Original Paper

Blood Pressure Estimation for Hemodynamic Monitoring Using Reduced Photoplethysmography Features in Intensive Care Units: Prospective Machine Learning Study

Jieun Lee[1,2], PhD; Saurabh Ashok Sujanyal[1], MBBS; Surbhi Dadwal[1], MBBS; Elcin Omercikoglu[1], MD; Heba M Alazab[1], MD; Megha Shashidhar Handral[1], MBBS; Mohammed Naveed Shariff[2,3], BTECH; Muhammed Asil Nechooli[4], BE; Saravanabhavan Chandrasekar[4], BE; Suganti Shivaram[3], MBBS; Vidhya C M[3], PhD; Shivaram P Arunachalam[1,2,3], PhD, DBA; Scott A Helgeson[1,2,5], MD

[1]Department of Medicine, Division of Pulmonary, Allergy, and Sleep Medicine, Mayo Clinic in Florida, Jacksonville, FL, United States

[2]Department of Critical Care Medicine, Digital Engineering & Artificial Intelligence Laboratory (DEAL), Mayo Clinic in Florida, Jacksonville, FL, United States

[3]Patfoci Technologies Inc., Rochester, MN, United States

[4]Qmax Systems India Private Limited Inc., Chennai, Tamil Nadu, India

[5]Department of Critical Care Medicine, Mayo Clinic in Florida, Jacksonville, FL, United States

Corresponding Author:

Scott A Helgeson, MD
Department of Medicine
Division of Pulmonary, Allergy, and Sleep Medicine, Mayo Clinic in Florida
4500 San Pablo Road S.
Jacksonville, FL 32224
United States
Phone: 1 (904) 953-2000
Email: helgeson.scott@mayo.edu

Abstract

Background: Continuous, noninvasive blood pressure (BP) monitoring is clinically important for managing hemodynamic instability in the intensive care unit (ICU). However, current ICU practice primarily relies on invasive arterial catheterization. While this is the clinical gold standard for continuous measurement, it carries risks such as infection and thrombosis and is typically limited to select critical care settings.

Objective: To reduce reliance on invasive arterial lines, we developed an interpretable estimation framework for systolic blood pressure (SBP), mean arterial pressure (MAP), and diastolic blood pressure (DBP) from single-sensor photoplethysmography (PPG).

Methods: Synchronized PPG and arterial line 10-minute recordings from 100 patients in the ICU were analyzed. Using a hybrid feature selection strategy (RReliefF and intrinsic model–specific rankings), we reduced 201 initial features to an optimized subset of 30 core features. Model development and evaluation were designed to minimize information leakage and overfitting, incorporating nested cross-validation and patient-wise testing to assess generalizability across heterogeneous ICU populations. We developed and validated predictive models using machine learning techniques, specifically XGBoost (XGB), random forest, and support vector regression, with XGB serving as our primary model.

Results: Among the evaluated models, the Hybrid Top 30 XGB configuration provided the best balance between estimation performance, model compactness, and computational efficiency. The model achieved mean absolute errors of 8.61 mm Hg (SBP), 3.50 mm Hg (MAP), and 2.62 mm Hg (DBP). For MAP and DBP, mean difference (SD) values of –0.09 (4.54) mm Hg and –1.07 (3.51) mm Hg, respectively, satisfied the Association for the Advancement of Medical Instrumentation numerical criteria, while performance corresponded to Grade A according to the British Hypertension Society grading criteria. Although SBP exhibited greater variability, estimation performance was largely preserved using a physiologically informed subset of 30 features (approximately 15% of the original feature set). Feature importance analysis revealed distinct physiological feature patterns across SBP, MAP, and DBP estimation.

Conclusions: Interpretability analysis identified key predictors associated with wave reflection, systolic timing, low-frequency vasomotor activity, and autonomic regulation. By explicitly linking predictive features to underlying cardiovascular

physiology, the proposed framework addresses a key limitation of existing black-box cuffless BP models. Ultimately, this study demonstrates that an optimized XGB framework using physiologically informed PPG features can provide accurate and computationally efficient BP estimation in ICU settings and may support the assessment of clinically relevant hemodynamic states using a single-sensor PPG platform. These findings provide a foundation for future development of more robust and generalizable interpretable BP estimation models and cuffless monitoring technologies.

JMIR AI 2026;5:e105057; doi: 10.2196/105057

Keywords: blood pressure estimation; cuffless blood pressure monitoring; photoplethysmography; intensive care unit; feature engineering; feature selection; interpretable machine learning; hemodynamic monitoring; mean arterial pressure; hypotension

Introduction

Blood pressure (BP) monitoring is a cornerstone of patient care, with Mean Arterial Pressure (MAP) serving as a primary clinical indicator of organ perfusion and a critical threshold for intervention in the intensive care unit (ICU) [1]. While traditional cuff-based methods are widely used, they provide only intermittent readings, whereas invasive arterial catheterization, the clinical gold standard, offers continuous monitoring but carries risks such as infection and thrombosis and is typically limited to critical care environments [2]. As health care shifts toward patient-centered and continuous monitoring paradigms, there is a pressing need for noninvasive solutions that approach invasive-level accuracy without the associated clinical burden.

Photoplethysmography (PPG), which captures peripheral arterial pulse waveforms through light absorption, has emerged as a promising alternative for cuffless BP estimation. With the increasing availability of pulse oximetry sensors and wearable technologies, PPG offers a practical platform for scalable, continuous monitoring. Furthermore, recent literature emphasizes that BP variability, the frequent fluctuations in pressure often overlooked by traditional average readings, is a critical, independent predictor of multisystem health risks. Such dynamic swings can trigger hemodynamic stress, leading to vascular injury and increased mortality across cardiovascular, renal, and neurological systems [3]. Consequently, there is an urgent need to shift clinical practice toward monitoring these complex patterns using AI-driven diagnostic tools rather than relying solely on static snapshots. However, existing research has largely focused on hypertension detection [4,5], while the lower end of the physiological spectrum, hypotension, remains comparatively understudied. Clinically, detecting hypotensive shifts is critical, as acute reductions in MAP signal rapid loss of hemodynamic reserve and require urgent intervention. At the same time, estimating BP under hypotensive conditions is technically challenging because diminished peripheral perfusion degrades PPG signal quality.

Traditional noninvasive estimation often relies on pulse transit time or pulse arrival time, requiring synchronized signals from both electrocardiogram and PPG [6]. However, recent efforts have shifted toward single-sensor PPG-based estimation to reduce sensor burden and improve wearable integration. As demonstrated in our preliminary work using publicly available datasets, AI-based frameworks can achieve high correlation with reference measurements by analyzing PPG morphology alone [7].

Current PPG-based BP estimation generally follows 2 methodologies: feature-based machine learning and end-to-end deep learning (DL), as described in prior studies [8,9]. Feature-based approaches use the original PPG signal and its derivatives (velocity plethysmogram and acceleration plethysmogram) to capture hemodynamic nuances. Studies using the MIMIC-II (Multiparameter Intelligent Monitoring in Intensive Care II) [10,11] and PPG-BP [12] datasets have successfully mapped multidomain features to BP values using models such as Gaussian process regression and support vector regression (SVR), often meeting validation standards set by the Association for the Advancement of Medical Instrumentation (AAMI) [13] and the British Hypertension Society (BHS) [14], as demonstrated in prior studies [15-17]. Complementing these, DL architectures, such as hybrid CNN-BiLSTM (convolutional neural network–bidirectional long short-term memory) models with attention mechanisms, have achieved exceptional accuracy (mean absolute error [MAE]<2 mm Hg) by bypassing manual feature engineering to capture complex spatial and temporal dependencies [8]. However, while DL models offer high performance, they often operate as “black boxes,” failing to clarify the key physiological factors driving BP fluctuations. Furthermore, their high computational complexity poses significant challenges for real-time deployment in resource-constrained clinical environments.

Recent prospective studies have demonstrated the clinical feasibility of commercially developed cuffless BP monitoring systems in ICUs and postoperative settings [18,19], while recent feature engineering investigations have shown that interpretable PPG-derived biomarkers can support BP assessment [20]. These studies demonstrate both the practical potential of cuffless monitoring and the value of physiologically meaningful signal representations. Nevertheless, the physiological basis underlying PPG-based BP estimation models remains incompletely understood, particularly in ICU populations. This knowledge gap remains particularly important because many clinically validated monitoring systems rely on proprietary algorithms that do not explicitly reveal the physiological basis underlying their predictions. Furthermore, identifying the most informative physiological features is essential for developing lightweight and computationally efficient models that can facilitate future real-time monitoring applications while maintaining clinical interpretability.

