

Original Article

Open Access



# Validation of deep learning models for cuffless blood pressure estimation on a large benchmarking dataset

Yu Huang<sup>1</sup>, Yonghu He<sup>1</sup>, Zhengbi Song<sup>1</sup>, Kun Gao<sup>2</sup>, Yali Zheng<sup>1\*</sup>

<sup>1</sup>Department of Biomedical Engineering, College of Health Science and Environmental Engineering, Shenzhen Technology University, Shenzhen 518118, China.

<sup>2</sup>Department of Critical Care Medicine, Jinling Hospital, Nanjing 210002, Jiangsu, China.

\*Correspondence to: Prof. Yali Zheng, Department of Biomedical Engineering, College of Health Science and Environmental Engineering, Shenzhen Technology University, Shenzhen 518118, China. E-mail: zhengyali@sztu.edu.cn

**How to cite this article:** Huang Y, He Y, Song Z, Gao K, Zheng Y. Validation of deep learning models for cuffless blood pressure estimation on a large benchmarking dataset. *Conn Health Telemed* 2024;3:300002. <https://dx.doi.org/10.20517/chatmed.2023.23>

**Received:** 7 Dec 2023 **First decision:** 30 Jan 2024 **Revised:** 22 Feb 2024 **Accepted:** 7 Mar 2024 **Published:** 19 Mar 2024

**Academic Editor:** Stefano Omboni **Copy Editor:** Yanbing Bai **Production Editor:** Yanbing Bai

## Abstract

**Objective:** This study aims to evaluate the effectiveness of deep learning techniques in estimating cuffless blood pressure (BP) across a diverse patient population in intensive care units (ICUs).

**Methods:** A comprehensive ICU benchmarking dataset encompassing 2,154 patients with a wide demographic range (18-97 years old) and varied cardiovascular status was employed to validate several deep learning models in predicting continuous BP waveforms. Three methods were developed to enhance the model's generalizability to this heterogeneous dataset. Ten-fold subject-independent cross-validation was performed and the model performance was assessed through mean absolute error (MAE), Pearson's correlation coefficient (PCC), and compared with significance analysis.

**Results:** The *UTransBPNet\_Demo\_In* model, which incorporated demographic and physiological signals as inputs, achieved a PCC of 0.89 and a MAE of 10.38 mmHg in predicting arterial BP waveforms, demonstrating the highest tracking capability among all models. Notably, the performance of *UTransBPNet\_Demo\_In* remained robust across variations in demographic and cardiovascular status.

**Conclusion:** The *UTransBPNet\_Demo\_In* model demonstrates robust predictive capabilities across a broad spectrum



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.



of demographics and cardiovascular conditions. Although the performance still needs further improvement, this study offers a benchmark in the field of cuffless BP monitoring in critical care settings for future studies.

**Keywords:** Cuffless blood pressure estimation, benchmark validation, demographic fusion, noninvasive intensive care monitoring

## INTRODUCTION

Cuffless blood pressure (BP) monitoring has great value in cardiovascular healthcare. Unlike traditional cuff-based methods, which can cause discomfort and are impractical for long-term monitoring, cuffless technology allows for continuous monitoring of BP changes without interrupting daily activities<sup>[1]</sup>. This continuous monitoring is particularly valuable for patients with hypertension and other cardiovascular conditions by providing a comprehensive picture of BP fluctuations in various situations<sup>[2]</sup>, thus facilitating early detection and timely intervention. Moreover, cuffless BP technology also empowers individuals to be aware of their BP variations and proactively manage their health<sup>[3]</sup>.

A plethora of machine and deep learning models for cuffless BP estimation have been proposed in recent years<sup>[4-9]</sup>; however, comparison of their performance is challenging as different models have been validated on diverse datasets. These models can be categorized into three types: subject-specific, population-based, and hybrid ones<sup>[1]</sup>. Subject-specific models, although accurate, require frequent calibrations and often rely on data from individuals, which may be impractical<sup>[4,10,11]</sup>. Furthermore, the lack of reported intra-subject BP variations in these studies raises concerns about potentially over-optimistic results. In contrast, population-based models, which do not need calibration, often sacrifice accuracy due to the complex relationship of the features and BP across diverse populations<sup>[12,13]</sup>. These models must consider basic demographic information in their design. The hybrid method seeks a balance between practicality and accuracy but has been limited by small datasets in previous validations<sup>[12-14]</sup>.

