



OPEN *UTransBPNet* for cuffless and calibration-free blood pressure estimation under dynamic conditions

Yali Zheng¹✉, Hongda Huang¹, Jiasheng Gao¹, Jingyuan Hong², Shenghao Wu¹,
Yuanting Zhang³ & Qing Liu⁴✉

Accurate cuffless blood pressure (BP) estimation remains challenging, particularly under dynamic conditions with significant intra-individual BP variations. This study introduces *UTransBPNet*, a novel, calibration-free model for cuffless BP estimation. It integrates a squeeze-and-excitation-enhanced Unet architecture for short-range feature extraction with a transformer and cross attention module to capture long-range dependencies from high-resolution, multi-channel physiological signals, further refined through an optimized fine-tuning scheme. Comprehensive validations were conducted across multiple dynamic datasets—Dataset_Drink, Dataset_Exercise, and Dataset_MIMIC—in both scenario-specific and cross-scenario settings. Results demonstrate that *UTransBPNet* outperformed existing models in tracking BP variations under dynamic conditions, achieving individual Pearson's correlation coefficients of 0.61 ± 0.17 and 0.62 ± 0.13 for systolic BP (SBP) and diastolic BP (DBP) in Dataset_Drink, 0.82 ± 0.11 and 0.72 ± 0.18 in Dataset_Exercise, and low mean absolute differences (MADs) of 4.38 and 2.25 mmHg in Dataset_MIMIC. The analysis also highlights the impact of dataset characteristics on model performance, such as distribution shift, distribution imbalance and individual BP variability, highlighting the need for well-curated data to ensure generalizability. This study advances the development of robust, cuffless BP estimation models for real-world applications.

Keywords Cuffless blood pressure estimation, Cross-scenario, Model generalizability, Calibration-free, Distribution shift, Distribution imbalance

It is estimated that there are around 1.28 billion adults aged 30–79 years worldwide with hypertension, and 46% were unaware of their conditions¹. Hypertension is also an important risk factor of some serious cardiovascular diseases such as heart attack and stroke. Although it is well known that blood pressure (BP) is intrinsically dynamic and the association of BP variability with cardiovascular outcomes has been well recognized², snapshot BP measurements are commonly adopted in current clinical practice of hypertension due to the limitation of existing cuff-based BP monitors, resulting in poor rates of control. Given the increasing burden of hypertension, there is a great demand to have a new way to diagnose and treat hypertension, while the first step to move is to manifest the “real” BP to the patients and clinicians.

Cuffless BP measurement technology has been attracting considerable attention due to its capability in providing continuous BP readings in a long period unobtrusively, which is very promising to improve the management of hypertension. Driven by the mature of cardiovascular sensing technologies³, a number of cuffless BP monitors are emerging in the marketplace in recent years, and some have been validated in clinical trials^{4–6}. The main challenge in cuffless BP monitoring technology lies in establishing the complex relationship between measurable cardiovascular signals or features and BP. Over the past decades, numerous models have been developed, which can be broadly categorized into physiological models and data-driven models. Physiological models, such as pulse wave velocity-based models⁷, have solid theoretical foundations but typically require cuff-based measurements for individual calibration. Despite their clear physiological basis and simplicity,

¹ Department of Biomedical Engineering, College of Health Science and Environmental Engineering, Shenzhen Technology University, Shenzhen, China. ² School of Imaging Sciences and Biomedical Engineering, King's College London, London, UK. ³ Chinese University of Hong Kong, Hong Kong, China. ⁴ Department of Communication and Networking, Xi'an Jiaotong-Liverpool University, Suzhou, China. ✉email: zhengyali@sztu.edu.cn; Qing.Liu@xjtu.edu.cn

these models struggle to accurately capture changes in the contraction status of the arterial wall^{8,9}. In contrast, data-driven models, such as those based on pulse wave analysis, lack strong theoretical foundations. These models generally extract morphological features from pulse waveforms and map them to BP using machine learning algorithms^{10–13}. Recent advances in deep learning have led to a surge in fully data-driven models, which can automatically learn temporal and cross-channel representations from physiological signals such as electrocardiogram (ECG), photoplethysmogram (PPG), ballistocardiogram (BCG), or their combination, without the need for handcrafted feature design¹⁴. Deep learning models typically require large amounts of data to train model parameters, with a practical way being to build *calibration-free* models based on population data. Compared to individualized models that require individual calibration, calibration-free models usually show degraded accuracy due to the challenge in learning the complex relationship from population data.