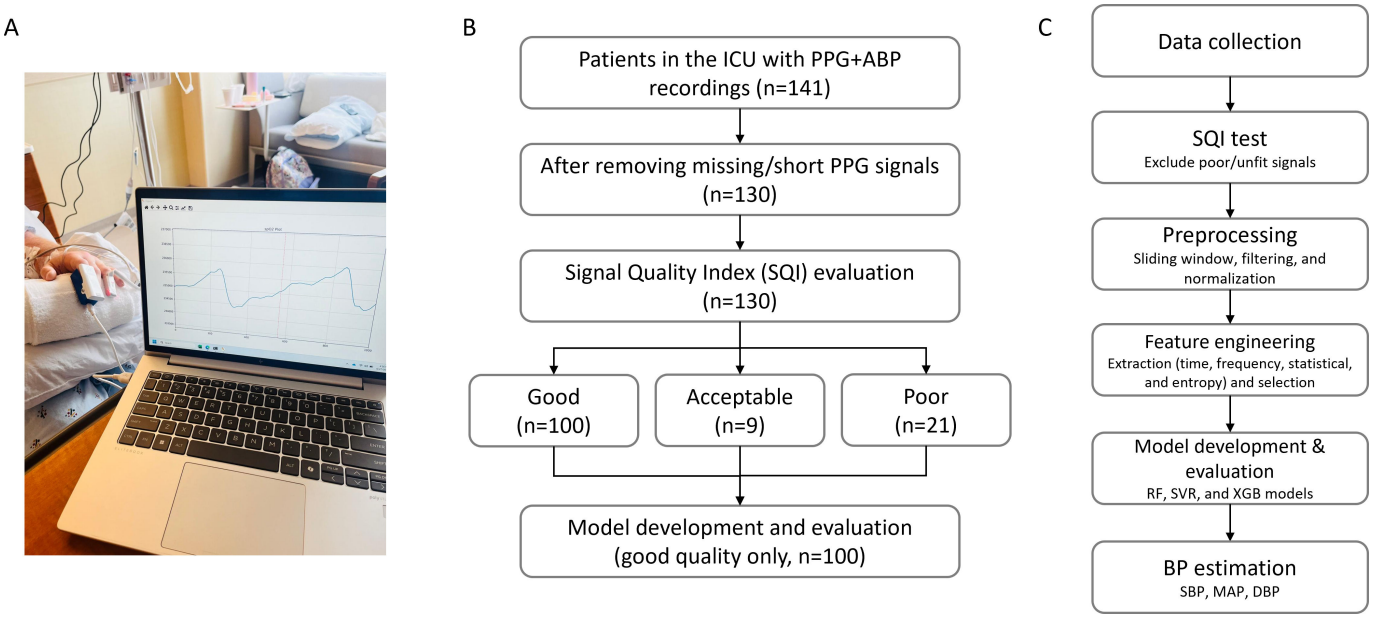
Despite these advancements, a significant challenge persists: current DL models are frequently too computationally demanding for real-time use and lack the interpretability required for clinical trust. Moreover, it remains a significant challenge to validate models against the extreme hemodynamic instability, ranging from profound hypotension to hypertensive crises, that is characteristic of the ICU population. To address these challenges, the aim of this study was to develop and evaluate an interpretable framework for systolic blood pressure (SBP), MAP, and diastolic blood pressure (DBP) estimation from single-sensor PPG signals in patients in the ICU. Specifically, we compared the performance of linear regression (LR), random forest (RF), SVR, and XGBoost (XGB) models and evaluated estimation accuracy, computational efficiency, and feature engineering strategies to identify an optimal model configuration for MAP, DBP, and SBP estimation. We further investigated whether a reduced set of physiologically informative PPG features could preserve estimation performance while improving model interpretability and computational efficiency.

Methods

Study Population and Overview

This study presents a framework for estimating SBP, MAP, and DBP from continuous PPG recordings obtained in patients in the ICU, paired with invasive arterial blood pressure (ABP) as the reference standard (Figure 1A). As illustrated in the study flowchart (Figure 1B), 141 ICU recordings were initially identified, of which 130 remained after excluding cases with missing or insufficient data. Following a multilevel Signal Quality Index (SQI) assessment, high-quality segments from a final cohort of 100 patients were selected for model development.

Figure 1. Study overview and computational workflow. (A) Clinical setup demonstrates the noninvasive photoplethysmography (PPG) device integrated with a real-time monitoring interface at the patient's bedside. (B) Flowchart of participant inclusion criteria, exclusion criteria, and signal quality stratification for the Mayo Clinic intensive care unit (ICU) cohort. (C) Computational pipelines including signal preprocessing, extraction of multidomain features, and model optimization using random forest (RF), support vector regression (SVR), and XGBoost (XGB) for the estimation of systolic blood pressure (SBP), mean arterial pressure (MAP), and diastolic blood pressure (DBP). ABP: arterial blood pressure.



The computational pipeline (Figure 1C) begins with the synchronized data collection of continuous PPG recordings and invasive ABP measurements. All PPG signals were first screened using a segment-level SQI to exclude unreliable data. Artifact-free signals then underwent detrending, filtering, normalization, and segmentation into 35-second windows with approximately 60% overlap.

From each segment, 201 physiologically relevant features spanning the time-domain, frequency-domain, statistical, and entropy-based domains were extracted. A hybrid feature selection strategy was then applied to reduce dimensionality from the 201-feature baseline to an optimized set of 30 predictors. Following feature reduction, regression models, including RF, SVR, and XGB, were trained and optimized using 5-fold nested cross-validation to estimate SBP, MAP, and DBP from unseen PPG segments.

Data Acquisition

This prospective study included 141 adult patients in the ICU at Mayo Clinic, Jacksonville, Florida, between February and September 2025. Eligible participants were ≥18 years of age with an indwelling arterial catheter.

For each participant, a continuous 10-minute PPG recording was obtained using a custom-developed data acquisition system. PPG signals were acquired using a custom pulse oximetry sensor with red and infrared optical channels. Signals were sampled at 120 Hz and recorded using SerialPlot software. As illustrated in Figure 1A, the device was secured

to the patient's index finger and interfaced with a dedicated laptop for real-time waveform visualization.

Simultaneously, reference SBP, MAP, and DBP measurements were obtained from a Philips IntelliVue MX500 bedside patient monitor (Philips) connected to a radial arterial catheter used for routine clinical care. The PPG sensor was positioned on the same side as the arterial catheter placement. Numerical BP values recorded by the monitoring system were extracted at 30-second intervals and synchronized with the corresponding PPG recordings for subsequent analysis. These monitor-derived numerical values served as the reference BP measurements throughout the study. Arterial-line setup, transducer leveling, and zeroing were performed according to standard ICU clinical protocols.

To increase the temporal density of the reference BP labels, linear interpolation was applied to generate BP values at 15 second intervals. Each 35-second PPG segment was paired with the corresponding BP reference value used for model training and evaluation. Prior to interpolation, numerical BP measurements underwent quality-control screening to identify missing or physiologically implausible values. Flagged observations were addressed using predefined quality-control procedures before subsequent analysis. Segments with missing data, physiologically implausible BP values, or poor PPG signal quality were excluded, yielding a final dataset of 3776 segments from 100 patients.

Ethical Considerations

This prospective study protocol was approved by the Mayo Clinic Institutional Review Board (IRB 24-001046). All methods were carried out in accordance with relevant guidelines and regulations, and written informed consent was obtained from all participants or their legally authorized representatives.

Signal Quality Assessment and Preprocessing

Preprocessing included detrending, filtering, and normalization. Segments were screened using an SQI to identify and exclude artifacts. Patient-level quality was determined from the proportion of acceptable and unfit segments across all windows, with only patients with high-quality data included in the analysis. A full description of the SQI methodology, segment exclusion criteria, and patient-level quality assignment is provided in the *Methods* section 1.1 in Multimedia Appendix 1.

Feature Extraction and Selection

Features were extracted across multiple domains: time-domain (morphology in PPG, velocity plethysmogram, and acceleration plethysmogram), frequency-domain (fast Fourier transform and power spectral density), statistical, and entropy-based. Multiscale features included multiscale entropy and multiscale frequency. Multiscale entropy was calculated by coarse-graining the PPG waveform at progressively increasing time scales and computing sample entropy at each scale, while multiscale frequency was estimated using instantaneous frequency obtained using the log-Gabor filter analysis. Empirical mode decomposition was applied to extract intrinsic mode functions (IMF1-IMF4), from which amplitude envelopes, instantaneous frequencies, relative energies, and fast Fourier transform–derived peak frequencies and amplitudes were computed. Beat-to-beat features were aggregated at the segment level using mean, SD, minimum, and maximum values, whereas heart rate variability measures summarize beat-to-beat variability. A comprehensive list of all features, along with their respective domains and physiological relevance, is provided in the *Methods* section 1.2 in Multimedia Appendix 1.