Emerging studies on cuffless BP estimation have validated various models on large-scale datasets. One study, using the VitalDB dataset, validated a pulse arrival time-based BP estimation model. They reported very satisfactory results<sup>[15]</sup>, which may be attributed to their focus solely on the accuracy of linear regression, without validation on an independent dataset. Another study proposed a lightweight deep learning model, KD-informer, and accurately predicted continuous BP on two separate datasets, but the intra-subject BP variations were not reported<sup>[8]</sup>. Additionally, a recent study assessed several machine learning models such as ridge regression, SVM, AdaBoost, and random forest, as well as deep learning methods including VGGNet16, ResNet50, BiLSTM, and ResLSTM, on a large-scale dataset gathered from 3,077 individuals without severe cardiovascular diseases using smartwatches<sup>[16]</sup>. They reported that the best-performing calibration-free model had estimation errors of  $-0.71 \pm 13.04$  and  $-0.29 \pm 8.78$  mmHg for SBP and DBP, respectively. However, the BP range of this literature was limited and they only tested under static conditions. Furthermore, their study highlighted significant performance degradation in models for aged ( $\geq 55$  years old) and hypertensive subgroups. One of our previous works proposed a high-performance deep learning model, *UTransBPNet*<sup>[17]</sup>. Due to its strong capability in short- and long-range feature representation due to the combined structure of Unet and Transformer, it demonstrated superior performance over popular models such as CNN-LSTM-attention<sup>[17]</sup> and CNN-BiGRU<sup>[18]</sup>, in scenarios with significant intra-subject BP variations. However, this validation was limited to a small, self-collected dataset with a narrow demographic range. The real-world effectiveness of *UTransBPNet*, particularly in clinical settings with diverse demographics and sufficient intra-subject BP variations, remains unknown.

Addressing this gap, Wang *et al.* recently proposed a large, cleaned dataset PulseDB with diverse demographics and sufficient intra-subject BP variations for benchmarking cuffless BP estimation methods<sup>[19]</sup>. Leveraging this resource, our study aims to validate the performance of several deep learning models in estimating arterial BP (ABP) across a diverse population in intensive care units (ICUs) and with sufficient intra-subject BP variations. We further explore three approaches to optimize the performance of *UTransBPNet*, incorporating demographic information and employing a hybrid methodology.

## METHODS

### Data

This study utilized the PulseDB dataset, which originates from the large-scale MIMIC-III and VitalDB datasets<sup>[19]</sup>, and includes several continuous physiological signals like electrocardiogram (ECG), photoplethysmogram (PPG), and arterial BP (ABP) acquired from ICU patients. To investigate the impact of demographic information on BP estimation, only the VitalDB dataset was used, as the MIMIC-III dataset did not include this information. A series of data preprocessing steps were conducted on the dataset as below:

*Band-pass filtering and resampling.* ECG, PPG, and ABP signals were filtered at 0.5-30 Hz, 0.5-15 Hz, and 0.5-15 Hz, respectively, and then resampled to 125 Hz.

*Normalization.* Maximum-minimum normalization was applied within each segment for ECG and PPG separately to scale the values to the range between 0 and 1.

*Removing segments with abnormal BP.* The segments with ABP values beyond the normal physiological range between 40 and 250 mmHg were excluded.

*Removing segments with abnormal ECG amplitudes.* The amplitude differences of adjacent ECG peaks within one segment were calculated, and those with a difference  $\leq 0.5$  were excluded.

*Removing segments with abnormal heart rate.* The segments with the R-R interval beyond the range of 0.4 to 2.4 seconds were removed.

*Excluding patients with short segment lengths.* The patients with a total duration of signal recording  $< 20$  min were excluded.

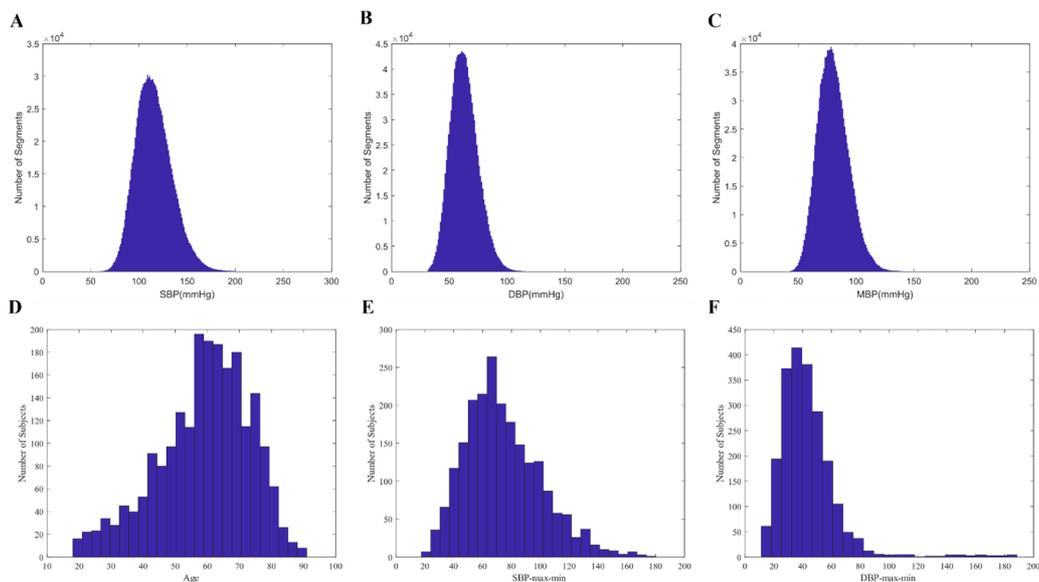
With preprocessing by these steps, a total of 1,257,141 10-sec signal segments of 2,154 patients finally remained in the following analysis. The demographic characteristics of these patients are presented in [Table 1](#). The dataset spans a diverse age range from 18 to 97 years old. [Figure 1](#) shows the histograms for systolic BP (SBP), diastolic BP (DBP), mean BP (MBP), the intra-subject max-min SBP and DBP variations (denoted as  $\Delta\text{SBP}_{\text{max-min}}$  and  $\Delta\text{DBP}_{\text{max-min}}$ , respectively) of the entire dataset. The dataset presents substantial intra-subject max-min SBP and DBP variations, ranging from 20 to 180 mmHg (mean  $\pm$  SD:  $74.2 \pm 26.3$ ) for SBP, and 20 to 180 mmHg ( $43.6 \pm 20.5$  mmHg) for DBP, respectively.

