

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

Ana Cisnal^{ID}, Graduate Student Member, IEEE, Yanke Li^{ID}, Bertram Fuchs^{ID}, Mehdi Ejtehad^{ID}, Robert Riener^{ID}, Senior Member, IEEE, and Diego Paez-Granados^{ID}, Member, IEEE

Abstract—Current blood pressure (BP) estimation methods have not achieved an accurate and adaptable approach for ambulatory diagnosis and monitoring applications of populations at risk of cardiovascular disease, generally due to a limited sample size. This paper introduces an algorithm for BP estimation solely reliant on photoplethysmography (PPG) signals and demographic features. It 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. This approach was validated with the Aurora-BP database, compromising ambulatory wearable cuffless BP measurements for over 500 individuals. After evaluating several machine-learning regression methods, Gradient Boosting emerged as the most effective. According to the MB feature selection, temporal, frequency, and demographic features ranked highest in importance, while statistical ones were deemed non-significant. A comparative assessment of a generic model (trained on unclassified BP data) and specialized models (tailored to each distinct BP population), demonstrated a consistent superiority of our proposed MB feature space with a mean absolute error of 10.2 mmHg(0.28) for systolic BP and 6.7 mmHg(0.18) for diastolic BP on the whole dataset. Moreover, we present a first comparison of in-clinic vs. ambulatory models, with performance significantly lower for the latter with a drop

of 2.85 mmHg in systolic ($p < 0.0001$) and 2.82 mmHg for diastolic ($p < 0.0001$) estimation errors. This work contributes to the resilient understanding of BP estimation algorithms from PPG signals, providing causal features in the signal and quantifying the disparities between ambulatory and in-clinic measurements.

Index Terms—Cuffless blood pressure, photoplethysmography, pulse wave analysis.

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 [1], and a direct biomarker in conditions such as hypertension. Traditionally, BP measurements have been obtained in clinical settings, offering valuable insights into cardiovascular health. However, the increasing recognition of the dynamic nature of BP calls for ambulatory measurements to capture 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 being the most widespread [2]. PPG is a non-invasive, non-occlusive, optical technique for measuring volumetric changes in the microvascular bed tissue, correlating with BP pulse wave propagation in arteries [3]. Models based on the pulse arrival time (PAT) [4], pulse transit time (PTT) [5], and pulse wave velocity (PWV) [6] are most commonly used. However, they require a PPG signal, either combined with an ECG signal or another PPG signal from a different peripheral site.

An alternative is developing models that rely on a single 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 multi-linear regression [7], regression trees [8], random forest [9], support vector machine [9] and artificial neural network [10], [11], [12]. However, PPG signals exhibit different morphologies, with Dawber's classification [13] delineating four PPG classes based on the diastolic phase (Fig. 1). The precise formation and location of the dicrotic notch (DN) remain uncertain, posing a challenge in extracting fiducial points [14]. As a result,

Received 8 September 2023; revised 21 December 2023 and 14 April 2024; accepted 14 May 2024. Date of publication 10 June 2024; date of current version 4 October 2024. This work was supported in part by the Schweizer Paraplegiker Stiftung (SPS) and in part by the ETH Zürich Foundation ETH-SPS Digital Transformation in Personalized Health Care for SCI. (Corresponding author: Diego Paez-Granados.)

Ana Cisnal is with Spinal Cord Injury & Artificial Intelligence Lab, ETH Zurich, 8092 Zürich, Switzerland, and also with Instituto de las Tecnologías Avanzadas de la Producción (ITAP), University of Valladolid, 47002 Valladolid, Spain (e-mail: cisnal@ieee.org).

Yanke Li and Mehdi Ejtehad are with Spinal Cord Injury & Artificial Intelligence Lab, ETH Zurich, 8092 Zürich, Switzerland (e-mail: yanke.li@hest.ethz.ch; mehdi.ejtehad@hest.ethz.ch).

Bertram Fuchs is with Spinal Cord Injury & Artificial Intelligence Lab, ETH Zurich, 8092 Zürich, Switzerland, and also with School of Computation, Information and Technology, Technical University of Munich, 80333 München, Germany (e-mail: bertram.fuchs@hest.ethz.ch).

Robert Riener is with Sensory-Motor Systems Lab, ETH Zurich, 8092 Zürich, Switzerland, and also with Spinal Cord Injury Center, University Hospital Balgrist, 8008 Zürich, Switzerland (e-mail: robert.riener@hest.ethz.ch).

Diego Paez-Granados is with Spinal Cord Injury & Artificial Intelligence Lab, ETH Zurich, 8092 Zürich, Switzerland, and also with Swiss Paraplegic Research (SPF), 6207 Nottwil, Switzerland (e-mail: diego.paez@hest.ethz.ch).

Digital Object Identifier 10.1109/JBHI.2024.3411693

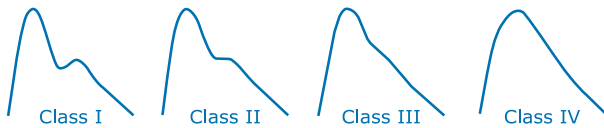


Fig. 1. Classification of PPG waveforms: I) A DN is evident; II) No evident DN but the descendent line becomes horizontal; III) No evident DN but there is a change in the angle of the descendent line; IV) No evidence of a DN is visible.

existing machine learning (ML) models have predominantly focused on optimizing their performance using a restricted set of fiducial points. This approach heavily relies on the dataset and features, leading to the absence of consensus regarding the optimal feature space for BP estimation. The presented work introduces a PWA-based algorithm that identifies all fiducial points for subsequent feature extraction, validated across all Dawber's classes to align with its intended medical application.

Previous research had primarily concentrated on in-clinic data, aiming to optimize performance for these specific datasets, thereby overlooking generalization across diverse populations. This is particularly critical for populations with limited available data, such as pregnant women, neonates, and spinal cord injury (SCI) patients, who could greatly benefit from ABP estimations, given their specialized cardiovascular care needs. To address this, personalized models and domain-transferred features are crucial, with transfer learning and feature selection aiding in identifying discriminatory features and enhancing learning quality. Moreover, since the BP waveform is not yet fully understood, meticulous feature extraction and selection are paramount for subsequent model development that prioritizes interpretability. Through the proposed feature selection method, leveraging the Markov Blanket (MB), discriminatory features pertinent to BP estimation can be pinpointed, facilitating the creation of precise and dependable models, even in the face of limited sample sizes. By eliminating less informative features and focusing on those crucial for BP estimation, the MB-based feature selection method elevates the quality of learning and enhances the interpretability of subsequent ML models, overcoming challenges associated with generalization across diverse populations.

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

II. MATERIALS AND METHODS

In contrast to previous works, the proposed 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 estimation accuracy for a single dataset. This emphasis on feature robustness contributes to the generalizability and reliability of the BP estimation method. The proposed PPG-based BP estimation method consists of five steps (Fig. 2): 1) signal pre-processing for noise removal, signal segmentation, and data cleaning; 2) identification of the fiducial points; 3) extraction of signal and demographic features; 4) robust feature selection. 5) estimation of BP using ML models.