Feature selection was performed in 2 stages to optimize model efficiency and interpretability. First, to remove redundant features, an interfeature correlation analysis was conducted; for each pair with a Pearson correlation coefficient>0.995, one feature was removed, resulting in a baseline set of 201 features. Second, the RReliefF algorithm was used to identify the top 50 model-agnostic features, via the sklearn-relief library (n_feat=50). RReliefF is an extension of the Relief algorithm for regression tasks, weighting features based on differences between neighboring instances with respect to the continuous BP targets [21-23]. The top 50 weighted features were consistently applied during model training and evaluation. RReliefF ranking was performed using the original engineered feature representations without additional scaling. Feature standardization was subsequently applied during model training and evaluation through a StandardScaler-based preprocessing pipeline for all regression models. This 201-feature baseline and the RReliefF-derived subsets served as the foundation for the comparative performance scenarios evaluated in this study.

To evaluate the trade-off between dimensionality and accuracy, we investigated four experimental scenarios: (1) Full Baseline (n_feat=201), (2) Model-Specific Top 30 (baseline-derived), (3) RReliefF (n_feat=50), and (4) Hybrid Top 30 (RReliefF-derived). For all model-specific scenarios (2 and 4), feature importance rankings were generated using multiple model-based and SHAP (Shapley additive explanations)–based ranking approaches. High-confidence predictors were identified through a consensus-voting strategy, and the resulting target-specific hybrid feature subsets were used for subsequent model training and evaluation. Interpretability was further validated using SHAP and multimodel consensus voting (see *Methods* sections 1.3-1.4 in Multimedia Appendix 1).

Model Training and Evaluation

Linear (LR) and nonlinear (RF, SVR, and XGB) regression models were developed using a patient-wise split to prevent data leakage, with 80% of the cohort (n=80) allocated for training and 20% (n=20) for independent testing. We used patient-wise nested 5-fold cross-validation, where the inner loop was used for hyperparameter optimization via grid search and the outer loop was used for performance and stability assessment.

A subject-specific one-point calibration procedure was incorporated into the proposed framework to account for

intersubject variability in baseline BP levels. For each subject, a constant offset was determined using the difference between the first available reference BP measurement and the corresponding model–estimated BP value (reference BP – predicted BP). The resulting offset remained fixed throughout the recording without further recalibration. This calibration strategy was applied consistently during model development, validation, and testing.

Model selection was performed using nested cross-validation within the training cohort, whereas the held-out test cohort was exclusively reserved for final performance evaluation. Following patient-wise train-test partitioning, all feature selection, model optimization, and model development procedures were exclusively performed using the training cohort. An initial correlation-based redundancy filtering step was used to remove highly correlated features prior to model development.

Estimation performance was quantified using MAE, mean squared error, the coefficient of determination (R^2), and mean difference (MD) with SD (see *Methods* sections, Equations 2-6 in Multimedia Appendix 1). Computational efficiency was assessed by measuring inference latency at both the aggregate and per-segment levels. Finally, the clinical accuracy of the estimated BP was evaluated using numerical criteria derived from the AAMI and BHS evaluation frameworks.

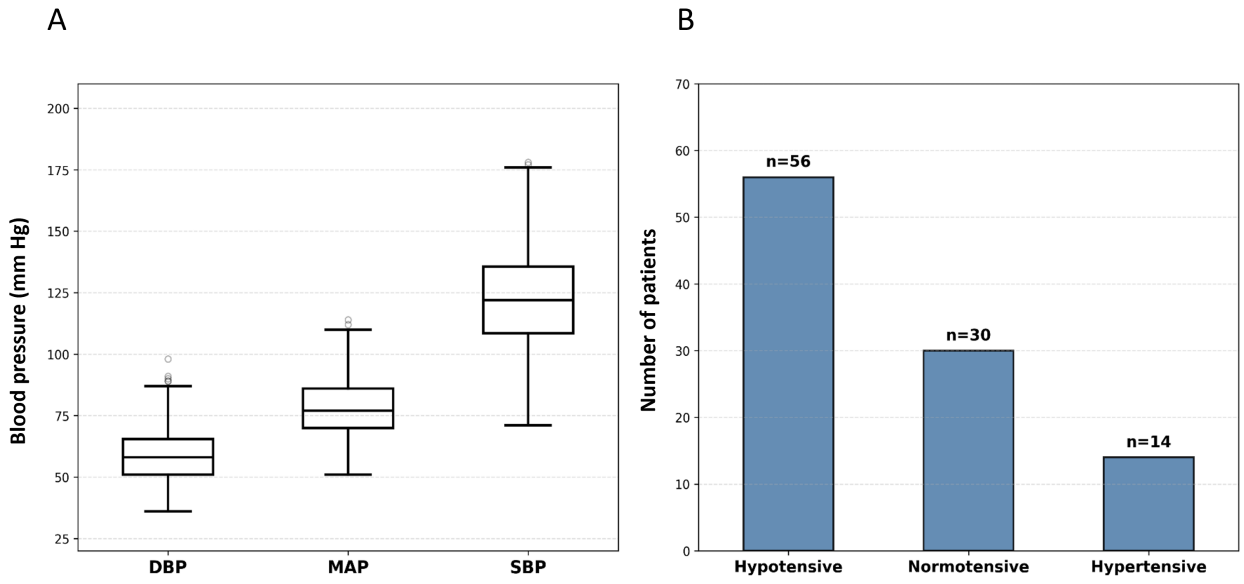
To evaluate hypotensive events, a hypotension episode was defined as MAP<65 mm Hg sustained for at least 1 minute [24]. Shorter periods of hypotension were not considered hypotensive episodes. This criterion was selected to reduce sensitivity to isolated threshold crossings and transient measurement variability while emphasizing hypotensive states of potential clinical relevance.

Results

Study Cohort BP Distribution

Following data quality filtering, the final analysis of the 100-patient cohort yielded reference BP measurements for distribution modeling. The distributions of invasive SBP, MAP, and DBP measurements are shown in Figure 2A, illustrating the broad hemodynamic range captured in the ICU setting.

Figure 2. Characterization of the patient cohort and reference blood pressure (BP) data. (A) Distribution of all invasive arterial line measurements for diastolic blood pressure (DBP), mean arterial pressure (MAP), and systolic blood pressure (SBP) across the study cohort (n=100 patients). (B) Patient distribution by BP category (n=100, per-patient mean), classified according to modified ACC/AHA (American College of Cardiology/American Heart Association) guidelines (detailed thresholds provided in Tables S2 and S3 in Multimedia Appendix 1).



Participants were categorized into hypotensive, normotensive, and hypertensive groups based on modified American College of Cardiology/American Heart Association adult guidelines and clinically relevant hypotension thresholds (Tables S2 and S3 in Multimedia Appendix 1) [25]. For patient-level categorization, mean SBP, DBP, and MAP values were calculated across the available recording period for each patient and used to determine BP category. Hypotensive was defined as SBP <90 mm Hg, DBP <60 mm Hg, or MAP< 65 mm Hg [24]. As shown in Figure 2B, the majority of the cohort was classified as hypotensive. To further characterize hemodynamic diversity, the number of patients meeting individual (SBP, MAP, and DBP) or combined hypotension criteria was quantified.

Feature Selection Framework and Model Architecture

Given the broad physiological range observed in the ICU cohort, we designed an experimental framework to identify the optimal balance between computational efficiency and predictive accuracy. We evaluated model performance across four feature selection scenarios: (1) Full Baseline (n_feat=201), (2) Model-Specific Top 30 (baseline-derived), (3) RReliefF (n_feat=50), and (4) Hybrid Top 30

(RReliefF-derived). The RReliefF algorithm was used as a model-agnostic filter to identify features with strong local dependencies and nonlinear associations with BP targets [21-23]. Scenario (2) was constructed by directly selecting the top 30 features based on the highest integrated importance rankings from RF and XGB models trained on the full 201-feature set.

In contrast, the "Hybrid Top 30" configuration (scenario 4) used a 2-stage selection process: initially, 50 features were preselected using the RReliefF algorithm to ensure robust clinical relevance independent of model architecture. Subsequently, the final 30 features were identified from these candidates using the same integrated model-importance criteria. This hybrid approach ensures that the final markers are both statistically significant through RReliefF and highly predictive across nonlinear architectures, effectively mitigating the risk of overfitting inherent in high-dimensional clinical data.