Various deep learning architectures have been adopted for cuffless BP estimation, including deep neural networks (DNN)¹⁵ and one-dimensional (1D) convolutional neural networks (CNN)¹⁶. However, CNN’s temporal representation capabilities are limited by its restricted receptive field, making it less effective at capturing long-range temporal dependencies from high-resolution (> 100 Hz) physiological signals. Efforts to enhance CNN-based models for better temporal feature representation include transforming physiological signals into phase space and using two-dimensional recurrence features through fuzzy recurrence plots¹⁷. Zhang et al. further improved CNN’s feature representation by incorporating a squeeze-and-excitation network to learn channel attention¹⁸. In addition, long short-term memory neural network (LSTM)^{12,19} and gated recurrent unit (GRU)^{14,20} have also been employed for BP estimation to enhance temporal feature representation. Combining the strengths of CNN with LSTM or GRU has led to CNN-LSTM or CNN-GRU models, which are now popular choices for BP estimation in recent studies^{14,21–23}. Recently, Unet and its variants have also gained attention for BP estimation due to their ability to capture contextual information by concatenating upsampled features from the expansive path with convolutional features from the contractive path, providing more precise outputs. Zhang et al. combined Unet with squeeze-and-excitation layers and LSTM for BP waveform reconstruction²⁴. Additionally, Ma et al. proposed KD-informer²⁵, a transformer-based model that estimates BP using single-channel PPG and has achieved state-of-the-art accuracy on both a private dataset and the MIMIC dataset.

Despite these advancements, a key issue remains: few studies have comprehensively validated models under dynamic conditions that involve sufficient intra-individual BP variations induced by activities or interventions, as required in the IEEE Standard 1708 and recommended in²⁶. Two recent studies tested under dynamic scenarios such as coffee drinking¹⁵ and water drinking¹⁴, reported significantly degraded performance compared to static conditions. Our previous work¹⁴ showed that existing models such as CNN, CNN-LSTM, were struggled to correctly estimate BP under dynamic situations. Considering BP presents both short- and long-range variation patterns under dynamic conditions due to various BP regulation mechanisms, existing deep models were not specifically designed to effectively learn these short- and long-term features from high-resolution physiological signals.

Inspired by the Unet-Transformer structure²⁷, which excels at learning fine-grained, local features through Unet and capturing global and long-range dependencies via transformer, we propose *UTransBPNet*, a novel BP estimation model for cuffless BP estimation. It is designed to effectively learn discriminative features from multi-channel and high-resolution physiological signals. The main contributions of this study are as follows:

- 1) A novel calibration-free model *UTransBPNet* was proposed, specifically designed to effectively learn short- and long-range features from multi-channel, high-resolution physiological signals;
- 2) An optimized fine-tuning scheme that leverages final-layer features of Unet and updates all parameters was found to yield the best results for estimating systolic and diastolic BP from BP waveforms;
- 3) *UTransBPNet* was comprehensively validated on multiple dynamic datasets, in both scenario-specific and cross-scenario settings. The findings offer key insights into the impact of dataset characteristics on model performance.

Methodology
Datasets

The basic information and distributions of systolic BP (SBP) and diastolic BP (DBP) of the three datasets are shown in Table 1; Fig. 1. Dataset_Drink originates from a previous study²⁸. 25 healthy subjects (aged 27 ± 3 years) were recruited; each participant was asked to rest for 5 min, drink 400 mL of water within 5 min, and then recover for 50 min. During the procedure, lead I ECG and PPG from left index finger of each subject were acquired continuously by an in-house multi-channel physiological acquisition system. Continuous arterial BP waveforms were measured by Finometer (Finapres Medical System BV, Netherlands). All data were sampled at 1 kHz by a data acquisition system (DI220, DATAQ Instruments WinDaq, USA).