A. Dataset

The Aurora-BP dataset [15] was chosen as the first large-scale collection of ambulatory and cuff-less BP measurements obtained over a 24-hour period using wearable technology, including ECG, tonometry, PPG, and reference BP. The dataset comprises two non-concurrent protocols: auscultatory and oscillometric. While the auscultatory protocol solely relied on in-clinic manual auscultation by trained observers using an aneroid sphygmomanometer, the oscillometric protocol employed an automated ABP monitor allowing both in-clinic and ambulatory measurements. This study focuses on ABP estimation using one PPG signal, utilizing data from the oscillometric protocol, in which participants underwent an initial clinical visit in the morning and a return visit after the 24-hour ambulatory recording. 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. The reference BP was measured using the Spacelabs Healthcare OnTrak 90227 ABP monitor fitted to the participant's dominant arm or according to their preference. The PPG optical sensor was based on the MAX30101, and it was worn on the opposite wrist. Both devices were time-synchronized before the initial visit for simultaneous measurements. For each pair of cuff-based BP measurements, 30-second PPG signal segments with a frequency of 500 Hz were stored in the database.

B. PPG Signal Pre-Processing

Some time-domain features are derived from time, amplitude, and area measurements extracted using PWA, which rely on the precise identification of fiducial points. However, their identification is challenging, leading many studies to only consider the systolic peak (S). To compensate for this limitation, additional parameters like PAT, PTT, and PWV are often incorporated [17], [18], [19], [20], with the inherent disadvantage of requiring a second signal. In contrast, this paper presents a method to automatically extract all fiducial points, leading to a broader spectrum of features, even in the presence of noisy ambulatory data. Identifying fiducial points within a PPG pulse waveform involves analyzing its derivatives, which can be severely distorted by noise. Hence, pulse segmentation, effective noise reduction and quality analysis are imperative for automatic and accurate detection.

First, high-frequency noise removal is required, and low-pass filtering should keep a balance between preserving the original features and mitigating noise. Moreover, some studies employ

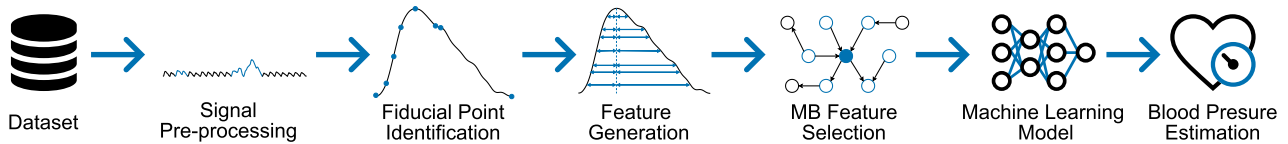


Fig. 2. Schematic diagram illustrating the methodology employed in this study for the estimation of BP using PPG.

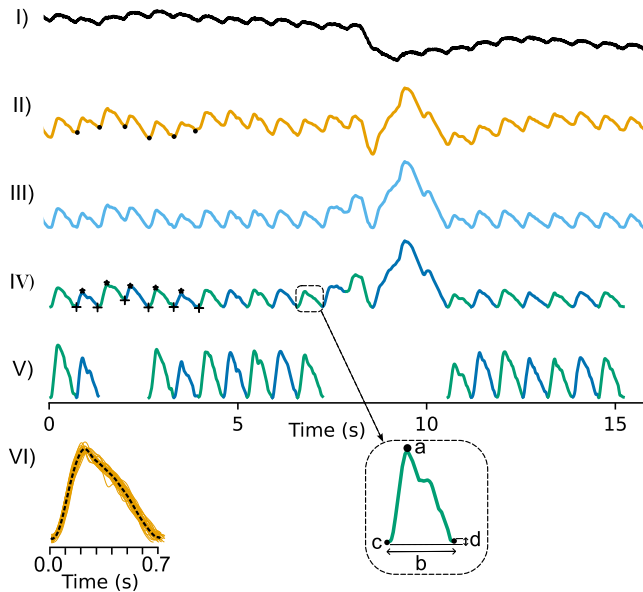


Fig. 3. PPG signal pre-processing steps: I) Raw PPG signal; II) Signal after Butterworth band-pass filter with DC drift, resulting in amplitude variation of the troughs (-); III) Signal after baseline removal; IV) Pulse segmentation by detecting peaks (*) and troughs (+); V) Remaining beat-to-beat waves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

polynomial interpolation [21] or filtering methods such as 25-ms moving average [22] to smoothen the signals and enhance derivative computations. Nonetheless, these approaches run the risk of altering the signal characteristics. In this 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, identifying individual beat-to-beat pulse waves is essential, requiring pulse peak and trough detection (Fig. 3). However, biological characteristics (tissue composition, respiration, vasomotor activity and thermoregulation) and external factors (light and acquisition device) can influence the signal baseline [23], making reliable trough detection more difficult. Hence, baseline removal is performed using an adaptive iteratively reweighted penalized least squares (airPLS) [24]. It autonomously adjusts to remove both linear and non-linear baselines without manual intervention or prior peak identification, and demonstrates resilience in noisy environments. It ensures accurate delineation of pulse waveforms and enhances trough detection by mitigating baseline-induced distortions. The method's dynamic thresholding improves detection of true peaks and troughs by evaluating local extrema against

the adaptively corrected baseline, thereby refining beat-to-beat pulse wave feature extraction.

The extraction of the beat-to-beat waves involves the detection of the peaks and troughs using an adaptive amplitude threshold [12]. 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 to 70% of the range between the median values of the identified maxima and minima within the 30-second window, respectively.

After identifying the pulse waves, the pulse morphology was examined based on four parameters with heuristically determined thresholds to ensure accurate segmentation by discarding any unreliable wave units. 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. The noise of the remaining PPG waves was further investigated using template-matching, which assesses pulse waveform similarities within a signal segment [26]. Regardless of the actual waveform morphology, this approach identifies irregularities caused by artifacts, as a clean segment exhibits similar pulse morphologies. The template is formed by averaging PPG waves in a segment, and correlation measurements are used to evaluate signal regularity (and quality). In this work, template-matching is applied to 30-second PPG segments. If a wave's Euclidean distance from the template surpassed a device-specific noise threshold, it was deemed unreliable and discarded. Finally, a segment is totally discarded if the number of valid waves is less than 5.