Model Optimization and Internal Validation

To ensure generalizability, we evaluated model performance using a nested 5-fold cross-validation scheme during the internal validation phase, allowing for unbiased hyperparameter tuning and feature set comparison. To examine the impact of dimensionality reduction, we compared the baseline performance (n_feat=201) against models trained on the preselected 50-feature RReliefF subset (scenario 1 vs 3).

Although reducing the feature set by 75% (201 to 50 features) did not yield a statistically significant change in MAE, it resulted in more stable and efficient models. As shown in Table 1, for MAP prediction, SVR and XGB maintained robust performance (MAE: 3.11, SD 0.92 and 3.81, SD 0.24 mm Hg, respectively) despite the drastically reduced input space. Initial evaluations included LR as a baseline; however, LR exhibited extreme sensitivity to signal variations, resulting in substantial outliers and negative R^2 values during testing (Table 1). This confirms that linear approaches lack the robustness required to capture the complex, nonlinear dynamics of PPG-BP signals. A detailed comparison of relief 50-feature model with the 201-feature models is provided in Table S5 in Multimedia Appendix 1. Based on these findings, the remainder of the study focused on nonlinear ensemble and kernel-based methods (RF, SVR, and XGB).

Table 1. Internal validation performance for mean arterial pressure prediction[a].

Model	Train, mean (SD)			Test, mean (SD)			
	MSE[b]	MAE[c]	R^2[d]	MSE	MAE	R^2	MD[e]
Full Baseline (n_feat=201)							
LR[f]	56.09 (5.55)	5.79 (0.33)	0.54 (0.05)	3488.75 (5300.88)	11.04 (5.46)	–25.81 (40.49)	–2.33 (59.03)
RF[g]	5.96 (2.26)	1.59 (0.4)	0.95 (0.02)	34.69 (10.21)	4.3 (0.75)	0.67 (0.19)	–1.28 (5.75)
SVR[h]	17.13 (2.75)	1.65 (0.38)	0.86 (0.03)	23.78 (16.63)	3.25 (1.31)	0.74 (0.27)	–0.36 (4.87)
XGB[i]	32.48 (18.32)	4.28 (1.52)	0.73 (0.16)	28.18 (7.49)	3.87 (0.57)	0.73 (0.15)	–1.38 (5.13)
RReliefF (n_feat=50)							
LR	97.04 (2.76)	7.8 (0.09)	0.2 (0.02)	12,111.11 (22,420.5)	11.17 (11.44)	–91.71 (171.73)	–5.14 (109.95)
RF	15.62 (3.07)	2.91 (0.26)	0.87 (0.03)	43.08 (12.57)	4.61 (0.56)	0.6 (0.18)	–1.14 (6.46)
SVR	19.67 (12.51)	1.78 (0.99)	0.84 (0.10)	21.49 (14.72)	3.11 (0.92)	0.81 (0.13)	–0.48 (4.62)
XGB	52.05 (23.67)	5.48 (1.64)	0.57 (0.19)	28.71 (6.29)	3.81 (0.24)	0.74 (0.08)	–0.84 (5.29)

[a]Values represent the mean (SD) across the nested 5-fold cross-validation for Full Baseline (n_feat=201) and RReliefF (n_feat=50) feature subsets.
[b]MSE: mean squared error.
[c]MAE: mean absolute error.
[d]R^2: coefficient of determination.
[e]MD: mean difference.
[f]LR: linear regression.
[g]RF: random forest.
[h]SVR: support vector regression.
[i]XGB: XGBoost.

Hyperparameter Tuning

Hyperparameters were optimized for each model and feature set within the inner loops of the nested cross-validation framework. For RF, SBP required deeper trees (max depth=20) when using the full feature set, while MAP and DBP were effectively modeled with shallower trees (max depth=10). The use of preselected RReliefF features allowed for a reduction in complexity, maintaining a max depth of 10 across all targets. SVR tuning revealed that reduced feature sets permitted higher *gamma* (γ) values (eg, increasing from 0.1 to 10 for SBP), enhancing the model's ability to capture nonlinear patterns in a lower-dimensional space without overregularization. XGB consistently used shallow trees (max depth=3) and a low *learning rate* (η) of 0.01 to ensure generalization; the only exception was SBP with preselected features, which benefited from a max depth of 9. These optimized selections balanced model complexity and feature dimensionality, ensuring stable performance while mitigating overfitting. Full hyperparameter details are provided in Table S6 in Multimedia Appendix 1.

Final Model Selection and Performance Evaluation

Following the identification of the optimal hyperparameters, performance was evaluated on the test cohort to assess generalization to unseen participants. We assessed generalization capability across 4 previously defined experimental scenarios (Table 2), ranging from the full 201-feature baseline to the optimized 30-feature subsets.

Table 2. Performance summary of the selected Hybrid Top 30 XGBoost model on the test cohort[a].

Target	MAE[b]	RMSE[c]	R^2[d]	Bias (SD, mm Hg)	95% LoA[e] (mm Hg)	BHS[f]
SBP[g]	8.61	11.75	0.56	–3.93 (11.07)	–25.64 to 17.77	Grade D
MAP[h]	3.50	4.54	0.80	–0.09 (4.54)	–8.99 to 8.80	Grade A
DBP[i]	2.62	3.67	0.85	–1.07 (3.51)	–7.95 to 5.82	Grade A

[a]Results obtained following one-point calibration.
[b]MAE: mean absolute error.
[c]RMSE: root mean squared error.
[d]R^2: coefficient of determination.
[e]LoA: limits of agreement.
[f]BHS: British Hypertension Society.
[g]SBP: systolic blood pressure.
[h]MAP: mean arterial pressure.
[i]DBP: diastolic blood pressure.

Performance was evaluated across the 4 previously defined feature selection scenarios on the test cohort. The RReliefF-derived feature subsets substantially reduced dimensionality while preserving estimation performance. The 50-feature RReliefF subset maintained performance comparable to the full 201-feature baseline, indicating that the selected features retained the most relevant physiological information while removing redundant variables (Table S7 in Multimedia Appendix 1).

Further refinement using model-specific feature importance identified the Hybrid Top 30 XGB configuration as a compact and robust solution while maintaining strong estimation performance. Although the feature space was reduced to 15% of the original feature set, no substantial decline in estimation accuracy was observed. In contrast, nontree-based approaches, particularly SVR, exhibited greater sensitivity to aggressive dimensionality reduction, whereas XGB maintained stable performance following feature reduction. Although the 50-feature SVR achieved lower MAEs for several BP targets, the Hybrid Top 30 XGB configuration maintained stable performance following feature reduction and demonstrated favorable agreement with invasive reference measurements while using a substantially smaller feature set (Tables S7-S10 in Multimedia Appendix 1). Accordingly, the Hybrid Top 30 XGB model was selected as the final model, and a performance summary on the test cohort is provided in Table 2.

Clinical Validation and Performance Standards

Figure 3 summarizes the estimation performance of the proposed XGB model on the test cohort. Results are shown after the application of the patient-specific one-point calibration procedure described in the *Methods* section. As shown in Figure 3A, representative time-series predictions demonstrate patient-level BP estimation performance across a range of physiological states and BP levels. Each vertical dashed line indicates a transition between the 20 different patients in the ICU. Despite the inherent variability in individual baselines, the model's ability to estimate BP variations using a fixed offset suggests that the selected morphological features capture generalized hemodynamic information across patients. The dataset encompasses a broad range of values recorded by invasive arterial line (A-line), including low-pressure states (SBP <90 mm Hg, DBP <60 mm Hg, or MAP <65 mm Hg) and elevated pressure ranges (SBP ≥140 mm Hg and DBP ≥90 mm Hg). While these ranges are broader than typical noninvasive cuff-based measurements (eg, American College of Cardiology/American Heart Association guidelines), they represent temporal BP variations observed during invasive monitoring in a clinical environment.

Figure 3. Performance and agreement analysis of the proposed XGBoost model on the test cohort for blood pressure (BP) estimation. (A) Representative time series illustrating patient-level BP estimation performance; vertical dashed lines indicate transitions between individual intensive care unit (ICU) patients. (B) Scatter plots demonstrate the correlation between predicted and reference BP values. (C) Bland-Altman plots evaluate agreement between the reference and predicted BP measurements, with the mean of the reference and predicted values shown on the horizontal axis and the difference (reference – predicted) shown on the vertical axis. The gray dashed line denotes the mean difference (MD), and red dashed lines indicate the 95% limits of agreement (MD ± 1.96 SD). DBP: diastolic blood pressure; MAP: mean arterial pressure; SBP: systolic blood pressure.