### Model

Two model structures proposed in our previous study were evaluated on this large dataset: one is the CNN and bidirectional LSTM network with attention mechanism (*CNN-LSTM-ATT*) and the other is *UTransBPNet*<sup>[17]</sup>. *UTransBPNet* comprises three primary modules: an enhanced U-net for capturing short-range features of the input signals, a transformer module for obtaining long-range contextual information,

**Table 1. Demographic and cardiovascular characteristics of the subjects included**

Characteristics	Values
Num of subject	2,154
Age, years old	Number (Proportion, %)
18-27	61 (2.83%)
28-37	120 (5.57%)
38-47	229 (10.63%)
48-57	441 (20.47%)
58-67	602 (27.95%)
68-77	495 (22.98%)
78-87	190 (8.82%)
88-97	16 (0.74%)
Female, %	55.2%
Male, %	44.8%
Height, cm (mean $\pm$ SD)	162.72 $\pm$ 8.70
Weight, kg (mean $\pm$ SD)	61.34 $\pm$ 11.42
BMI, kg/m <sup>2</sup> (mean $\pm$ SD)	23.09 $\pm$ 3.46

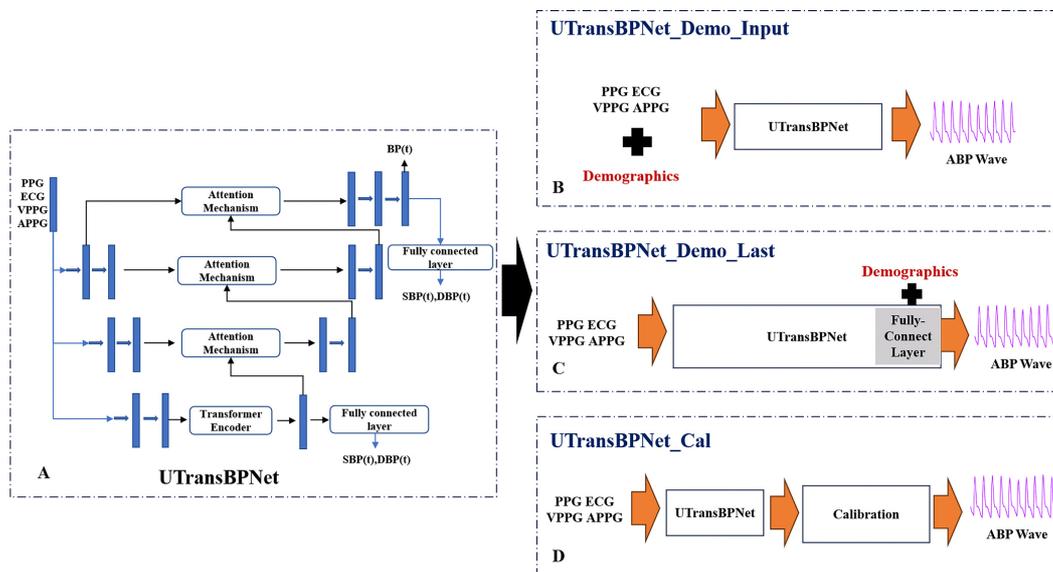


**Figure 1.** (A-C) Distributions of SBP, DBP, and MBP; (D-F) distributions of the subject age, and intra-subject max-min variations of SBP and DBP; The y-axis of (A-C) represents the number of 10-second signal segments, while the y-axis of (D-F) indicates the number of subjects across the entire dataset.

and a cross-attention module to further refine feature learning. Based on *UTransBPNet*, three methods were further introduced, as shown in [Figure 2](#):

*UTransBPNet* model: it uses ECG, PPG, and the first and second derivatives of PPG signals (VPPG and APPG) as inputs to estimate continuous ABP.

*UTransBPNet* model with individual calibration (*UTransBPNet\_Cal*): similar to *UTransBPNet*, *UTransBPNet\_Cal* incorporates calibration using the initial 20 seconds of individual data. Specifically, the BP bias that minimizes the difference between the reference and predicted BP during the initial 20 seconds was identified and used to calibrate the entire set of predictions for each subject.



**Figure 2.** Model structure. (A) *UTransBPNet*; (B) *UTransBPNet\_Demo\_In*; (C) *UTransBPNet\_Demo\_Last*; (D) *UTransBPNet\_Cal*.