C. Fiducial Point Identification

The fiducial points were identified for each pulse wave using derivative analysis (Fig. 4(a)). 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

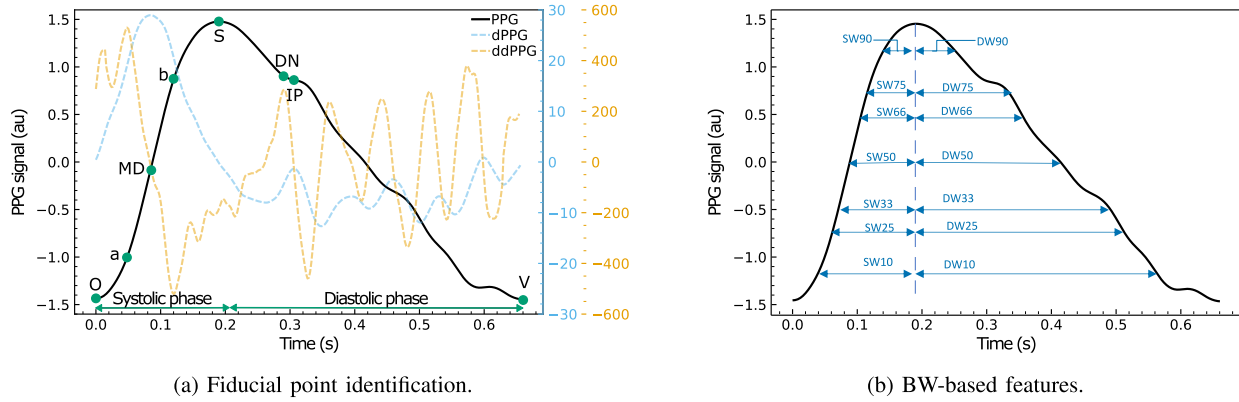


Fig. 4. (a) Identification of the fiducial points based on the analysis of PPG, dPPG (first derivative or velocity of PPG), and ddPPG (second derivative or acceleration of PPG) signals. (b) BW-based features: Systolic branch width (SBWx) and diastolic branch width (DBWx).

change of sign in the first derivative from negative to positive occurs [22]. Likewise, the point 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].

This work presents an 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 point S was 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 point b denotes the 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 segment 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 pre-processing and feature extraction stages was made readily accessible [16] with this publication. The package accepts a raw signal and produces a feature vector that represents the observed time interval. The package allows

Algorithm 1: Algorithm for Fiducial Point Identification.

Input: wave

Output: O, V, S, MD, D, IP, DN, a, b

```

1: d_wave = first_derivative(wave)
2: dd_wave = second_derivative(wave)
3: O = wave[first point]
4: V = wave[last point]
5: S = maximum_peak(wave)
6: systolic_phase = wave[from O to S]
7: diastolic_phase = wave[from S to V]
8: MD = maximum_peak(wave in systolic_phase)
9: search_zone = [diastolic_phase from 80 ms to
0.6 × duration(diastolic_phase)]
10: n = number_of_local_maxima(wave in
search_zone)
11: if n = 0 then
12:   D = maximum_peak(d_wave in search_zone)
13:   IP = D
14:   DN = maximum_peak(dd_wave right before IP)
15: else if n = 1 then
16:   D = maximum_peak(wave in search_zone)
17:   IP = local_maximum(d_wave right before D)
18:   DN = local_minimum(wave right before IP)
19: else
20:   IP = maximum_peak(d_wave in search_zone)
21:   D = local_maximum(wave right after IP)
22:   DN = local_minimum(wave right before IP)
23: end if
24: a = maximum_peak(dd_wave in systolic_phase)
25: b = minimum_peak(d_wave in systolic_phase)
26: return O, V, S, MD, D, IP, DN, a, b

```

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 subsequently 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 fiducial point combinations.

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 [8], [9], [11], [12], [17]. A total of 28 BW-based features were extracted (Fig. 4(b)), including the systolic branch width (SBW_x), the diastolic branch width (DBW_x), the branch width (DW_x = SBW_x+DBW_x), and the branch width ratio (BWR_x=DBW_x/SBW_x) at x% of the pulse height (x = 10, 25, 33, 50, 66, 75, 90).

The morphology of PPG signals is intricately linked to various physiological factors, including arterial tone, peripheral resistance, and blood viscosity [3]. Previous research has identified specific metrics strongly correlated with BP, which have subsequently been employed as features in ML models [17], [21], [31], [32]. Reflection index (RI), also known as augmentation index (AI), measures the pulse reflection, which is related to the arterial tone and it is calculated as the intensity's ratio between the intensity of the S and IP. Inflection point area (IPA) is defined as the ratio of the area between O, MS, S, IP, V and was proved to be an indicator of peripheral resistance. Crest time (CT) is the time difference between O and S, which is related to the PWV. PPGK, also known as PPG characteristic value or K value, 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. Large Artery Stiffness Index (LASI) is an indicator of the stiffness of the arteries and is inversely related to the time interval between S and IP. While the aforementioned features were initially included in the analysis as intensity, time, or area-related features, the additional features of PPGK, mNpV, 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 (FF)*: Using the Fast Fourier Transform (FFT), the frequency and magnitude for the first, second, and third harmonics were extracted. Relative power [34] and its quantification 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 [9], [12], [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. The objective was to analyse the best transferability across sub-populations. Therefore, a wrapped selection was conducted 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, the minimum redundancy maximum relevance (mRMR) method was employed. It is a filter-based feature selection approach that reduces the feature set by selecting features highly relevant to the target variable while minimizing redundancy among them. By prioritizing relevant features and penalizing redundancy, mRMR ensures the selection of a diverse and complementary subset, improving model performance and interpretability [36].

Subsequently, the Predictive Permutation Feature Selection (PPFS) [37], a wrapped feature selection method using the MB derived from a graphical model, was applied. PPFS statistically determines the sufficient set of features for estimation across different subgroups. While this approach may potentially introduce bias to subsequent models, it helps identify the most robust feature space for potential transfer to ambulatory settings in future studies involving smaller populations.

MB Discovery refers to finding the set of variables that are sufficient for estimating 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 (Fig. 5(b)). This paper applies the concept of graphical models, where nodes and edges represent variables and their statistical relations. This framework enables a deeper understanding of complex systems due to its compact statistical relations and clear structural

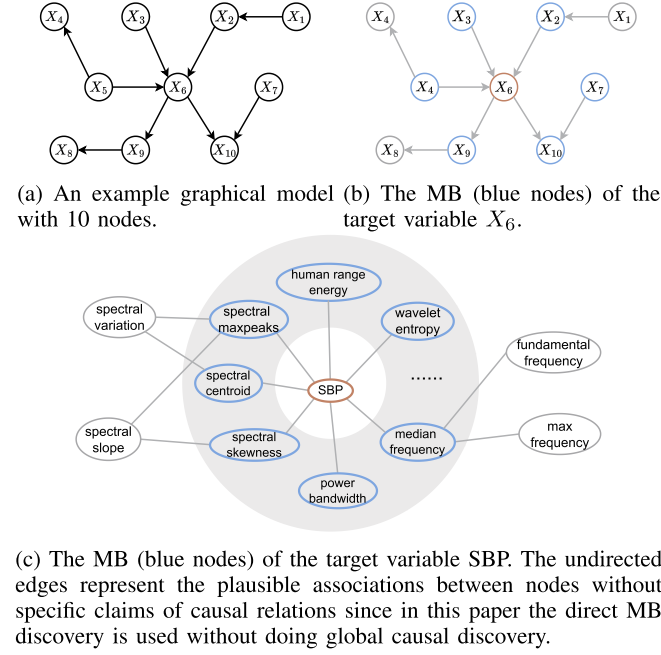


Fig. 5. Illustration of (a) graphical models; (b) Markov blanket (MB); (c) the potential application of using MB for SBP estimation.

representation. Fig. 5(a) shows a graphical model with 10 variables and its factorized distribution is derived as:

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

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.

