracy of results.

Moreover, in [15] incorporated three features that were left out in this work: acceleration, which was omitted to only employ one signal modality, and the sin and cosine of the time of the day, which introduces a bias by assuming a uniform sleep pattern for all subjects.

As we observed, BP estimation performance is highly influenced by the characteristics of the dataset used, which has been observed in several works where the same BP estimation algorithm applied to different datasets yielded substantial differences in accuracy [4], [8], [10], resulting from a number of factors, such as differences in the population characteristics, data acquisition protocols, signal quality, and noise levels present in the datasets. On our work we measured the significant deviation of the model accuracy when using in-clinic data in contrast to ambulatory data, even though the population characteristics and acquisition devices were the same.

Nevertheless, it is crucial for subsequent research to prioritize the development of BP estimation algorithms that are robust across diverse populations and reliable enough to handle suboptimal conditions in PPG signals. Therefore, despite our comparatively lower accuracy results, we believe that our findings provide a more realistic representation of a BP estimation algorithm for ambulatory BP monitoring, aligning with the future goal of achieving continuous ambulatory BP monitoring.

Code Availability

The code for preprocessing and feature extraction generated during this study is available under the GPLv3 License on GitHub [16].

APPENDIX I
FEATURE SELECTION ROBUSTNESS

We evaluated the robustness of the MB-based feature selection method (PFFS) and compared it with the baseline of recursive feature elimination (RFE) an algorithm commonly used in ML to identify the most relevant features
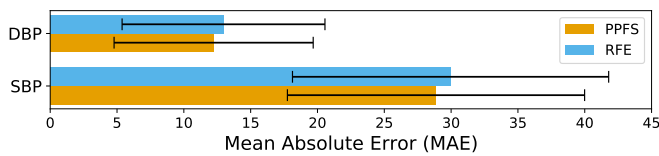
Fig. 13: Performance comparison between PPFS and RFE method in the task of domain generalization: both feature selection methods were tested using a cross-population validation strategy where the prediction results are obtained using the selected feature space with ML model finetuning.

for a given variable [54]. It is an iterative process that aims to determine the optimal subset of features by eliminating less essential or redundant features based on the ranking such as feature importance or model-specific coefficient. To this end, the GB model was trained on the normal, elevated and HTN S1 populations, and tested on the HTN S2 population. The model developed using the PPFS method exhibited superior generalization compared to the model relying on the RFE selection method, as evidenced by the results for both SBP and DBP estimation (Fig. 13).

REFERENCES

[1] M. Desai, N. Stockbridge, and R. Temple, "Blood pressure as an example of a biomarker that functions as a surrogate," *The AAPS Journal*, vol. 8, no. 1, pp. E146–E152, Mar. 2006. [Online]. Available: http://link.springer.com/10.1208/aapsj080117

[2] S. M. S. Islam, C. K. Chow, R. Daryabeygikhotbehsara, N. Subedi, J. Rawstorn, T. Tegegne, C. Karmakar, M. U. Siddiqui, G. Lambert, and R. Maddison, "Wearable cuffless blood pressure monitoring devices: a systematic review and meta-analysis," *European Heart Journal - Digital Health*, vol. 3, no. 2, pp. 323–337, Jul. 2022. [Online]. Available: https://academic.oup.com/ehjdh/article/3/2/323/6576559

[3] A. Esmaili, M. Kachuee, and M. Shabany, "Nonlinear Cuffless Blood Pressure Estimation of Healthy Subjects Using Pulse Transit Time and Arrival Time," *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 12, pp. 3299–3308, Dec. 2017. [Online]. Available: http://ieeexplore.ieee.org/document/8032000/

[4] V. Figini, S. Galici, D. Russo, I. Centonze, M. Visintin, and G. Pagana, "Improving Cuff-Less Continuous Blood Pressure Estimation with Linear Regression Analysis," *Electronics*, vol. 11, no. 9, p. 1442, Apr. 2022. [Online]. Available: https://www.mdpi.com/2079-9292/11/9/1442