*UTransBPNet* model with demographic inputs (*UTransBPNet\_Demo\_In*): it integrates demographic information along with the ECG and PPG as inputs. Each demographic feature was duplicated to match the length of the input signal. The pure feature values were used for the continuous variables such as age, height, weight, and BMI, while binary coding was used for discrete variables such as gender.

*UTransBPNet* fused with demographics (*UTransBPNet\_Demo\_Last*): the demographic features were directly concatenated with the deep features learned by the model at the last fully-connected layer of *UTransBPNet*.

**Training & validation.** The training and validation of all models were conducted on Pytorch using an Nvidia Tesla V100 graphics card with 32 GB of memory. During the training of each model, the hyperparameters were configured as follows. The learning rate was set as  $1e-5$  with a batch size of 64 to balance the trade-off between computational efficiency and the accuracy of gradient estimates. The training was scheduled to run for 100 epochs, with an early stopping mechanism activated after 10 epochs without improvement to prevent overfitting. The L1 Loss function and the Adam optimizer were employed. A weight decay parameter of 0.1 was employed to regularize the model. Ten-fold subject-independent cross-validation was performed. In each fold, the training and testing datasets did not include data from the same subject.

### Performance metrics

The model performance was assessed by two key metrics, i.e., the mean absolute error (MAE) and Pearson's correlation coefficient (PCC) between the reference and predicted BP, to evaluate the averaged estimation bias and the tracking capability, respectively. MAE and PCC were determined individually for each subject ( $MAE_{ind}$ ,  $r_{ind}$ ) as well as for one fold ( $MAE_{fold}$ ,  $r_{fold}$ ), and the mean  $\pm$  SD of these metrics were calculated across all subjects and ten folds, respectively. Additionally, the correlation ( $r$ ) between the MAEs and demographic/cardiovascular variables, and the significance of this correlation ( $P$ ) were evaluated to explore how the model's performance might be influenced by each variable. The Wilcoxon signed-rank test was utilized to compare the MAE and PCC values across models to determine the statistical significance of

their performance differences.

## RESULTS

Table 2 and Table 3 show the predictive errors and Pearson's correlation between the reference and predicted BP for the five models, respectively. As seen in Table 2, *UTransBPNet* achieved the lowest MAEs, significantly outperforming *CNN-LSTM-ATT*, *UTransBPNet\_Cal*, and *UTransBPNet\_Demo\_Last*. *UTransBPNet\_Demo\_In* achieved comparable performance with *UTransBPNet* for ABP, DBP and MBP, except for an increase of around 0.5 mmHg of MAE for SBP prediction ( $P = 0.0073$ ). From Table 3, among all models, *UTransBPNet\_Demo\_In* demonstrated the highest correlation between the reference and predicted BP at both individual and population levels, highlighting the importance of integrating demographic information as input data. All models presented notable individual differences in performance, as reflected by the SDs of *MAE\_ind* and *r\_ind*. Figure 3 provides a visualization of the estimation results of the four *UTransBPNet* models for two representative individuals. For one subject [Figure 3A], *UTransBPNet\_Demo\_In* demonstrated the most accurate tracking of the reference ABP, while all models failed to track the reference accurately in the other subject [Figure 3B].

Table 4 presents the correlation and significant analysis results between the intra-subject MAEs of ABP predictions and relevant influencing factors including demographic and cardiovascular characteristics. There is no significant correlation between the demographic features and the MAEs for all four models, suggesting that the performance of the four models does not vary with demographic factors. The performance across different age groups for the four models is further presented in Figure 4. Among all models, *UTransBPNet\_Demo\_In* presented the most robust performance among different age groups, exhibiting no significant differences between groups. In contrast, the other models presented one or more significant differences between groups. The scatter plots and the fitted linear relationship between the demographics and the MAEs given by *UTransBPNet\_Demo\_In*, as presented in Figure 5, further highlight *UTransBPNet\_Demo\_In*'s superior ability to maintain consistent accuracy irrespective of age, height, and weight.

On the other hand, all models except *UTransBPNet\_Demo\_In* presented significant correlations between the MAEs and the cardiovascular features, indicating that the model performance was influenced by cardiovascular features (defined as significant when  $r > 0.15$  and  $P < 0.05$ ). Specifically, the *UTransBPNet* model shows a notable positive correlation between the MAE and two cardiovascular features: mean and SD of SBP (denoted as *SBP\_mean* and *SBP\_std*). This indicates relatively poorer performance in individuals with high SBP levels and fluctuations. In contrast, the *UTransBPNet\_Demo\_Last* model demonstrates poorer performance primarily in individuals with faster heart rates. The *UTransBPNet\_Cal* model tends to be influenced by most factors, performing less effectively in individuals with significant SBP and DBP fluctuations, high maximum SBP and DBP, broad SBP and DBP ranges, as well as high heart rate variations. The scatter plots and fitted linear relationship graphs in Figure 6 further highlight these trends. Out of all models, the *UTransBPNet\_Demo\_In* model stands out with the most stable performance, exhibiting robust predictive performance unaffected by individual cardiovascular features.