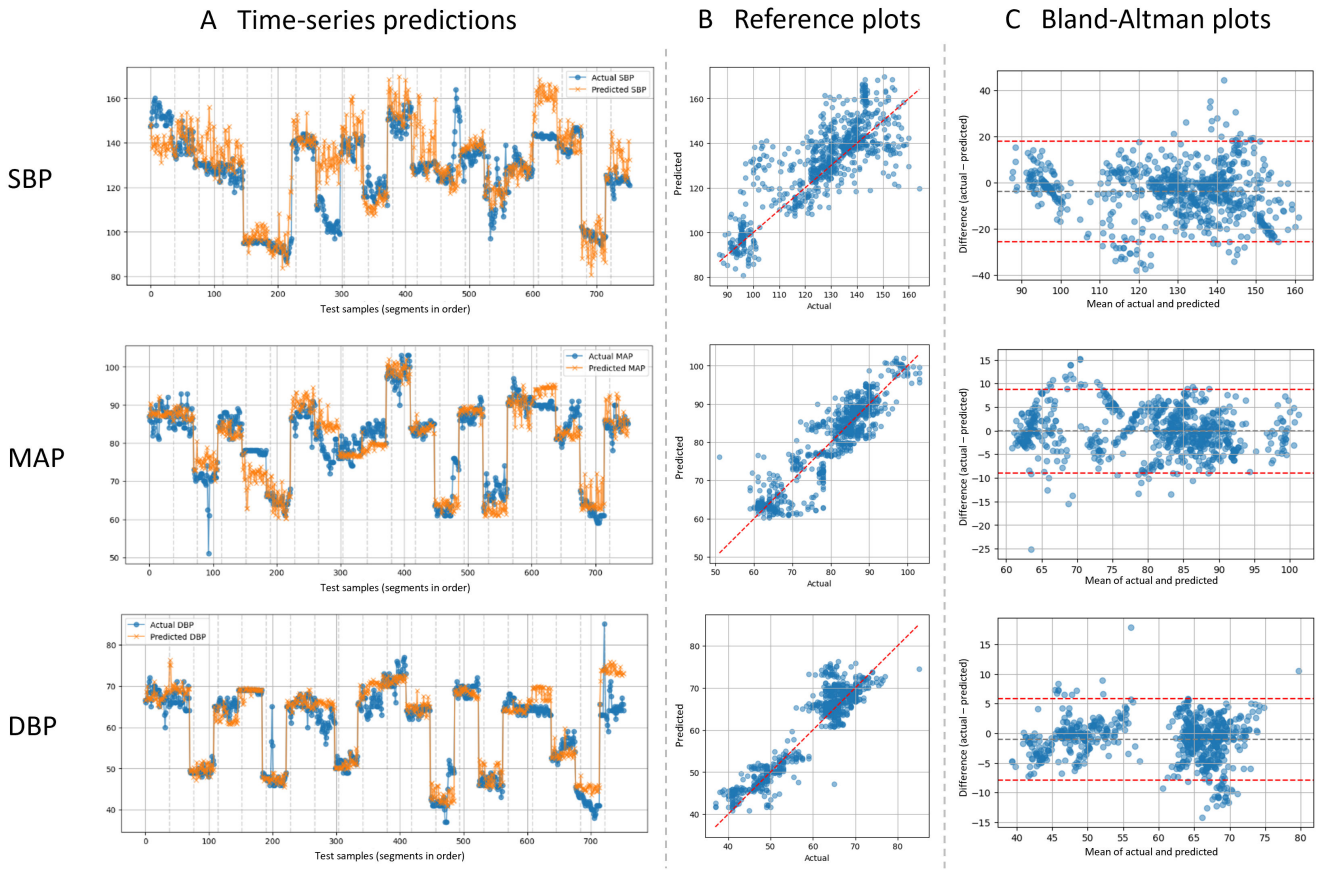


Figure 3B and C demonstrate strong agreement between predicted and invasive reference BP measurements (R^2=0.80-0.85 for MAP and DBP). The scatter plots show close correspondence between predicted and reference values, while the Bland-Altman plots indicate limited systematic bias and acceptable agreement across the evaluated BP targets. Although performance decreased slightly at extreme BP values, the model remained stable within clinically relevant pressure ranges, and the majority of observations fell within the 95% limits of agreement.

Clinical performance was further evaluated using numerical criteria derived from the AAMI and BHS evaluation frameworks [13,14]. As summarized in Table 2 and detailed in Tables S8-S10 in Multimedia Appendix 1, the proposed XGB model satisfied the AAMI numerical criteria (mean error ≤5 mm Hg, SD ≤8 mm Hg) for MAP and DBP and achieved grade A performance according to the BHS grading criteria (cumulative error percentages within 5, 10, and 15 mm Hg) for both targets. Although SBP exhibited lower agreement metrics, particularly with respect to variability, the mean error remained below the AAMI threshold. Overall, these findings support the selection of the Hybrid Top 30 configuration based on its performance, robustness to feature reduction, and model compactness in this ICU cohort. In a clinical setting, the accurate estimation of MAP and DBP may support the assessment of low BP states (eg, MAP <65 mm Hg), which are important for maintaining organ perfusion.

Hypotension Detection

Using the predefined episode criterion, 4 hypotensive episodes were identified in the test cohort. The proposed model successfully detected all 4 episodes while generating one false-positive hypotensive alert, corresponding to an episode-level sensitivity of 100% (4/4). Detailed episode-level results are provided in Table S11 in Multimedia Appendix 1.

Additional analyses of BP-change estimation performance are provided in Figure S5 in Multimedia Appendix 1. Bland-Altman analyses of successive BP changes (ΔBP) demonstrated minimal systematic bias across SBP, MAP, and DBP.

Model Interpretability and Feature Contribution

To evaluate the physiological consistency of the framework, we analyzed feature importance through a combination of SHAP and intrinsic model–specific rankings. Consensus

voting (Table S18 in Multimedia Appendix 1) identified a core set of high-confidence features supported by both SHAP attribution and internal model scoring. Key contributors included systolic reflection–related amplitudes (Ab/Aa) for SBP, IMF3 energy ratio for MAP, and standard deviation of normal-to-normal intervals for DBP. Based on prior physiological interpretations, these features may be associated with wave reflection, low-frequency vascular regulation, and autonomic control, respectively.

The identification of these physiologically relevant markers across disparate architectures suggests that the framework identified features that were consistently informative across modeling approaches. This alignment between attribution-based and intrinsic rankings confirms that the internal logic of the models is grounded in robust physiological signatures rather than dataset-specific artifacts.

Inference Time and Computational Stability

To evaluate computational cost, latency measurements were performed on the 20-patient test cohort, and the results are summarized in Table S19 in Multimedia Appendix 1. Each 10-minute recording yielded an average of 37.76 (SD 1) segments.

Preprocessing and feature extraction required an average of 4023.5 (SD 268.8) seconds per 10-minute recording, corresponding to 107.1 (SD 7.7) seconds per segment. Because the current implementation extracts the complete 201-feature set prior to application of the Hybrid Top 30 subset, this preprocessing time should be interpreted as a conservative upper-bound estimate rather than the expected runtime of an optimized deployment pipeline. Following preprocessing and feature extraction, XGB inference required only 20.82, 22.73, and 20.55 ms per approximately 10-minute recording for SBP, MAP, and DBP estimation, respectively.

Discussion

Principal Findings

This study developed and evaluated an interpretable framework for SBP, MAP, and DBP estimation using PPG-derived features, with XGB serving as the primary modeling approach. Our principal findings are that feature engineering strategies enabled substantial dimensionality reduction while preserving estimation performance, and that the resulting XGB framework provided accurate estimation of MAP and DBP using a compact set of interpretable PPG features. These results were achieved using a single-sensor PPG platform, reducing hardware complexity compared with multisensor approaches. We also identified substantial redundancy among the initial 201 PPG features: accuracy was maintained after reducing the feature set by 75% (to 50 features) using RReliefF preselection and further compressed to 30 features (about 15% of the original) using model-specific feature importance. This significant reduction enhances interpretability and supports computational efficiency and practical integration into ICU monitors and wearable devices, where low-latency inference is critical.