[5] R. Byfield, M. Miller, J. Miles, G. Guidoboni, and J. Lin, "Towards Robust Blood Pressure Estimation From Pulse Wave Velocity Measured by Photoplethysmography Sensors," *IEEE Sensors Journal*, vol. 22, no. 3, pp. 2475–2483, Feb. 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9646921/

[6] S. Haddad, A. Boukhayma, and A. Caizzone, "Continuous PPG-Based Blood Pressure Monitoring Using Multi-Linear Regression," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 5, pp. 2096–2105, May 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9616476/

[7] S. G. Khalid, H. Liu, T. Zia, J. Zhang, F. Chen, and D. Zheng, "Cuffless Blood Pressure Estimation Using Single Channel Photoplethysmography: A Two-Step Method," *IEEE Access*, vol. 8, pp. 58 146–58 154, 2020. [Online]. Available: https://ieeexplore.ieee.org/document/9042315/

[8] S. Maqsood, S. Xu, M. Springer, and R. Mohawesh, "A Benchmark Study of Machine Learning for Analysis of Signal Feature Extraction Techniques for Blood Pressure Estimation Using Photoplethysmography (PPG)," *IEEE Access*, vol. 9, pp. 138 817–138 833, 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9558767/

[9] 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, p.

5668, Oct. 2020. [Online]. Available: https://www.mdpi.com/1424-8220/20/19/5668

[10] Z. Qiu, D. Chen, B. W.-K. Ling, Q. Liu, and W. Li, "Joint Regression Network and Window Function-Based Piecewise Neural Network for Cuffless Continuous Blood Pressure Estimation Only Using Single Photoplethesmogram," *IEEE Transactions on Consumer Electronics*, vol. 68, no. 3, pp. 236–260, Aug. 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9774278/

[11] P. Yao, N. Xue, S. Yin, C. You, Y. Guo, Y. Shi, T. Liu, L. Yao, J. Zhou, J. Sun, C. Dong, C. Liu, and M. Zhao, "Multi-Dimensional Feature Combination Method for Continuous Blood Pressure Measurement Based on Wrist PPG Sensor," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 8, pp. 3708–3719, Aug. 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9756927/

[12] T. R. Dawber, H. E. Thomas, and P. M. McNamara, "Characteristics of the Dicrotic Notch of the Arterial Pulse Wave in Coronary Heart Disease," *Angiology*, vol. 24, no. 4, pp. 244–255, Apr. 1973. [Online]. Available: http://journals.sagepub.com/doi/10.1177/000331977302400407

[13] M. A. Gamrah, J. Xu, A. El Sawy, H. Aguib, M. Yacoub, and K. H. Parker, "Mechanics of the dicrotic notch: An acceleration hypothesis," *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, vol. 234, no. 11, pp. 1253–1259, Nov. 2020. [Online]. Available: http://journals.sagepub.com/doi/10.1177/0954411920921628

[14] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiological Measurement*, vol. 28, no. 3, pp. R1–R39, Mar. 2007. [Online]. Available: https://iopscience.iop.org/article/10.1088/0967-3334/28/3/R01

[15] R. Mieloszyk, H. Twede, J. Lester, J. Wander, S. Basu, G. Cohn, G. Smith, D. Morris, S. Gupta, D. Tan, N. Villar, M. Wolf, S. Malladi, M. Mickelson, L. Ryan, L. Kim, J. Kepple, S. Kirchner, E. Wampler, R. Terada, J. Robinson, R. Paulsen, and T. S. Saponas, "A Comparison of Wearable Tonometry, Photoplethysmography, and Electrocardiography for Cuffless Measurement of Blood Pressure in an Ambulatory Setting," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 7, pp. 2864–2875, Jul. 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9721156/

[16] A. Cisnal, B. Fuchs, and D. Paez-Granados, "VitalPy: A Vital Signal Analysis Python Package," Jul. 2023. [Online]. Available: https://github.com/SCAI-Lab/VitalPy