## DISCUSSION

While numerous studies have developed new models for cuffless BP estimation, the validation was insufficient in terms of the demographic diversity of subjects and the intra-subject variations. This study investigated the effectiveness of four *UTransBPNet*-based methods and a *CNN-LSTM-ATT* model for ABP estimation in a large ICU population characterized by diverse demographics and cardiovascular conditions. Among these methods, the *UTransBPNet\_Demo\_In* model, which integrates demographic information with

**Table 2. Predictive errors of the five models**

Errors, mmHg	CNN-LSTM-ATT		UTransBPNet		UTransBPNet_Demo_In		UTransBPNet_Demo_Last		UTransBPNet_Cal	
	MAE-fold	MAE-ind	MAE-fold	MAE-ind	MAE-fold	MAE-ind	MAE-fold	MAE-ind	MAE-fold	MAE-ind
ABP	10.85 ± 0.28	11.02 ± 4.11**	<b>10.24 ± 0.23</b>	<b>10.40 ± 4.09</b>	10.38 ± 0.22	10.55 ± 4.74	10.41 ± 0.31	10.56 ± 4.26*	12.10 ± 0.58	11.74 ± 6.10**
SBP	13.56 ± 0.39	13.67 ± 6.47**	<b>12.46 ± 0.45</b>	<b>12.50 ± 5.80</b>	12.91 ± 0.39	13.04 ± 6.92*	12.84 ± 0.60	12.94 ± 6.13**	13.11 ± 0.55	12.55 ± 5.84
DBP	8.50 ± 0.35	8.56 ± 4.27**	<b>8.19 ± 0.19</b>	<b>8.32 ± 4.08</b>	8.21 ± 0.32	8.30 ± 4.53	8.33 ± 0.25	8.44 ± 4.03**	10.94 ± 0.70	8.47 ± 4.25**
MBP	9.25 ± 0.38	9.32 ± 4.42**	<b>8.84 ± 0.23</b>	<b>8.93 ± 4.22</b>	9.04 ± 0.28	9.15 ± 4.90	9.10 ± 0.33	9.19 ± 4.37**	10.89 ± 0.58	9.06 ± 4.33**

MAE of *UTransBPNet* was compared with that of the other models at the individual level. \* indicates significance ( $P < 0.01$ ); \*\* indicates high significance ( $P < 0.001$ ).

**Table 3. Pearson's correlation of the reference and predicted BP for the five models**

Pearson Correlation	CNN-LSTM-ATT		UTransBPNet		UTransBPNet_Demo_In		UTransBPNet_Demo_Last		UTransBPNet_Cal	
	r_fold	r_ind	r_fold	r_ind	r_fold	r_ind	r_fold	r_ind	r_fold	r_ind
ABP	0.85 ± 0.01	0.85 ± 0.08**	0.87 ± 0.00	0.87 ± 0.08**	<b>0.89 ± 0.01</b>	<b>0.89 ± 0.08</b>	0.87 ± 0.01	0.87 ± 0.08**	0.88 ± 0.01	0.87 ± 0.08**
SBP	0.58 ± 0.03	0.58 ± 0.23**	0.61 ± 0.01	0.61 ± 0.21**	<b>0.67 ± 0.04</b>	<b>0.67 ± 0.22</b>	0.61 ± 0.02	0.61 ± 0.20**	0.60 ± 0.01	0.61 ± 0.21**
DBP	0.56 ± 0.03	0.55 ± 0.23**	0.56 ± 0.02	0.55 ± 0.22**	<b>0.65 ± 0.03</b>	<b>0.64 ± 0.22</b>	0.56 ± 0.02	0.54 ± 0.22**	0.55 ± 0.02	0.55 ± 0.22**
MBP	0.60 ± 0.03	0.59 ± 0.22**	0.61 ± 0.02	0.61 ± 0.21**	<b>0.68 ± 0.03</b>	<b>0.68 ± 0.21</b>	0.61 ± 0.02	0.60 ± 0.21**	0.60 ± 0.02	0.61 ± 0.21**

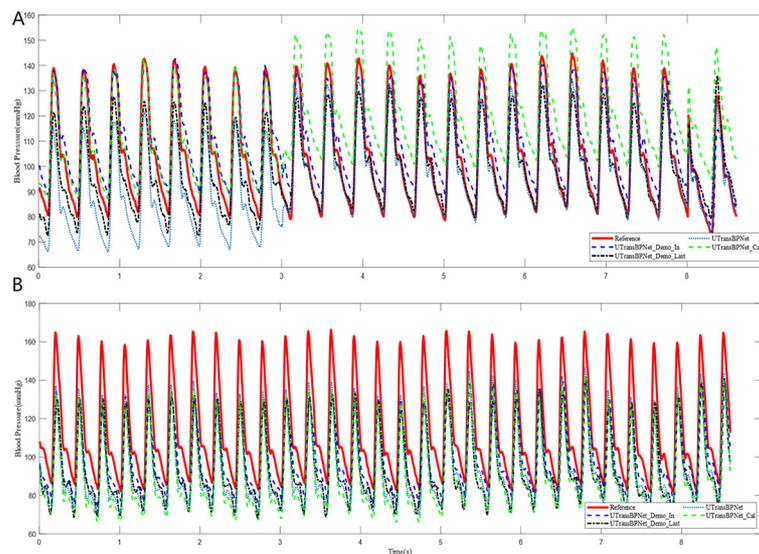
PCC of *UTransBPNet\_Demo\_In* was compared with that of the other models at the individual level. \* Indicates significance ( $P < 0.01$ ); \*\* indicates high significance ( $P < 0.001$ ).