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 estimating the target variable. The MB, shown in Fig. 5(c), identifies sparse structures and essential features for SBP estimation from generated PPG feature space. Predictive models trained on these selected features are aim to enhance robustness against spurious correlations and covariate shifts in the generated feature space across diverse populations [39]. This paper focuses on approaches that can handle mixed-type features without distribution assumptions using non-parametric methods, leading to the adoption of the PPFS [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.

To ensure robust and generalizable performance evaluation, a nested cross-validation (CV) was employed. This approach surpasses the conventional train/test split by mitigating overfitting and bias inherent in relying on a single static split. It iterates through various test datasets within an outer loop, resulting in a more realistic assessment of model performance across diverse data subsets. Furthermore, to address the challenge of accurately estimating classification performance for each test fold, an inner (nested) CV loop iterates over different validation datasets within each outer loop iteration, optimizing the model's hyperparameters based on the validation data's classification performance. This two-layered process ensures the model's true generalizability on unseen test data in the outer loop, guarding against overfitting to a specific training set [40]. In this study, a 10-fold inner CV, optimized hyperparameters for each test fold of the outer leave-one-participant-out CV, ensuring each participant served as unseen test data exactly once. The overall estimation error was given by the average of the individual model scores of every participant.

The assessment followed metrics and visualizations recommended by the IEEE Std 1708-2014 standard [41], as well as ANSI/AAMI SP 10-1987 standard [42]. 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. It also introduces the MAE and mean absolute percentage error (MAPE) as an alternatives to the mean \pm standard deviation proposed by the ANSI/AAMI and the cumulative percentage errors suggested by the British Hypertension Society (BHS) protocol [43].

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 (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 lines at 5, 6, and 7 mmHg to delineate these grades.

Since the BP data was collected from different populations, the primary interest lies 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 recommendations on how to utilize these metrics for further applications. In this work, the measurement of the distance between two empirical distributions is accomplished using the Wasserstein distance metric [44]. The p -Wasserstein distance between probability measures μ and ν on \mathbb{R}^d , given with a hyperparameter p , 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 [45]. For a better focus of the paper, details of this algorithm are omitted.

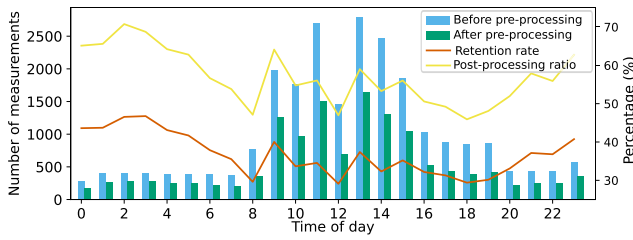


Fig. 6. Measurements before and after signal pre-processing, retention rate and post-processing ratio by time of day.

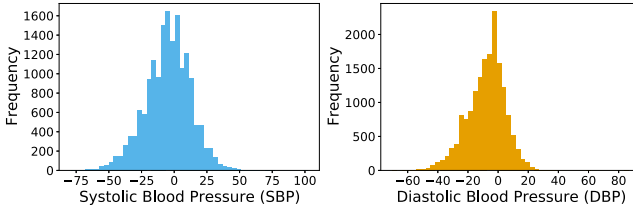


Fig. 7. BP distribution in the dataset per measurement from the individual resting point (mmHg).

III. RESULTS

The performance of different ML algorithms for BP estimation with MB-selected features was initially compared to identify the most accurate 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 evaluated by training and testing several models specialized for different groups and a generic model for all data. Additionally, the performance of the BP estimation algorithm is evaluated for ambulatory and in-clinic data separately.

A. Generated Dataset

From the initial cohort of 548 participants, in-clinic BP in a resting position (specifically sitting with the arm down) was available for only 534 participants, serving as the reference BP. Each participant contributed a mean (SD) of 55.70 (4.11) measurements, with 14.15 (0.68) and 41.55 (4.07) measurements conducted during in-clinic and ambulatory settings. After signal pre-processing, the cohort was reduced to 523 participants. The signal retention rate, calculated as the ratio of the duration of valid pulses to the segment duration, was 35.22 (35.00)% with rates of 32.54 (34.63)% and 41.25 (35.06)% for in-clinic and ambulatory data. The post-processing ratio, representing the available measurements after pre-processing compared to before, was 56.97 (7.22)% (Fig. 6).¹

Assessing the algorithm's performance requires a broad distribution of BP data relative to the resting position. According to the IEEE standard [41], BP values at the calibration point should vary by at least 10 mmHg for DBP and 15 mmHg for SBP. Fig. 7 shows the distribution of BP changes. The mean

¹Further details regarding these metrics are available in [16].

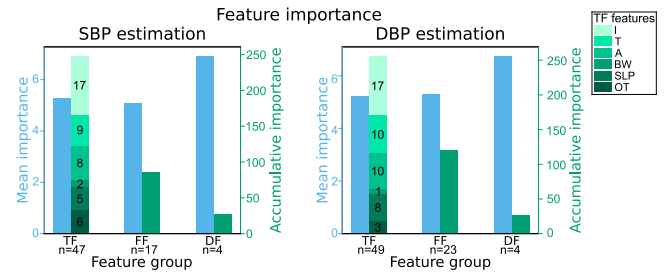


Fig. 8. Number of selected features, along with their mean and cumulative importance, segregated into the three main categories: Time (TF), frequency (FF), and demographic (DF) features. The time domain is subdivided into intensity (I), time (T), area (A), slope (SLP), branch width (BW), and others (OT), illustrating their cumulative importance and quantity.

TABLE I
PERFORMANCE OF ML ALGORITHMS FOR ESTIMATION

ML Model	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

of DBP and SBP at rest position are 86.21 (10.88) and 131.15 (15.53) mmHg, respectively, with delta DBP and SBP of 40.72 (11.87) and 54.03 (16.25) mmHg, respectively.

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, with 38 features in common². Among the selected features, no SFs were included, but four DFs were deemed significant for both DBP and SBP estimation, ranking among the top 11 features in terms of importance. While DF features have the highest mean importance, their lower count (n=4) results in TF and FF features with lower mean importance, contributing more significantly to the overall cumulative importance (Fig. 8). An evaluation of the feature selection process robustness through MB compared to recursive feature elimination is detailed in the Appendix.