[17] S. Yang, J. Sohn, S. Lee, J. Lee, and H. C. Kim, "Estimation and Validation of Arterial Blood Pressure Using Photoplethysmogram Morphology Features in Conjunction With Pulse Arrival Time in Large Open Databases," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 4, pp. 1018–1030, Apr. 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9142317/

[18] B. De Marchi, M. Frigerio, S. De Nadai, G. Longinotti-Buitoni, and A. Aliverti, "Blood Pressure Continuous Measurement through a Wearable Device: Development and Validation of a Cuffless Method," *Sensors*, vol. 21, no. 21, p. 7334, Nov. 2021. [Online]. Available: https://www.mdpi.com/1424-8220/21/21/7334

[19] F. Miao, Z.-D. Liu, J.-K. Liu, B. Wen, Q.-Y. He, and Y. Li, "Multi-Sensor Fusion Approach for Cuff-Less Blood Pressure Measurement," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 1, pp. 79–91, Jan. 2020. [Online]. Available: https://ieeexplore.ieee.org/document/8667638/

[20] Z.-D. Liu, J.-K. Liu, B. Wen, Q.-Y. He, Y. Li, and F. Miao, "Cuffless Blood Pressure Estimation Using Pressure Pulse Wave Signals," *Sensors*, vol. 18, no. 12, p. 4227, Dec. 2018. [Online]. Available: http://www.mdpi.com/1424-8220/18/12/4227

[21] N. Hasanzadeh, M. M. Ahmadi, and H. Mohammadzade, "Blood Pressure Estimation Using Photoplethysmogram Signal and Its Morphological Features," *IEEE Sensors Journal*, vol. 20, no. 8, pp. 4300–4310, Apr. 2020. [Online]. Available: https://ieeexplore.ieee.org/document/8938751/

[22] O. Vardoulis, T. S. Saponas, D. Morris, N. Villar, G. Smith, S. Patel, and D. Tan, "In vivo evaluation of a novel, wrist-mounted arterial pressure sensing device versus the traditional hand-held tonometer," *Medical Engineering & Physics*, vol. 38, no. 10, pp. 1063–1069, Oct. 2016. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1350453316301461

[23] J. Park, H. S. Seok, S.-S. Kim, and H. Shin, "Photoplethysmogram Analysis and Applications: An Integrative Review," *Frontiers in Physiology*, vol. 12, p. 808451, Mar. 2022. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fphys.2021.808451/full

[24] Z.-M. Zhang, S. Chen, and Y.-Z. Liang, "Baseline correction using adaptive iteratively reweighted penalized least squares," *The Analyst*, vol. 135, no. 5, p. 1138, 2010. [Online]. Available: http://xlink.rsc.org/?DOI=b922045c

[25] J. A. Sukor, S. J. Redmond, and N. H. Lovell, "Signal quality measures for pulse oximetry through waveform morphology analysis," *Physiological Measurement*, vol. 32, no. 3, pp. 369–384, Mar. 2011. [Online]. Available: https://iopscience.iop.org/article/10.1088/0967-3334/32/3/008

[26] C. Orphanidou, T. Bonnici, P. Charlton, D. Clifton, D. Vallance, and L. Tarassenko, "Signal Quality Indices for the Electrocardiogram and Photoplethysmogram: Derivation and Applications to Wireless Monitoring," *IEEE Journal of Biomedical and Health Informatics*, pp. 1–1, 2014. [Online]. Available: http://ieeexplore.ieee.org/document/6862843/

[27] M. J. Oppenheim and D. F. Sittig, "An Innovative Dicrotic Notch Detection Algorithm Which Combines Rule-Based Logic with Digital Signal Processing Techniques," *Computers and Biomedical Research*, vol. 28, no. 2, pp. 154–170, Apr. 1995. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0010480985710117

[28] A. H. Hafifah and N. Nazrul Anuar, "Methods of Extracting Feature from Photoplethysmogram Waveform for Non-Invasive Diagnostic Applications," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 16, no. 09, p. 39, Aug. 2020. [Online]. Available: https://online-journals.org/index.php/i-joe/article/view/13577

[29] Y. K. Qawqzeh, A. M. A. Mohd, M. Reaz, and O. Maskon, "Photoplethysmography analysis of artery properties in patients presenting with established erectile dysfunction," in *2010 2nd International Conference on Electronic Computer Technology*, May 2010, pp. 165–168.