ECG and PPG as inputs, demonstrated superior performance in tracking BP variations. Although *UTransBPNet\_Demo\_In* presented notable individual variations in performance, it demonstrates robust accuracy across different demographics and cardiovascular conditions. The superior performance of *UTransBPNet\_Demo\_In* may be due to the capability of learning effective features representing the complex interplay between demographics and cardiovascular signals through the deep learning process, which ensures robust BP estimation performance across diverse ICU demographics and cardiovascular states. In contrast, the *UTransBPNet\_Demo\_Last* method, which directly merges demographic data with deep representations of physiological signals without an extensive deep learning process, fails to effectively learn relevant features. Additionally, the *UTransBPNet\_Cal* method, which utilized a bias calibration technique, exhibited reduced performance. This decline is likely due to the unrealistic assumption that vascular states remain constant over long periods.

For monitoring patients in the ICU, continuous BP waveform is invaluable as it offers detailed insights into the patient's hemodynamic status. Therefore, we have reported the accuracy of ABP to cater to this critical requirement. However, we also recognize that other clinical settings may prioritize discrete

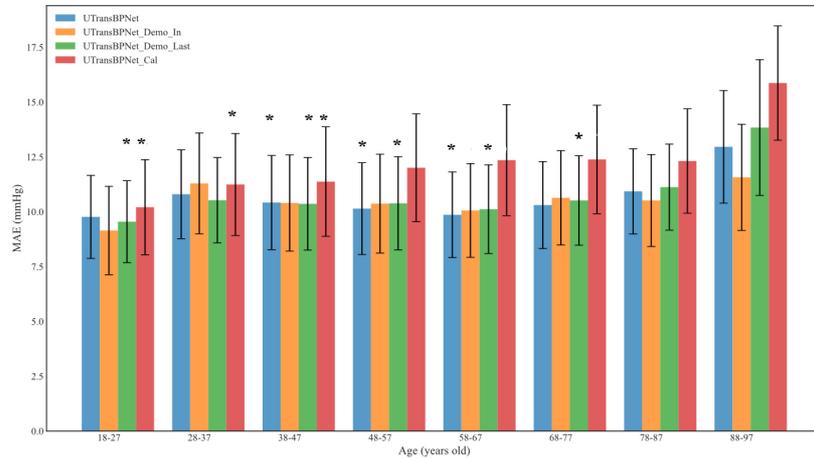
**Table 4. Correlation and significance analysis between the MAEs of ABP predictions by the four models and demographic/cardiovascular features**

Features	UTransBPNet		UTransBPNet_Demo_In		UTransBPNet_Demo_Last		UTransBPNet_Cal	
	r	P	r	P	r	P	r	P
Age	0.028	0.190	0.018	0.408	0.046	0.033	0.052	0.016
Height	-0.018	0.402	-0.021	0.339	-0.022	0.312	-0.042	0.049
Weight	-0.052	0.016	-0.050	0.019	-0.092	0.000	-0.035	0.101
BMI	-0.053	0.015	-0.048	0.026	-0.098	0.000	-0.013	0.553
Gender	0.006	0.779	-0.000	0.988	0.020	0.360	0.004	0.868
SBP_mean	<b>0.162</b>	<b>0.000</b>	0.090	0.000	0.084	0.000	0.066	0.002
SBP_std	<b>0.152</b>	<b>0.000</b>	0.089	0.000	0.100	0.000	<b>0.318</b>	<b>0.000</b>
DBP_mean	0.106	0.000	0.064	0.003	0.029	0.180	0.034	0.116
DBP_std	0.125	0.000	0.080	0.000	0.083	0.000	<b>0.308</b>	<b>0.000</b>
SBP_max	0.134	0.000	0.086	0.000	0.069	0.001	<b>0.278</b>	<b>0.000</b>
SBP_min	0.062	0.004	0.045	0.038	0.010	0.655	-0.146	0.000
DBP_max	0.074	0.001	0.061	0.005	0.026	0.224	<b>0.191</b>	<b>0.000</b>
DBP_min	0.060	0.006	0.043	0.047	0.000	0.986	-0.130	0.000
$\Delta$ SBP <sub>max-min</sub>	0.094	0.000	0.058	0.007	0.060	0.005	<b>0.336</b>	<b>0.000</b>
$\Delta$ DBP <sub>max-min</sub>	0.049	0.022	0.044	0.043	0.026	0.221	<b>0.248</b>	<b>0.000</b>
Heart_Rate_mean	0.138	0.000	0.119	0.000	<b>0.163</b>	<b>0.000</b>	0.114	0.000
Heart_Rate_std	0.048	0.025	0.006	0.792	0.034	0.111	<b>0.156</b>	<b>0.000</b>