In clinical practice, MAP is the primary parameter used to guide hemodynamic interventions, as a threshold of <65 mm Hg typically triggers critical decisions regarding vasopressor initiation or fluid resuscitation [24]. While standard guidelines focus on intermittent cuff measurements, our use of a continuous A-line reference enabled evaluation across a broad range of BP values observed during clinical monitoring. Our results (both baseline and lightweight) achieved an MAE of 3.46 and 3.5 mm Hg for MAP, consistently falling below the 5-mm Hg AAMI threshold. The analysis of patient BP distributions revealed that hypotension in this cohort was most frequently characterized by low DBP. This reflects typical ICU physiology, where diastolic pressure may decrease due to vasodilation or sedation even when systolic pressure remains relatively preserved [26-28]. By providing high-precision estimation for both MAP and DBP, our framework may support the assessment of clinically relevant hemodynamic states while illuminating the underlying hemodynamic shifts in hypotensive patients.

Beyond conventional prediction accuracy (Figure S5 in Multimedia Appendix 1), additional analyses demonstrated that the proposed framework was able to capture temporal BP changes with minimal bias in ΔBP Bland-Altman analyses and successfully detected all hypotensive episodes identified in the test cohort (Table S11 in Multimedia Appendix 1). Although the present dataset was not designed to evaluate rapid hemodynamic perturbations or prospective early-warning performance, these findings suggest that the framework may provide useful support for hemodynamic monitoring and assessment of evolving physiological states.

Comparison With Prior Work and Clinical Implications

A comparative analysis against established benchmarks (Table 3) underscores the robustness of this framework within a real-world clinical context. While DL approaches (eg, Mohammadi et al [8]) report lower error rates on curated subsets, they typically use complex, multilayered architectures (CNN-BiLSTM) and auxiliary signals.

Table 3. Performance comparison of the proposed photoplethysmography (PPG)–based blood pressure (BP) estimation model against established literature benchmarks.

Authors	Data environment	Input signals	Segment size, patients, records, duration	Method	MAE[a] (mm Hg)		
					SBP[b]	MAP[c]	DBP[d]
Chowdhury et al [16] (2020)	PPG-BP: controlled/stable	PPG + discrete cuff-referenced BP	222 records, 126 patients	GPR[e]	3.02	—[f]	1.74
Nishan et al [17] (2024)	PPG-BP: controlled/stable	—	2.1 s, 657 PPG	SVR[g]	2.47	1.62	1.43
Kachuee et al [29] (2017)	MIMIC-II[h]: ICU[i]	PPG + ABP[j] + ECG[k]	3663 records, 1000 patients	DWT[l], AdaBoost, and RF[m]	11.17	5.92	5.35
Mohammadi et al [8] (2025)	MIMIC-II via UCI[n]: ICU	PPG + ABP + ECG	8.192 s (75% overlap), for 8 min	CNN-BiLSTM[o]	1.88	—	1.34
Our study (Hybrid Top 30)	ICU	Single-sensor PPG	35 s (~60% overlap), 10 min, 100 patients	XGB[p]	8.61	3.5	2.62

[a]MAE: mean absolute error.
[b]SBP: systolic blood pressure.
[c]MAP: mean arterial pressure.
[d]DBP: diastolic blood pressure.
[e]GPR: Gaussian process regression.
[f]Not available.
[g]SVR: support vector regression.
[h]MIMIC-II: Multiparameter Intelligent Monitoring in Intensive Care II.
[i]ICU: intensive care unit.
[j]ABP: arterial blood pressure.
[k]ECG: electrocardiogram.
[l]DWT: discrete wavelet transform.
[m]RF: random forest.
[n]UCI: University of California, Irvine.
[o]BiLSTM: bidirectional long short-term memory network.
[p]XGB: XGBoost.

In contrast, our Hybrid Top 30 XGB model achieved an MAE of 3.5 mm Hg for MAP and 2.62 mm Hg for DBP using only a single-sensor PPG. These results represent an advancement over previous ICU-focused research; for instance, our MAP estimation outperformed the results reported by Kachuee et al [29] (5.92 mm Hg) on the MIMIC-II dataset, despite our model relying solely on single-sensor PPG data without auxiliary electrocardiogram signals.

While studies using laboratory-controlled datasets (eg, PPG-BP [12]) or curated subsets (eg, University of California, Irvine-MIMIC [10]) report lower error rates, these "precleaned" environments often lack the signal artifacts and extreme hemodynamic shifts inherent to critical care. By achieving performance consistent with AAMI/BHS numerical criteria for MAP and DBP, our model demonstrates promising performance in this ICU cohort. Although SBP estimation remains a challenge due to the inherent variance of systolic peaks, the precision achieved for MAP and DBP remains clinically vital for guiding vasopressor and fluid resuscitation therapy.

Physiological Interpretation

The physiological interpretability of our models is summarized in Table 4, which categorizes key predictors by their physiological domain. Our results indicate that BP estimation is not driven by a single dominant feature; it relies on a distributed set of morphological and energy-specific descriptors that capture multifactorial hemodynamic determinants. Amplitude ratios (Ab/Aa, Ac/Aa) and systolic timing (Tsys) were consistently influential for SBP and MAP, as they may be associated with systolic strength and pulse-wave augmentation [30]. Conversely, DBP relied more heavily on diastolic decay (AUCdia) and autonomic regulation (standard deviation of normal-to-normal intervals), which may reflect autonomic regulatory processes [31]. Furthermore, IMF3-related energy features, which may reflect low-frequency vasomotor activity, emerged as important contributors to MAP estimation [32]. These findings align with previous Feat2Lab approaches [9], which emphasized the importance of elapsed times between PPG points of interest (eg, dicrotic notch and diastolic rise) and area-related or slope deviation curve features for robust SBP and DBP estimation. Similar physiological domains, including waveform morphology, spectral characteristics, area-related descriptors, and markers of arterial stiffness and wave reflection, were also identified in a recent interpretable PPG feature engineering study for BP-stage classification [20]. Although that study focused on classification, the overlap in selected feature domains supports their physiological relevance for BP assessment. In contrast, our regression framework enabled target-specific interpretation, revealing distinct physiological signatures for SBP, MAP, and DBP estimation.

Table 4. Summary of feature importance by physiological domain[a].

Physiological domain	Key features categories	Why important for BP[b] estimation	SBP[c]	MAP[d]	DBP[e]
Amplitude ratios and reflection strength	Amplitude ratios and reflection-related indices	May capture systolic strength and reflected-wave characteristics; associated with arterial stiffness and pulse-wave augmentation	Ab/Aa_sd, Ac/Aa_max, Asp_min	Ac/Aa_sd, Ab/Aa_max, Ac/Aa_mean	Ab/Aa_sd, Ac/Aa_sd, Aoff_sd
Systolic timing and PWV[f]	Timing and PWV-related metrics	Shorter systolic timing is associated with higher PWV and BP; timing variability may reflect vascular tone	Tsys_sd, deltaT_min	Tsys_sd, Tu_min, Tv_min	Tsys_sd, Ta_min
Wave reflection (derivatives)	VPG[g]-derived reflection indices	May capture reflected-wave characteristics associated with stiffness and peripheral resistance	—[h]	RIp1_mean, RIp1_min	RIp1_mean, RIp1_sd
EMD[i]-based energy and envelopes	EMD energy, FFT[j], and envelope metrics	IMF3=low-frequency vasomotor-related activity; IMF1/2=systolic waveform morphology; IMF4=respiratory/autonomic modulation	IMF2_FFT_Features_peak_amp	IMF3_Energy_Features_energy_ratio	IMF1_FFT_Features_peak_amp
Frequency-domain metrics	Spectral power and ratio metrics	Associated with autonomic regulation, respiratory modulation, and vascular resistance	psd_hf_power	psd_cr_ratio	psd_lf_hf_ratio
Area features	Area-under-curve metrics	May represent systolic and diastolic waveform energy associated with stroke volume and vascular resistance	AUCpi_mean	AUCpi_sd	AUCsys_mean, AUCdia_mean
Beat-to-beat variability/HRV[k]	HRV metrics	May reflect baroreflex and autonomic regulation associated with DBP variability	—	—	SDNN[l]

[a]Features listed represent the top contributors for each respective BP target (SBP, MAP, and DBP) based on SHAP analysis and consensus voting across models (the *Methods* sections 2.3-2.4 in Multimedia Appendix 1).
[b]BP: blood pressure.
[c]SBP: systolic blood pressure.
[d]MAP: mean arterial pressure.
[e]DBP: diastolic blood pressure.
[f]PWV: pulse wave velocity.
[g]VPG: velocity plethysmogram.
[h]Not available.
[i]EMD: empirical mode decomposition.
[j]FFT: fast Fourier transform.
[k]HRV: heart rate variability.
[l]SDNN: standard deviation of normal-to-normal intervals.