[30] B. Liu, Z. Zhang, X. Di, X. Wang, L. Xie, W. Xie, and J. Zhang, "The Assessment of Autonomic Nervous System Activity Based on Photoplethysmography in Healthy Young Men," *Frontiers in Physiology*, vol. 12, p. 733264, Sep. 2021. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fphys.2021.733264/full

[31] M. Kachuee, M. M. Kiani, H. Mohammadzade, and M. Shabany, "Cuffless Blood Pressure Estimation Algorithms for Continuous Health-Care Monitoring," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 4, pp. 859–869, Apr. 2017. [Online]. Available: http://ieeexplore.ieee.org/document/7491263/

[32] P. Li and T.-M. Laleg-Kirati, "Central Blood Pressure Estimation From Distal PPG Measurement Using Semiclassical Signal Analysis Features," *IEEE Access*, vol. 9, pp. 44 963–44 973, 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9374974/

[33] M. Barandas, D. Folgado, L. Fernandes, S. Santos, M. Abreu, P. Bota, H. Liu, T. Schultz, and H. Gamboa, "TSFEL: Time Series Feature Extraction Library," *SoftwareX*, vol. 11, p. 100456, Jan. 2020. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S2352711020300017

[34] M. Elgendi, "Optimal Signal Quality Index for Photoplethysmogram Signals," *Bioengineering*, vol. 3, no. 4, p. 21, Sep. 2016. [Online]. Available: http://www.mdpi.com/2306-5354/3/4/21

[35] Q. Hu, X. Deng, A. Wang, and C. Yang, "A novel method for continuous blood pressure estimation based on a single-channel photoplethysmogram signal," *Physiological Measurement*, vol. 41, no. 12, p. 125009, Dec. 2020. [Online]. Available: https://iopscience.iop.org/article/10.1088/1361-6579/abc8dd

[36] Hanchuan Peng, Fuhui Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005. [Online]. Available: http://ieeexplore.ieee.org/document/1453511/

[37] A. Hassan, J. H. Paik, S. Khare, and S. A. Hassan, "PPFS: Predictive Permutation Feature Selection," 2021. [Online]. Available: https://arxiv.org/abs/2110.10713

[38] C. Glymour, K. Zhang, and P. Spirtes, "Review of Causal Discovery Methods Based on Graphical Models," *Frontiers in Genetics*, vol. 10, p. 524, Jun. 2019. [Online]. Available: https://www.frontiersin.org/article/10.3389/fgene.2019.00524/full

[39] M. B. Sesen, A. E. Nicholson, R. Banares-Alcantara, T. Kadir, and M. Brady, "Bayesian Networks for Clinical Decision Support in Lung Cancer Care," *PLoS ONE*, vol. 8, no. 12, p. e82349, Dec. 2013. [Online]. Available: https://dx.plos.org/10.1371/journal.pone.0082349

[40] N. Friedman, "Inferring Cellular Networks Using Probabilistic Graphical Models," *Science*, vol. 303, no. 5659, pp. 799–805, Feb. 2004. [Online]. Available: https://www.science.org/doi/10.1126/science.1094068

[41] I. Ebert-Uphoff and Y. Deng, "Causal Discovery for Climate Research Using Graphical Models," *Journal of Climate*, vol. 25, no. 17, pp. 5648–5665, Sep. 2012. [Online]. Available: http://journals.ametsoc.org/doi/10.1175/JCLI-D-11-00387.1

[42] C. F. Aliferis, I. Tsamardinos, and A. Statnikov, "HITON: a novel Markov Blanket algorithm for optimal variable selection," *AMIA ... Annual Symposium proceedings. AMIA Symposium*, vol. 2003, pp. 21–25, 2003.