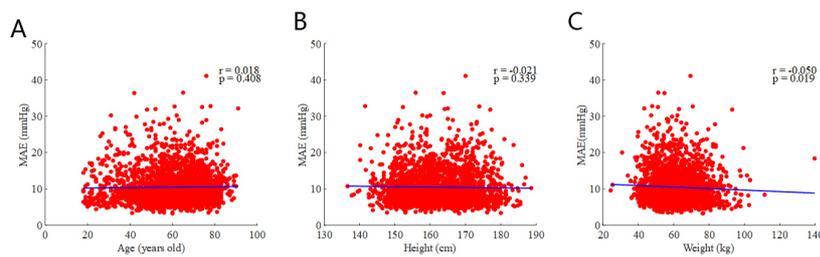


**Figure 3.** Typical examples of ABP predictions given by the four models.

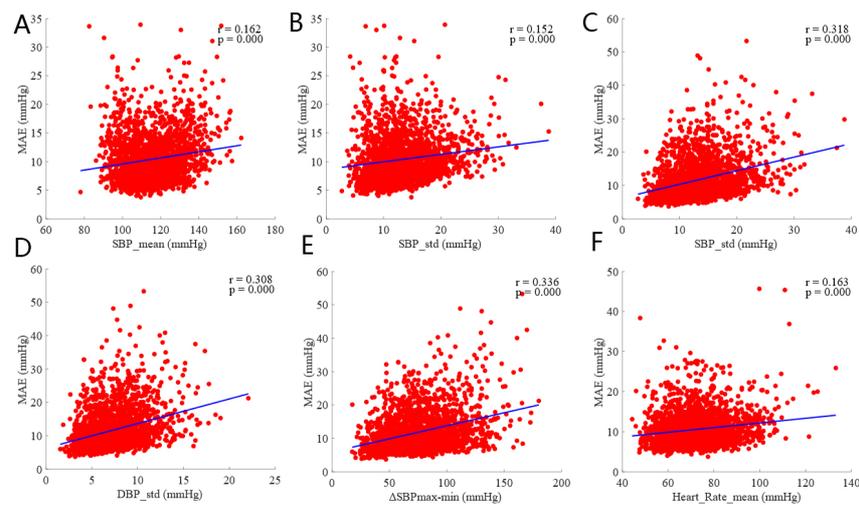
measurements of SBP, DBP, and MBP over continuous waveforms, and many previous studies have only reported the prediction performance of models for SBP and DBP<sup>[8,16]</sup>. To facilitate a broader comparison with existing research, we have included performance metrics for both SBP and DBP predictions in this study.



**Figure 4.** Analysis of the model performance across various age groups, with error bars representing the SDs of MAEs for each group. Significant testing was performed among age groups for each of the four models. \*: the predicted Mean Absolute Errors (MAEs) for the last age group are significantly different from those of the other groups across each model, achieving statistical significance with a  $P$ -value of less than 0.01 ( $P < 0.01$ ).



**Figure 5.** Scatter plots and insignificant linear relations between the demographics and the MAEs of *UTransBPNet\_Demo\_In*.



**Figure 6.** Scatter plots and linear relations between the intra-subject cardiovascular features and MAEs of the three models.

This study has several limitations. Firstly, there is a noticeable gap between the performance of the models and the international standards for BP measurement. This discrepancy may be attributed to the complexity of the PulseDB dataset, characterized by highly dynamic BP variations in ICU settings. Additionally, the data preprocessing methods employed, such as normalization within segments, might neglect useful

temporal information in physiological signals. Another significant limitation is the relatively intuitive approach used to integrate demographic features into the deep learning model. Future research should focus on developing more effective and dedicated methods for integrating demographic data and deep features extracted from continuous physiological signals. Moreover, a statistical analysis of the data distribution for the entire dataset after excluding certain segments found that, the distribution did not align with the data distribution requirements set by the AAMI standard, notably with less than 5% of the measurements falling into the very high categories of BP (> 160 and 100 mmHg for SBP and DBP, respectively). Therefore, further validation that strictly adheres to the AAMI test standard is needed in future work.

## CONCLUSION

In conclusion, this study comprehensively validated two deep learning models for cuffless BP estimation on a large-scale ICU dataset, and designed different approaches to enhance the generalization capability across diverse demographic and cardiovascular conditions. Among these, the *UTransBPNet\_Demo\_In* model, leveraging the integration of demographic data with ECG and PPG signals as model inputs, has emerged as the most effective one. Although there is still considerable room for improvement due to the challenging nature of cuffless BP estimation, especially within the complex context of ICU settings, this work provides a benchmark for future studies. Future advancement may focus on enhancing model accuracy by more effectively leveraging the temporal trends of physiological signals.

## DECLARATIONS

### Authors' contributions

Writing and analyzing experimental results: Huang Y, Zheng Y

Model implementation and experimental design: He Y, Song Z

Conceptualization of research ideas: Zheng Y, Gao K

### Availability of data and materials

The data used in this study are from the public dataset.