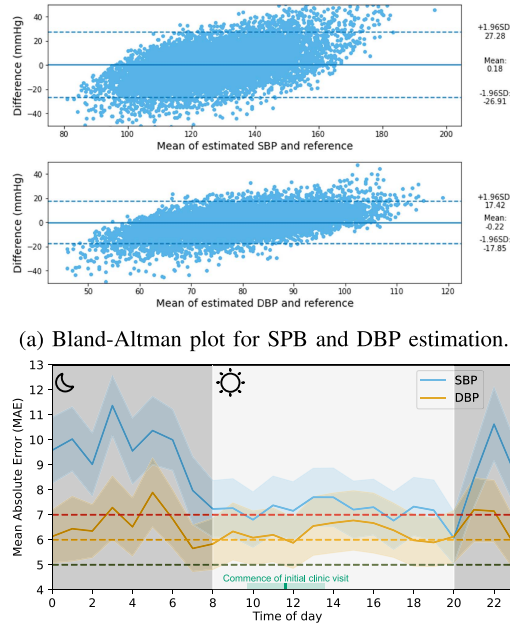
Following the feature selection process, two datasets were created. The systolic BP dataset consists of 68 features and a reference BP value, and the DBP dataset comprises 76 features and a reference BP value. Various scalers were tested, and the MinMaxScaler yielded the highest accuracy. Table I presents the results for BP estimation using each algorithm while applying the MinMaxScaler. Among the algorithms, GB achieved the highest accuracy with its hierarchical structure, comprising an ensemble of decision trees, each sequentially constructed to rectify predecessor errors. Through a hyperparameter tuning process, the optimal configuration was determined, featuring a learning rate of 0.01, Huber loss function, maximum tree depth of 8, minimum samples per leaf of 1, minimum samples for a split of 2, 500 estimators, and a subsample ratio of 0.5.

²The selected features ranked by their importance are available in [16].

TABLE II
PERFORMANCE, EXPRESSED AS MD (SD), MAE (MD) AND MAPE, OF THE GENERIC MODEL AND SPECIALIZED MODELS BASED ON BP CATEGORY

	Available measurements			SBP						DBP					
				Individual			Generic			Individual			Generic		
	Total	Percentage (%)	Mean (SD)	MD (SD) (mmHg)	MAE (SD) (mmHg)	MAPE (%)	MD (SD) (mmHg)	MAE (SD) (mmHg)	MAPE (%)	MD (SD) (mmHg)	MAE (SD) (mmHg)	MAPE (%)	MD (SD) (mmHg)	MAE (SD) (mmHg)	MAPE (%)
Normal	2057	16.4%	23.38 (11.55)	0.40 (12.98)	9.82 (0.29)	8.89	0.10 (13.59) ²	10.27 (0.29) ²	8.35 ²	-0.20 (7.80)	5.91 (0.34)	8.99	-0.20 (8.82) ²	6.71 (0.22) ²	9.23 ²
Elevated	3082	24.6%	24.66 (11.45)	0.17 (12.54)	9.63 (0.41)	7.92	0.13 (13.65) ²	10.32 (0.24) ²	8.37 ²	-0.01 (8.50)	6.58 (0.19)	9.03	-0.23 (8.91) ²	6.76 (0.14) ²	9.31 ²
HTN S1	4359	34.8%	23.19 (11.42)	0.04 (3.95)	10.58 (0.50)	8.49	0.16 (13.63) ²	10.27 (0.32) ²	8.33 ²	0.00 (9.12)	7.01 (0.29)	9.57	-0.22 (8.86) ²	6.73 (0.18) ²	9.25 ²
HTN S2	3012	24.1%	24.69 (12.22)	-1.38 (15.57)	11.77 (0.47)	8.95	0.12 (13.65) ²	10.31 (0.24) ²	8.34 ²	-1.36 (10.39)	7.95 (0.19)	10.46	-0.24 (8.88) ²	6.74 (0.17) ²	9.28 ²
All	12510	100%	23.92 (11.36)	-0.20 (13.83) ¹	10.50 (0.91) ¹	8.53 ¹	0.13 (13.63)	10.29 (0.28)	8.35	-0.35 (9.05) ¹	6.94 (0.72) ¹	9.55 ¹	-0.22 (8.87)	6.74 (0.18)	9.27

¹ Indicates combined results from individual models. ² Indicates split results from the generic model by BP group.



(b) Temporal evolution of MAE for SBP and DBP estimation.

Fig. 9. GB model performance for BP estimation. (a) Bland-Altman plot showing the MD and SD of the reference and estimated BP. The ANSI/AAMI SP 10 requires to have MD \pm SD less 5 ± 8 mmHg. (b) MAE across day and night hours, indicating the start of the measurements. Lines represent the average MAE, and the shaded area depicts the SD.

The GB model exhibited a MAE of 10.29 and 6.74 mmHg and a MD (SD) of 0.18 (13.82) and -0.22 (8.71) for SBP and DBP (Fig. 9(a)). Moreover, Fig. 9(b) presents the evolution of the MAE for SBP and DBP estimation throughout day hours. Note that the initial clinical visit commenced at a mean time of $11:30:35 \pm 01:40:11$ (range: 09:07:47 to 15:24:23). Subsequent analyses exclusively report the performance of the GB model.

C. Results for Sub-Populations

Participants were categorized into four BP categories following ACC/AHA guidelines [46], based on their resting SBP and DBP (Fig. 10(a)): normal, elevated, hypertension stage 1 (HTN S1), and hypertension stage 2 (HTN S2). To assess the distributional disparity of features among BP categories, the normalized Wasserstein distance matrix were estimated. Contrary to the inter-group distance observed in the original feature space (Fig. 10(b)), the feature space chosen by PPFS for SBP (Fig. 10(c)) and DBP (Fig. 10(d)) estimation exhibits a more pronounced discrepancy across categories. These features, with

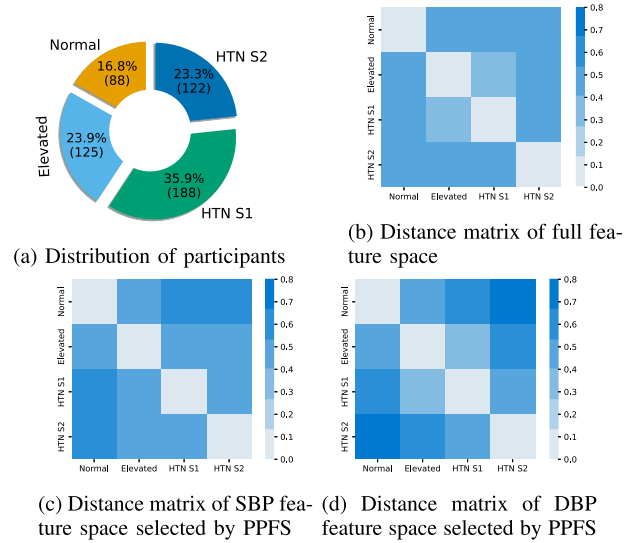


Fig. 10. (a) Distribution of participants' BP profile. (b) Normalized distance matrix for the full feature space (after standardization) across different sub-populations by BP profile, normalized to the number of dimensions. (c) Normalized distance matrix of selected SBP feature space by PPFS. (d) Normalized distance matrix of selected DBP feature space by PPFS.

better discrimination, hold promise for training robust generic models against feature shift over BP categories.

One generic model was developed using the information coming from all the participants, while various specialized models were trained and tested with each sub-population based on the BP category. The MAE of the generic model along with the MAE of the specialized models 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 SD for each participant is also provided in Table II.