Recent prospective clinical studies have demonstrated that commercially developed cuffless BP monitoring systems can achieve clinically acceptable agreement with invasive arterial pressure measurements in ICU and postoperative populations [18,19]. These investigations provide important evidence supporting the feasibility of continuous noninvasive BP monitoring. Complementing these efforts, the present study focused on model development and interpretable feature discovery rather than the clinical performance evaluation of deployable cuffless BP monitoring systems. By identifying a compact set of physiologically meaningful PPG features capable of maintaining estimation performance, our findings provide insight into the physiological mechanisms underlying BP estimation and may support the future development of lightweight, transparent, and computationally efficient cuffless BP monitoring systems.

Practical Implementation Considerations

The reported preprocessing and feature extraction latency reflects the current implementation, which computes the complete set of 201 handcrafted features before applying Hybrid Top 30 feature selection. This exhaustive feature generation strategy was intentionally adopted to facilitate the broad exploration of physiologically relevant predictors across multiple feature domains and to identify the most informative biomarkers for SBP, MAP, and DBP estimation. Although computationally intensive, this process enabled systematic feature discovery and model optimization. Since only the final Hybrid Top 30 feature subset is required after model development, the reported latency should be regarded as a conservative upper-bound estimate. Future deployment may substantially reduce preprocessing time by directly extracting only the selected features while preserving the demonstrated estimation performance.

A key design consideration in this work was the trade-off between model simplicity and maximum achievable accuracy. Recent advances in multimodal and deep-learning-based approaches have demonstrated promising performance by leveraging additional physiological signals and more complex model architectures. Such approaches may be particularly valuable in patients who are critically ill and experiencing rapid hemodynamic changes or poor peripheral perfusion. However, increased model complexity may also reduce interpretability, increase computational requirements, and complicate implementation across diverse monitoring environments.

The present study therefore prioritized a single-sensor PPG framework combined with physiologically interpretable feature selection. This design choice was motivated by the goal of identifying robust physiological markers and developing a computationally efficient estimation framework suitable for practical implementation. While this simplicity may impose an upper limit on achievable accuracy compared with more complex multimodal systems, it provides important advantages in transparency, portability, and implementation feasibility.

Limitations

Limitations of this study include the handling of long-term temporal dependencies, limited population diversity, and reliance on high-quality PPG signals. As a proof-of-concept investigation, model development and evaluation were performed using recordings that met predefined signal-quality criteria to minimize the influence of severe artifacts. While this approach facilitated the assessment of the proposed methodology under controlled signal-quality conditions, it may limit the generalizability to real-world monitoring environments where signal quality is more variable. In addition, signal quality was assessed using the complete recordings rather than a dynamic real-time framework. While the model does not replicate every short-term BP oscillation in the ABP signal, it captures clinically relevant variations, including transitions between hypotensive and hypertensive states. Rather than indicating a lack of learning, this behavior reflects robustness to transient PPG artifacts, ensuring that predicted trends remain stable and actionable.

An additional limitation is the relatively short monitoring duration available for each patient (approximately 10 min). Consequently, the present dataset was not designed to evaluate long-term calibration stability, prolonged continuous monitoring performance, or responses to controlled hemodynamic interventions such as vasopressor adjustment or fluid administration. Although additional ΔBP and hypotension analyses were performed, longer recordings and intervention-driven BP changes will be required to fully assess dynamic monitoring performance.

Future work should investigate algorithm-level calibration-free BP estimation strategies, adaptive approaches for suboptimal signal quality, and the development of domain-invariant estimators to improve robustness across diverse clinical cohorts. Additional studies should also evaluate whether the physiological feature set identified in this work remains stable across external datasets, broader patient populations, and clinically deployed cuffless monitoring platforms [33].

Conclusions

Despite these limitations, the study demonstrates that carefully designed machine learning frameworks with multidomain, physiologically interpretable features can provide accurate and computationally efficient MAP and DBP estimation in ICU settings. Furthermore, in accordance with recent recommendations [34], our evaluation incorporates standardized error metrics to support clinical interpretability and comparability.

Ultimately, this study demonstrates that an optimized XGB framework using physiologically informed PPG features can provide accurate and computationally efficient BP estimation in ICU settings and may support the assessment of clinically relevant hemodynamic states using a single-sensor PPG platform. These findings provide a foundation for future development of more robust and generalizable interpretable BP estimation models and cuffless monitoring technologies.

Generative AI Use

The authors confirm that no generative AI tools were used to generate scientific content, analyses, results, figures, or conclusions.

Funding

This study was supported by the Mayo Clinic Gerstner Scholars Program in AI Translation and by the resources within the Digital Engineering & Artificial Intelligence Laboratory (DEAL), Department of Critical Care Medicine, Mayo Clinic, Jacksonville, FL, United States. No external funding was received.

Data Availability

The datasets generated and/or analyzed during the current study are not publicly available due to the inclusion of protected health information but are available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: SAH, SPA, JL
Data curation: SAH, JL, SAS, SD, EO, HMA, MH
Formal analysis: JL
Funding acquisition: SAH
Investigation: SAH, SPA, JL
Methodology: SAH, SPA, JL, MNS
Project administration: SAH, SPA
Resources: SAH, SPA, MNS, MAN, SC, SS, VCM
Supervision: SAH, SPA

Conflicts of Interest

None declared.

Multimedia Appendix 1

Supplementary methods and supporting analyses.
[DOCX File (Microsoft Word File), 2324 KB-Multimedia Appendix 1]

References

1. LeDoux D, Astiz ME, Carpati CM, Rackow EC. Effects of perfusion pressure on tissue perfusion in septic shock. Crit Care Med. Aug 2000;28(8):2729-2732. [doi: 10.1097/00003246-200008000-00007] [Medline: 10966242]
2. Lehman LH, Saeed M, Talmor D, Mark R, Malhotra A. Methods of blood pressure measurement in the ICU. Crit Care Med. Jan 2013;41(1):34-40. [doi: 10.1097/CCM.0b013e318265ea46] [Medline: 23269127]
3. Pugalenthi LS, Senapati SG, Gohri J, et al. Blood pressure variability (BPV) as a novel digital biomarker of multisystem risk and diagnostic insight: measurement, mechanisms, and emerging artificial intelligence methods. Biomedicines. Jan 30, 2026;14(2):317. [doi: 10.3390/biomedicines14020317] [Medline: 41751215]
4. Elgendi M, Fletcher R, Liang Y, et al. The use of photoplethysmography for assessing hypertension. NPJ Digit Med. 2019;2(1):60. [doi: 10.1038/s41746-019-0136-7] [Medline: 31388564]
5. Liang Y, Chen Z, Ward R, Elgendi M. Photoplethysmography and deep learning: enhancing hypertension risk stratification. Biosensors (Basel). Oct 26, 2018;8(4):101. [doi: 10.3390/bios8040101] [Medline: 30373211]
6. Kachuee M, Kiani MM, Mohammadzade H, Shabany M. Cuff-less high-accuracy calibration-free blood pressure estimation using pulse transit time. Proc IEEE Int Symp Circuits Syst (ISCAS). 2015:1006-1009. [doi: 10.1109/ISCAS.2015.7168806]
7. Shariff MN, Lee J, Sujanyal SA, et al. Cuffless photoplethysmogram based beat-beat AI powered continuous blood pressure estimation towards precision medicine. IEEE Int Conf Electro Inf Technol (eIT). 2026:270-275. [doi: 10.1109/eIT68936.2026.11670374]
8. Mohammadi H, Tarvirdizadeh B, Alipour K, Ghamari M. Cuff-less blood pressure monitoring via PPG signals using a hybrid CNN-BiLSTM deep learning model with attention mechanism. Sci Rep. Jul 1, 2025;15(1):22229. [doi: 10.1038/s41598-025-07087-2] [Medline: 40593051]
9. González S, Hsieh WT, Chen TPC. A benchmark for machine-learning based non-invasive blood pressure estimation using photoplethysmogram. Sci Data. Mar 21, 2023;10(1):149. [doi: 10.1038/s41597-023-02020-6] [Medline: 36944668]
10. UCI Irvine Machine learning repository. URL: https://archive.ics.uci.edu/ [Accessed 2026-10-24]