### Financial support and sponsorship

This work was supported by Guangdong Basic and Applied Basic Research Foundation [2021A1515110025]; Young Scientists Fund from National Natural Science Foundation of China (NSFC) [62301333]; Research Foundation of Education Department of Guangdong Province [2022ZDJS115]; and the Common University Innovation Team Project of Guangdong [2021KCXTD041].

### Conflicts of interest

Not applicable.

### Ethical approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Copyright

© The Author(s) 2024.

## REFERENCES

1. Mukkamala R, Stergiou GS, Avolio AP. Cuffless blood pressure measurement. *Annu Rev Biomed Eng* 2022;24:203-30. DOI PubMed
2. Schutte AE, Kollias A, Stergiou GS. Blood pressure and its variability: classic and novel measurement techniques. *Nat Rev Cardiol* 2022;19:643-54. DOI PubMed PMC
3. Avolio A. The reality and serendipity of cuffless blood pressure monitoring. *Hypertens Res* 2023;46:1609-11. DOI PubMed
4. Tanveer MS, Hasan MK. Cuffless blood pressure estimation from electrocardiogram and photoplethysmogram using waveform based ANN-LSTM network. *BIOMED SIGNAL PROCES* 2019;51:382-92. DOI
5. Sadrawi M, Lin YT, Lin CH, et al. Genetic deep convolutional autoencoder applied for generative continuous arterial blood pressure via photoplethysmography. *Sensors* 2020;20:3829. DOI PubMed PMC
6. Landry C, Hedge ET, Hughson RL, Peterson SD, Arami A. Accurate blood pressure estimation during activities of daily living: a wearable cuffless solution. *IEEE J Biomed Health Inform* 2021;25:2510-20. DOI PubMed
7. Vardhan KR, Vedanth S, Poojah G, Abhishek K, Kumar MN, Vijayaraghavan V. BP-Net: efficient deep learning for continuous arterial blood pressure estimation using photoplethysmogram. 20th IEEE International Conference on Machine Learning and Applications (ICMLA); 2021. p. 1495- 500.
8. Ma C, Zhang P, Song F, et al. KD-informer: a cuff-less continuous blood pressure waveform estimation approach based on single photoplethysmography. *IEEE J Biomed Health Inform* 2023;27:2219-30. DOI
9. Mahmud S, Ibtehad N, Khandakar A, et al. NABNet: A nested attention-guided BiConvLSTM network for a robust prediction of blood pressure components from reconstructed arterial blood pressure waveforms using PPG and ECG signals. *Biomed Signal Proces* 2023;79:104247. DOI
10. Hsu YC, Li YH, Chang CC, Harfiya LN. Generalized deep neural network model for cuffless blood pressure estimation with photoplethysmogram signal only. *Sensors* 2020;20:5668. DOI PubMed PMC
11. Senturk U, Polat K, Yuicedag I. A non-invasive continuous cuffless blood pressure estimation using dynamic recurrent neural networks. *Appl Acoust* 2020;170:107534. DOI
12. Lee D, Kwon H, Son D, et al. Beat-to-beat continuous blood pressure estimation using bidirectional long short-term memory network. *Sensors* 2020;21:96. DOI PubMed PMC
13. Song K, Chung K, Chang J. Cuffless deep learning-based blood pressure estimation for smart wristwatches. *IEEE Trans Instrum Meas* 2020;69:4292-302. DOI
14. Zhang L, Hurley NC, Ibrahim B, et al. Developing personalized models of blood pressure estimation from wearable sensors data using minimally-trained domain adversarial neural networks. Proceedings of the 5th Machine Learning for Healthcare Conference; 2020. p. 97-120.
15. Lee J, Yang S, Lee S, Kim HC. Analysis of pulse arrival time as an indicator of blood pressure in a large surgical biosignal database: recommendations for developing ubiquitous blood pressure monitoring methods. *J Clin Med* 2019;8:1773. DOI PubMed PMC
16. Liu ZD, Li Y, Zhang YT, et al. Cuffless blood pressure measurement using smartwatches: a large-scale validation study. *IEEE J Biomed Health Inform* 2023;27:4216-27. DOI
17. Zheng Y, Liu Q, Hong J, Wu S, Zhang Y. UTransBPNet: a general deep learning model for cuffless blood pressure estimation under activities. Available from: <https://www.techrxiv.org/users/688692/articles/680841-utransbpnet-a-general-deep-learning-model-for-cuffless-blood-pressure-estimation-under-activities> [Last accessed on 13 Mar 2024].
18. Hong J, Gao J, Liu Q, Zhang Y, Zheng Y. Deep learning model with individualized fine-tuning for dynamic and beat-to-beat blood pressure estimation. 2021 IEEE 17th International Conference on Wearable and Implantable Body Sensor Networks (BSN), Athens, Greece; 2021. pp. 1-4.
19. Wang W, Mohseni P, Kilgore KL, Najafizadeh L. PulseDB: A large, cleaned dataset based on MIMIC-III and VitalDB for benchmarking cuff-less blood pressure estimation methods. *Front Digit Health* 2022;4:1090854. DOI PubMed PMC