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; 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 showed 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

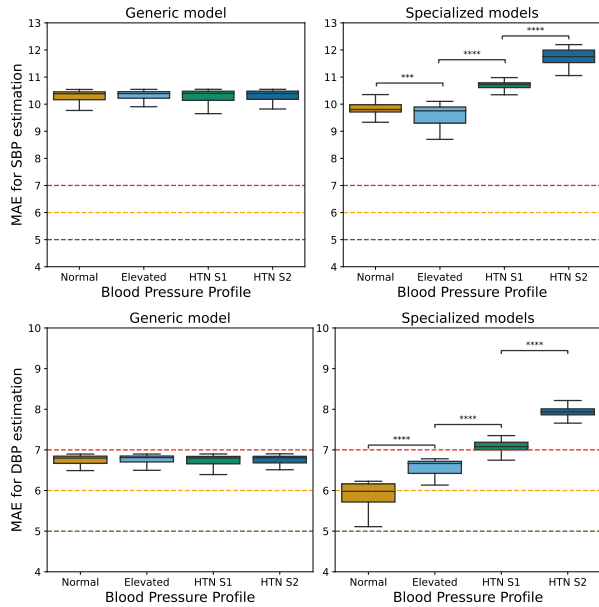


Fig. 11. Performance evaluation of the generic model and specialized models for BP estimation across different BP categories: Normal, elevated, HTN S1, and HTN S2. * denotes significant difference by a paired t-test at the level $p < 0.0001$.

(BP profile) had no significant impact on the results of the generic model for estimating both SBP (p -level > 0.1 , F value = 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

To evaluate the impact of data type on BP estimation, the database was separated into two subsets: in-clinic and ambulatory. For each subset, a GB model was trained and tested separately after the feature selection. For SBP estimation, 61 and 49 features were selected for ambulatory and in-clinic datasets; for DBP estimation, counts were 62 and 46. 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. The MD (SD) values for SBP and DBP are 0.63 (11.73) and 0.01 (6.58) in in-clinic, and -0.33 (14.54) and -0.22 (10.10) in ambulatory measurements. The MAPE for SBP and DBP is 6.34% and 5.03% in in-clinic, and 9.54% and 11.15% in ambulatory measurements. Statistically significant differences were found between ambulatory and in-clinic data in terms of MAE for both SBP (p -level < 0.0001 , F-value = 32764.76) and DBP (p -level < 0.0001 , F-value 65675.36) estimation (Fig. 12).

IV. DISCUSSION

A. Feature Selection

Prior studies focused on selecting the optimal features for BP estimation, yet many of them proposed a limited range of

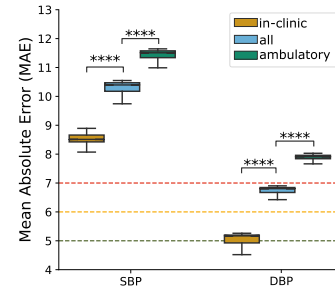


Fig. 12. Performance of the BP estimation algorithm using in-clinic, ambulatory, and all measurements. * Denotes a significant difference by a paired t-test at the level $p < 0.0001$.

features. For instance, [10], [17], [47] proposed feature sets confined to TF, with 24, 42 and 65 features. These restricted set overlooked valuable information encoded in other feature domains. Similarly, [12] proposed 74 features encompassing TF, SF and DF features, while disregarding FF. Likewise, [19] proposed 34 features in the TF, FF and SF domains, omitting DF. In contrast, this work offers a comprehensive set of features belonging to four main categories (TF, FF, ST, and DF), ensuring a robust feature selection process through the MB with causal and effect nodes included.

Feature selection results revealed that over half of the chosen features belong to the TF category, including all six subcategories. This aligns with prior studies [9], [12], which underscored the significance of TF in BP estimation. The results also indicated the importance of DF features, in line with other works [12], [48], although BMI was excluded. FF were identified as well, including the wavelet entropy for both SBP and DBP estimation. This contradicts the results of [19], which ranked approximate entropy in the top 15, but no wavelet domain features were considered influential. Lastly, SF features were found to offer less informative value compared to other feature types, consistent with previous works [9], [12], but not with [19], which identified SK and SD among the 15 top features.

B. Domain Transfer

The study conducted by Miao et al. [49] underscored substantial variation in feature importance across individuals, highlighting the subject-specific nature of feature selection and the need for individualized approaches in accurate BP estimation. In this line, the work in [8] introduced a two-step algorithm classifying BP into categories and applying specific algorithms for each. The two-step algorithm outperformed the generic approach with statistically significant differences in SBP/DBP estimation. Additionally, the BP profile significantly influenced the DBP/SBP estimation accuracy using either model. Similarly, [50] reported significantly higher estimation errors in HTN subjects than in normotensive individuals.

However, in practical applications, subject-specific feature selection may not be feasible due to the need for extensive individual data for model pre-training. This work proposes adopting an MB-based feature selection method to obtain a reliable and consistent feature set across individuals, enabling robust BP estimation without requiring subject-specific customization. As

shown in Fig. 10(b)–(d), compared with the normalized distance matrix of full feature space, the normalized distances of selected feature spaces for SBP/DBP estimation appear more salient. PPFS identifies the set of features that are more informative and differentiated across sub-populations, enabling the consequent generic modeling to capture those individualized distinctions. These findings reflect that the generic model exhibited significantly superior estimation accuracy compared to the overall accuracy of the specialized models for both SBP and DBP (Table II). Moreover, the BP profile did not have a significant impact on the accuracy of DBP/SBP estimation when using the generic model (Fig. 11). Additionally, the feature space selected by MB had better transferability in terms of estimation accuracy and robustness. Indeed, the outcomes (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 populations, such as pregnant women, neonates and SCI individuals, where data to train large learning models may be insufficient.

C. Ambulatory vs In-Clinic Data

Performance validation for ambulatory measurements showed that the estimation deviated more from the reference BP compared to in-clinic ones (Fig. 12). Ambulatory data, both for SBP and DBP, did not satisfy the requirements of either the IEEE 1708 and the ANSI/AAMI standards, while in-clinic data met the standards for DBP. Nonetheless, in-clinic data, while falling short of meeting the standards for SBP estimation, exhibited a closer proximity to the benchmark than the ambulatory data. This discrepancy underscores the impact of the data source (ambulatory vs. in-clinic) on meeting BP standards. Previous research has pointed out the inadequacy of those methods to follow the dynamic changes in BP elicited by the cardiovascular autonomic nervous activities [50]. Nevertheless, it is important to remark that in-clinic data included not only resting measurements (supine, sitting arm down/lap/up) but also dynamic BP changes induced by various cardiac activities (walking, cooling down, and running).

It can be therefore concluded that BP estimation accuracy from in-clinic data is superior to the 24-hour ambulatory recordings due to the latter's increased susceptibility to noise. PPG signals are highly influenced by both biological characteristics (respiration, vasomotor activity, thermoregulation) and by external factors, such as motion or light [3]. However, the potential of ABP estimation based on PPG signals lies in its capacity to continuously monitor dynamic BP changes, providing valuable insights into cardiovascular health biomarkers. Moreover, its non-invasive and wearable design ensure high patient acceptability, promoting widespread use.