[43] "IEEE Standard for Wearable Cuffless Blood Pressure Measuring Devices," IEEE, Tech. Rep. [Online]. Available: http://ieeexplore.ieee.org/document/6882122/

[44] A. for the Advancement of Medical Instrumentation, "American National Standards for Electronic or Automated Sphygmomanometers," *Association for the Advancement of Medical Instrumentation*, 1987. [Online]. Available: https://cir.nii.ac.jp/crid/1572261550560818560

[45] E. O'Brien, J. Petrie, W. Little, M. de Swiet, P. L. Padfield, D. G. Altma, M. Bland, A. Coats, and N. Atkins, "Short report: An outline of the revised British Hypertension Society protocol for the evaluation of blood pressure measuring devices:," *Journal of Hypertension*, vol. 11, no. 6, pp. 677–679, Jun. 1993. [Online]. Available: http://journals.lww.com/00004872-199306000-00013

[46] V. M. Panaretos and Y. Zemel, "Statistical aspects of Wasserstein distances," *Annual Review of Statistics and Its Application*, vol. 6, pp. 405–431, Mar. 2019.

[47] M. Cuturi, "Sinkhorn Distances: Lightspeed Computation of Optimal Transport," in *Sinkhorn Distances: Lightspeed Computation of Optimal Transport*, 2013.

[48] P. K. Whelton, R. M. Carey, W. S. Aronow, D. E. Casey, K. J. Collins, C. Dennison Himmelfarb, S. M. DePalma, S. Gidding, K. A. Jamerson, D. W. Jones, E. J. MacLaughlin, P. Muntner, B. Ovbiagele, S. C. Smith, C. C. Spencer, R. S. Stafford, S. J. Taler, R. J. Thomas, K. A. Williams, J. D. Williamson, and J. T. Wright, "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, Jun. 2018. [Online]. Available: https://www.ahajournals.org/doi/10.1161/HYP.0000000000000065

[49] D. Wu, L. Xu, R. Zhang, H. Zhang, L. Ren, and Y.-T. Zhang, "Continuous Cuff-Less Blood Pressure Estimation Based on Combined Information Using Deep Learning Approach," *Journal of Medical Imaging and Health Informatics*, vol. 8, no. 6, pp. 1290–1299, Aug. 2018. [Online]. Available: http://www.ingentaconnect.com/content/10.1166/jmihi.2018.2474

[50] M. M. R. Khan Mamun and A. T. Alouani, "Cuffless Blood Pressure Measurement Using Linear and Nonlinear Optimized Feature Selection," *Diagnostics*, vol. 12, no. 2, p. 408, Feb. 2022. [Online]. Available: https://www.mdpi.com/2075-4418/12/2/408

[51] F. Miao, N. Fu, Y.-T. Zhang, X.-R. Ding, X. Hong, Q. He, and Y. Li, "A Novel Continuous Blood Pressure Estimation Approach Based on Data Mining Techniques," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 6, pp. 1730–1740, Nov. 2017. [Online]. Available: https://ieeexplore.ieee.org/document/7914684/

[52] X. Ding, B. P. Yan, Y.-T. Zhang, J. Liu, N. Zhao, and H. K. Tsang, "Pulse Transit Time Based Continuous Cuffless Blood Pressure Estimation: A New Extension and A Comprehensive Evaluation," *Scientific Reports*, vol. 7, no. 1, p. 11554, Sep. 2017. [Online]. Available: https://www.nature.com/articles/s41598-017-11507-3

[53] D.-K. Kim, Y.-T. Kim, H. Kim, and D.-J. Kim, "DeepCNAP: A Deep Learning Approach for Continuous Noninvasive Arterial Blood Pressure Monitoring Using Photoplethysmography," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 8, pp. 3697–3707, Aug. 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9769903/

[54] P. M. Granitto, C. Furlanello, F. Biasioli, and F. Gasperi, "Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products," *Chemometrics and Intelligent Laboratory Systems*, vol. 83, no. 2, pp. 83–90, Sep. 2006. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0169743906000232