11. Goldberger AL, Amaral LA, Glass L, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation. Jun 13, 2000;101(23):E215-E220. [doi: 10.1161/01.cir.101.23.e215] [Medline: 10851218]
12. Liang Y, Liu G, Chen Z, Elgendi M. PPG-BP database. Figshare. 2018. URL: https://figshare.com/articles/dataset/PPG-BP_Database_zip/5459299 [Accessed 2021-04-06]
13. Stergiou GS, Alpert B, Mieke S, et al. A universal standard for the validation of blood pressure measuring devices: Association for the Advancement of Medical Instrumentation/European Society of Hypertension/International Organization for Standardization (AAMI/ESH/ISO) collaboration statement. Hypertension. Mar 2018;71(3):368-374. [doi: 10.1161/HYPERTENSIONAHA.117.10237] [Medline: 29386350]
14. O'Brien E, Waeber B, Parati G, Staessen J, Myers MG. Blood pressure measuring devices: recommendations of the European Society of Hypertension. BMJ. Mar 3, 2001;322(7285):531-536. [doi: 10.1136/bmj.322.7285.531] [Medline: 11230071]
15. Duan K, QianZ, Atef M, Guoxing W. A feature exploration methodology for learning based cuffless blood pressure measurement using photoplethysmography. Annu Int Conf IEEE Eng Med Biol Soc. Aug 2016;2016:6385-6388. [doi: 10.1109/EMBC.2016.7592189] [Medline: 28269709]
16. Chowdhury MH, Shuzan MNI, Chowdhury MEH, et al. Estimating blood pressure from the photoplethysmogram signal and demographic features using machine learning techniques. Sensors (Basel). Jun 1, 2020;20(11):3127. [doi: 10.3390/s20113127] [Medline: 32492902]
17. Nishan A, Taslim Uddin Raju SM, Hossain MI, et al. A continuous cuffless blood pressure measurement from optimal PPG characteristic features using machine learning algorithms. Heliyon. Mar 30, 2024;10(6):e27779. [doi: 10.1016/j.heliyon.2024.e27779] [Medline: 38533045]
18. Chen Y, Xu Z, Ma Y, et al. Clinical performance evaluation of a cuffless continuous non-invasive blood pressure monitoring device in ICU patients. Digit Health. 2026;12:20552076261456613. [doi: 10.1177/20552076261456613] [Medline: 42221509]
19. Hellou E, Jamal T, Zuroff E, et al. Performance of a cuffless photoplethysmography-based device for continuous monitoring of blood pressure after cardiac surgery: a preliminary validation study. J Hum Hypertens. Dec 2025;39(12):894-902. [doi: 10.1038/s41371-025-01082-w] [Medline: 41109863]
20. Msokar S, Davydov R, Davydov V. Interpretable photoplethysmography feature engineering for multi-class blood pressure staging. Computers. 2026;15(4):209. [doi: 10.3390/computers15040209]
21. Kira K, Rendell LA. A practical approach to feature selection. In: Sleeman D, Edwards P, editors. Machine Learning Proceedings 1992. Morgan Kaufmann; 1992:249-256. [doi: 10.1016/B978-1-55860-247-2.50037-1]
22. Kira K, Rendell LA. The feature selection problem: traditional methods and a new algorithm. In: AAAI'92: Proceedings of the Tenth National Conference on Artificial Intelligence. AAAI Press; 1992:129-134. URL: https://dl.acm.org/doi/10.5555/1867135.1867155 [Accessed 2026-10-24]
23. Robnik-Šikonja M, Kononenko I. An adaptation of relief for attribute estimation in regression. Presented at: 14th International Conference on Machine Learning (ICML 1997); Jul 8-12, 1997; Vanderbilt University, Nashville, Tennessee, USA. URL: https://mlanthology.org/icml/1997/robniksikonja1997icml-adaptation/ [Accessed 2026-10-24]
24. Evans L, Rhodes A, Alhazzani W, et al. Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock 2021. Crit Care Med. Nov 1, 2021;49(11):e1063-e1143. [doi: 10.1097/CCM.0000000000005337] [Medline: 34605781]
25. Writing Committee Members*, Jones DW, Ferdinand KC, et al. 2025 AHA/ACC/AANP/AAPA/ABC/ACCP/ACPM/AGS/AMA/ASPC/NMA/PCNA/SGIM 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 Joint Committee on Clinical Practice Guidelines. Circulation. Sep 16, 2025;152(11):e114-e218. [doi: 10.1161/CIR.0000000000001356] [Medline: 40811497]
26. Zhao J, Lin ST, Qin AH, et al. Impact of diastolic blood pressure time under range on mortality and acute kidney injury in septic patients: a retrospective cohort study. BMC Anesthesiol. Oct 21, 2025;25(1):512. [doi: 10.1186/s12871-025-03382-7] [Medline: 41120869]
27. Gao Z, Li C, Chen H, et al. Association between diastolic blood pressure during the first 24 h and 28-day mortality in patients with septic shock: a retrospective observational study. Eur J Med Res. Sep 9, 2023;28(1):329. [doi: 10.1186/s40001-023-01315-z] [Medline: 37689707]
28. Cinel I, Kasapoglu US, Gul F, Dellinger RP. The initial resuscitation of septic shock. J Crit Care. Jun 2020;57:108-117. [doi: 10.1016/j.jcrc.2020.02.004] [Medline: 32135409]
29. Kachuee M, Kiani MM, Mohammadzade H, Shabany M. Cuffless blood pressure estimation algorithms for continuous health-care monitoring. IEEE Trans Biomed Eng. Apr 2017;64(4):859-869. [doi: 10.1109/TBME.2016.2580904] [Medline: 27323356]

30. Nichols WW, Denardo SJ, Wilkinson IB, McEniery CM, Cockcroft J, O'Rourke MF. Effects of arterial stiffness, pulse wave velocity, and wave reflections on the central aortic pressure waveform. J Clin Hypertens (Greenwich). Apr 2008;10(4):295-303. [doi: 10.1111/j.1751-7176.2008.04746.x] [Medline: 18401227]
31. Sanders JS, Ferguson DW. Diastolic pressure determines autonomic responses to pressure perturbation in humans. J Appl Physiol (1985). Feb 1989;66(2):800-807. [doi: 10.1152/jappl.1989.66.2.800] [Medline: 2708209]
32. Takalo R, Korhonen I, Turjanmaa V, Majahalme S, Tuomisto M, Uusitalo A. Frequency shift in baroregulatory oscillation in borderline hypertensive subjects. Am J Hypertens. May 1997;10(5 Pt 1):500-504. [doi: 10.1016/s0895-7061(96)00490-6] [Medline: 9160759]
33. Joshi DD, Saraboji K, Shaik AT, et al. Artificial intelligence enabled diagnostics using photoplethysmography (PPG): beyond SpO2 and heart rate. Life (Basel). Sep 15, 2026;16(9):1530. [doi: 10.3390/life16091530] [Medline: 42795395]
34. Elgendi M, Haugg F, Fletcher RR, et al. Recommendations for evaluating photoplethysmography-based algorithms for blood pressure assessment. Commun Med (Lond). Jul 12, 2024;4(1):140. [doi: 10.1038/s43856-024-00555-2] [Medline: 38997447]

Abbreviations

AAMI: Association for the Advancement of Medical Instrumentation
ABP: arterial blood pressure
BHS: British Hypertension Society
BP: blood pressure
CNN-BiLSTM : convolutional neural network–bidirectional long short-term memory
DBP: diastolic blood pressure
DL: deep learning
ICU: intensive care unit
IMF: intrinsic mode function
LR: linear regression
MAE: mean absolute error
MAP: mean arterial pressure
MIMIC-II: Multiparameter Intelligent Monitoring in Intensive Care II
PPG: photoplethysmography
RF: random forest
SBP: systolic blood pressure
SHAP: Shapley additive explanations
SQI: Signal Quality Index
SVR: support vector regression
XGB: XGBoost

Edited by Ivan Steenstra; peer-reviewed by Elias Hellou, Roman Davydov; submitted 19.Jun.2026; final revised version received 07.Oct.2026; accepted 07.Oct.2026; published 17.Nov.2026

Please cite as:
Lee J, Sujanyal SA, Dadwal S, Omercikoglu E, Alazab HM, Handral MS, Shariff MN, Nechooli MA, Chandrasekar S, Shivaram S, C M V, Arunachalam SP, Helgeson SA
Blood Pressure Estimation for Hemodynamic Monitoring Using Reduced Photoplethysmography Features in Intensive Care Units: Prospective Machine Learning Study
JMIR AI 2026;5:e105057
URL: https://ai.jmir.org/2026/1/e105057
doi: 10.2196/105057

© Jieun Lee, Saurabh Ashok Sujanyal, Surbhi Dadwal, Elcin Omercikoglu, Heba M Alazab, Megha Shashidhar Handral, Mohammed Naveed Shariff, Muhammed Asil Nechooli, Saravanabhavan Chandrasekar, Suganti Shivaram, Vidhya C. M., Shivaram P Arunachalam, Scott A. Helgeson. Originally published in JMIR AI (https://ai.jmir.org), 17.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR AI, is properly cited. The complete bibliographic information, a link to the original publication on https://www.ai.jmir.org/, as well as this copyright and license information must be included.