D. Comparative Analysis

The generic model lacks below the IEEE 1708 or ANSI/AAMI/ISO standards for SBP estimation in 24-h data. However, for DBP estimation, it demonstrated compliance with the IEEE standard but fell short of meeting the ANSI/AAMI

criteria. This work is the first of 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 proves challenging due to the disparities in the sample size [51], data and target population. As observed, the dataset's characteristics have a profound impact on BP estimation performance. This influence is well-documented across various studies [5], [9], [11], where the same BP estimation algorithm applied to different datasets has yielded notable variations in accuracy. These variations can be attributed to a multitude of factors, including distinctions in population characteristics, data acquisition protocols, signal quality, and noise levels.

To date the only work using the Aurora-BP dataset is [15]. Their BP estimation algorithm based on PPG achieved a SBP and DBP estimation accuracy of 0.42 (8.98) and 0.54 (5.95) mmHg, respectively. In contrast, this work showed lower accuracy with values of 0.48 (13.93) and -0.22 (8.71) mmHg for SBP and DBP estimation. Nonetheless, while their study only used an average of 19.3 (10.1) measurements for a total of 227 participants, without providing specific details regarding the resulting BP distribution, this study incorporated data from 523 participants, with a mean of 23.92 (11.36) measurements per participant and provided the BP distribution (Fig. 10). These discrepancies highlight the influence of data processing and cleaning techniques on outcomes. Three additional features (acceleration, sine, and cosine of the time of day) were included in [15], but omitted in this study to rely only on a single signal modality. Consequently, the model exhibits increased error during nighttime hours (Fig. 9(b)), as it does not account for night BP dipping during sleep. The imbalanced data, with more measurements available during the day than at night (Fig. 6), coupled with the use of rest SBP and DBP as input features, likely contribute to the higher MAE observed during nighttime hours. Unlike prior work, the time of day was deliberately omitted as a feature in the model to avoid bias from uniform sleep pattern assumptions and prevent introducing a correlated but non-causal factor to BP estimation.

V. CONCLUSION

A comprehensive BP estimation method from PPG signals has been validated with ambulatory measurements across diverse populations, presenting a stable algorithm for PPG feature extraction and robust feature selection. The proposed method, available in the public library [16], incorporates signal pre-processing to ensure accurate pulse segmentation and quality analysis before employing the fiducial point identification algorithm for subsequent feature extraction. Leveraging the extracted features, the MB method proved to effectively determine features for robust transfer among populations. We believe these findings provide a more realistic representation of a BP estimation algorithm in ambulatory diagnosis and monitoring applications for populations at risk of cardiovascular diseases, which often face data limitations. Significant deviations in model performance between in-clinic and ambulatory data highlight the need for more reliable models and data for ABP estimation.

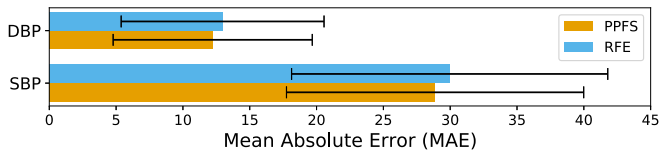


Fig. 13. Performance comparison between PPFS and RFE method in the task of domain generalization: Both feature selection methods were tested using a cross-population validation strategy where the estimation results are obtained using the selected feature space with ML model finetuning.

Subsequent research should prioritize consistent data collection and developing robust algorithms capable of handling suboptimal conditions in PPG signals across diverse populations.

Data Availability: The Aurora-BP dataset [15] is utilized under a data transfer agreement signed between Microsoft and ETH Zurich.

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

APPENDIX FEATURE SELECTION ROBUSTNESS

The robustness of the MB-based feature selection method (PPFS) was evaluated and compared with the baseline of recursive feature elimination (RFE), an algorithm commonly used in ML to identify the most relevant features for a given variable [52]. It is an iterative process that aims to determine the optimal subset of features by eliminating less essential or redundant features based on a ranking such as feature importance or model-specific coefficients. 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," *AAPS J.*, vol. 8, no. 1, pp. E146–E152, Mar. 2006. [Online]. Available: <http://link.springer.com/10.1208/aapsj080117>
- [2] S. M. S. Islam et al., "Wearable cuffless blood pressure monitoring devices: A systematic review and meta-analysis," *Eur. Heart J.-Digit. Health*, vol. 3, no. 2, pp. 323–337, Jul. 2022. [Online]. Available: <https://academic.oup.com/ehjdh/article/3/2/323/6576559>
- [3] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiol. Meas.*, vol. 28, no. 3, pp. R1–R39, Mar. 2007. [Online]. Available: <https://iopscience.iop.org/article/10.1088/0967-3334/28/3/R01>
- [4] A. Esmaili, M. Kachuee, and M. Shabany, "Nonlinear cuffless blood pressure estimation of healthy subjects using pulse transit time and arrival time," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 12, pp. 3299–3308, Dec. 2017. [Online]. Available: <https://ieeexplore.ieee.org/document/8032000/>
- [5] 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, Apr. 2022, Art. no. 1442. [Online]. Available: <https://www.mdpi.com/2079-9292/11/9/1442>
- [6] 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 J.*, vol. 22, no. 3, pp. 2475–2483, Feb. 2022. [Online]. Available: <https://ieeexplore.ieee.org/document/9646921/>
- [7] S. Haddad, A. Boukhayma, and A. Caizzone, "Continuous PPG-Based blood pressure monitoring using multi-linear regression," *IEEE J. Biomed. Health Inform.*, vol. 26, no. 5, pp. 2096–2105, May 2022. [Online]. Available: <https://ieeexplore.ieee.org/document/9616476/>
- [8] 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. 58146–58154, 2020. [Online]. Available: <https://ieeexplore.ieee.org/document/9042315/>
- [9] 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. 138817–138833, 2021. [Online]. Available: <https://ieeexplore.ieee.org/document/9558767/>
- [10] 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, Oct. 2020, Art. no. 5668. [Online]. Available: <https://www.mdpi.com/1424-8220/20/19/5668>
- [11] 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 photoplethysmogram," *IEEE Trans. Consum. Electron.*, vol. 68, no. 3, pp. 236–260, Aug. 2022. [Online]. Available: <https://ieeexplore.ieee.org/document/9774278/>
- [12] P. Yao et al., "Multi-dimensional feature combination method for continuous blood pressure measurement based on wrist PPG sensor," *IEEE J. Biomed. Health Inform.*, vol. 26, no. 8, pp. 3708–3719, Aug. 2022. [Online]. Available: <https://ieeexplore.ieee.org/document/9756927/>
- [13] 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, doi: [10.1177/000331977302400407](https://doi.org/10.1177/000331977302400407).
- [14] M. A. Gamrah, J. Xu, A. E. Sawy, H. Aguib, M. Yacoub, and K. H. Parker, "Mechanics of the dicrotic notch: An acceleration hypothesis," *Proc. Inst. Mech. Eng. Part H: J. Eng. Med.*, vol. 234, pp. 1253–1259, Nov. 2020, doi: [10.1177/0954411920921628](https://doi.org/10.1177/0954411920921628).
- [15] R. Mieloszyk et al., "A comparison of wearable tonometry, photoplethysmography, and electrocardiography for cuffless measurement of blood pressure in an ambulatory setting," *IEEE J. Biomed. Health Inform.*, vol. 26, no. 7, pp. 2864–2875, Jul. 2022. [Online]. Available: <https://ieeexplore.ieee.org/document/9721156/>
- [16] A. Cissal, 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 J. Biomed. Health Inform.*, vol. 25, no. 4, pp. 1018–1030, Apr. 2021. [Online]. Available: <https://ieeexplore.ieee.org/document/9142317/>
- [18] B. D. Marchi, M. Frigerio, S. D. 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, Nov. 2021, Art. no. 7334. [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 J. Biomed. Health Inform.*, 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, Dec. 2018, Art. no. 4227. [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 J.*, vol. 20, no. 8, pp. 4300–4310, Apr. 2020. [Online]. Available: <https://ieeexplore.ieee.org/document/8938751/>
- [22] O. Vardoulis et al., "In vivo evaluation of a novel, wrist-mounted arterial pressure sensing device versus the traditional hand-held tonometer," *Med. Eng. Phys.*, 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," *Front. Physiol.*, vol. 12, Mar. 2022, Art. no. 808451. [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," *Analyst*, vol. 135, no. 5, 2010, Art. no. 1138. [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," *Physiol. Meas.*, 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 J. Biomed. Health Inform.*, vol. 19, no. 3, pp. 832–838, May 2015. [Online]. Available: <https://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," *Comput. Biomed. Res.*, vol. 28, no. 2, pp. 154–170, Apr. 1995. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0010480985710117>
- [28] A. H. Hafifah and N. N. Anuar, "Methods of extracting feature from photoplethysmogram waveform for non-invasive diagnostic applications," *Int. J. Online Eng.*, vol. 16, no. 09, Aug. 2020, Art. no. 39. [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 *Proc. IEEE 2nd Int. Conf. Electron. Comput. Technol.*, 2010, pp. 165–168.
- [30] B. Liu et al., "The assessment of autonomic nervous system activity based on photoplethysmography in healthy young men," *Front. Physiol.*, vol. 12, Sep. 2021, Art. no. 733264. [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 Trans. Biomed. Eng.*, vol. 64, no. 4, pp. 859–869, Apr. 2017. [Online]. Available: <https://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. 44963–44973, 2021. [Online]. Available: <https://ieeexplore.ieee.org/document/9374974/>
- [33] M. Barandas et al., "TSFEL: Time series feature extraction library," *SoftwareX*, vol. 11, Jan. 2020, Art. no. 100456. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S2352711020300017>
- [34] M. Elgendi, "Optimal signal quality index for photoplethysmogram signals," *Bioengineering*, vol. 3, no. 4, Sep. 2016, Art. no. 21. [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," *Physiol. Meas.*, vol. 41, no. 12, Dec. 2020, Art. no. 125009. [Online]. Available: <https://iopscience.iop.org/article/10.1088/1361-6579/abc8dd>
- [36] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005. [Online]. Available: <https://ieeexplore.ieee.org/document/1453511/>
- [37] A. Hassan, J. H. Paik, S. Khare, and S. A. Hassan, "PPFS: Predictive permutation feature selection," 2021, *arXiv:2110.10713*. [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," *Front. Genet.*, vol. 10, Jun. 2019, Art. no. 524. [Online]. Available: <https://www.frontiersin.org/article/10.3389/fgene.2019.00524/full>
- [39] C. F. Aliferis, I. Tsamardinos, and A. Statnikov, "HITON: A novel Markov blanket algorithm for optimal variable selection," in *Proc. AMIA Annu. Symp. Proc.*, 2003, pp. 21–25.
- [40] J. Wainer and G. Cawley, "Nested cross-validation when selecting classifiers is overzealous for most practical applications," *Expert Syst. Appl.*, vol. 182, no. 115222, pp. 1–10, 2021, doi: [10.1016/j.eswa.2021.115222](https://doi.org/10.1016/j.eswa.2021.115222).
- [41] *IEEE standard for wearable cuffless blood pressure measuring devices*, Standard IEEE 1708-2014, 2014. [Online]. Available: <https://ieeexplore.ieee.org/document/6882122/>
- [42] "American national standards for electronic or automated sphygmomanometers," Assoc. Adv. Med. Instrum., Tech. Rep. ANSI/AAMI SP 10-1987, 1987. [Online]. Available: <https://cir.nii.ac.jp/crid/1572261550560818560>
- [43] E. O'Brien et al., "Short report: An outline of the revised British hypertension society protocol for the evaluation of blood pressure measuring devices," *J. Hypertension*, vol. 11, no. 6, pp. 677–679, Jun. 1993. [Online]. Available: <http://journals.lww.com/00004872-199306000-00013>
- [44] V. M. Panaretos and Y. Zemel, "Statistical aspects of Wasserstein distances," *Annu. Rev. Statist. Appl.*, vol. 6, pp. 405–431, Mar. 2019.
- [45] M. Cuturi, "Sinkhorn distances: Lightspeed computation of optimal transport," in *Proc. Annu. Conf. Neural Inf. Process. Syst.*, 2013, pp. 1–9.
- [46] 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. 19, pp. e127–e248, Jun. 2018, doi: [10.1161/HYP.0000000000000065](https://doi.org/10.1161/HYP.0000000000000065).
- [47] M. M. R. K. Mamun and A. T. Alouani, "Cuffless blood pressure measurement using linear and nonlinear optimized feature selection," *Diagnostics*, vol. 12, no. 2, Feb. 2022, Art. no. 408. [Online]. Available: <https://www.mdpi.com/2075-4418/12/2/408>
- [48] 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," *J. Med. Imag. Health Inform.*, vol. 8, no. 6, pp. 1290–1299, 2018. [Online]. Available: <http://www.ingentaconnect.com/content/10.1166/jmihi.2018.2474>
- [49] F. Miao et al., "A novel continuous blood pressure estimation approach based on data mining techniques," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 6, pp. 1730–1740, Nov. 2017. [Online]. Available: <https://ieeexplore.ieee.org/document/7914684/>
- [50] 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," *Sci. Rep.*, vol. 7, no. 1, Sep. 2017, Art. no. 11554. [Online]. Available: <https://www.nature.com/articles/s41598-017-11507-3>
- [51] 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 J. Biomed. Health Inform.*, vol. 26, no. 8, pp. 3697–3707, Aug. 2022. [Online]. Available: <https://ieeexplore.ieee.org/document/9769903/>
- [52] P. M. Granitto, C. Furlanello, F. Biasioli, and F. Gasperi, "Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products," *Chemometrics Intell. Lab. Syst.*, vol. 83, no. 2, pp. 83–90, Sep. 2006. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0169743906000232>