Dataset	Subject	Number of subjects or recordings	Number of segments	BP Reference	Mean ± SD of intra-subject BP changes, mmHg		Mean ± SD of BP, mmHg	
					SBP	DBP	SBP	DBP
Drinking	Healthy Young	25 subjects	62,678	Noninvasive BP, Finapres	47.77 ± 13.64	36.65 ± 9.87	131.14 ± 14.96	79.37 ± 9.76
Exercise	Healthy Young	20 subjects	28,814	Noninvasive BP, Finapres	67.24 ± 24.68	46.70 ± 11.42	129.09 ± 18.00	72.37 ± 12.91
MIMIC	ICU patients	163 recordings	15,474	Invasive BP	13.54 ± 6.44	7.04 ± 3.09	122 ± 18.56	63.28 ± 8.27

Table 1. Basic information of the three datasets.

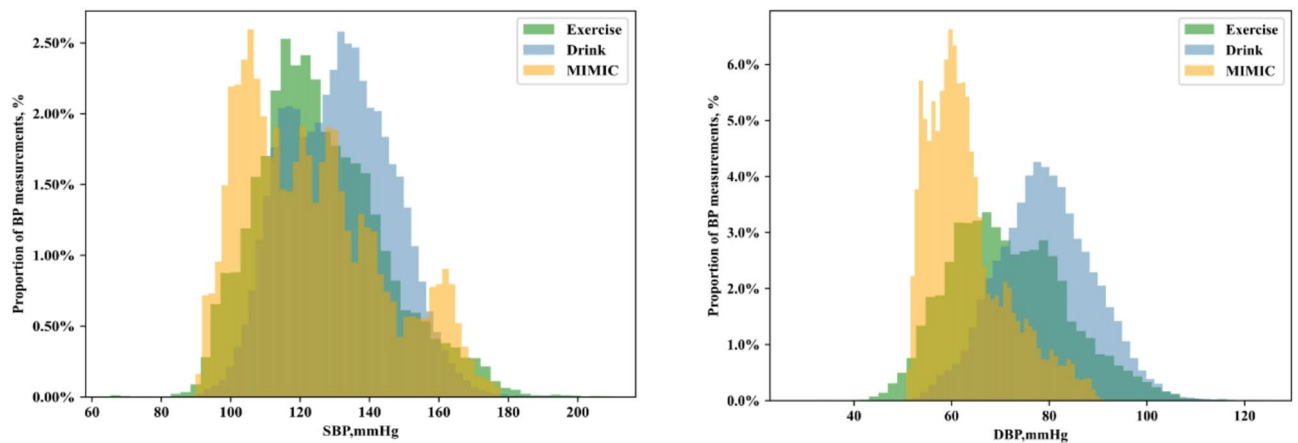


Fig. 1. The histograms of SBP and DBP of the three datasets.

Dataset_Exercise is from another previous study²⁹, involving 20 healthy subjects aged 26 ± 4 years. Each subject was asked to lie on a tilted bed and perform lower limb exercise by cycling, with the workload increased by 25 watts every 2 min from an initial load of 25 watts until the target heart rate of $85\% \times (220 - \text{Age})$ or exhaustion was reached. This setup ensures minimal interferences to fingertip PPG and ECG signals. The experimental setup and devices for ECG, PPG, and continuous BP signal acquisition were identical to those in Dataset_Drink.

Dataset_MIMIC is an online public dataset, a subset of the Multiparameter Intelligent Monitoring in Intensive Care (MIMIC) II waveform database. ECG, fingertip PPG and invasive arterial BP waveforms were recorded from patients across various hospitals, with a sampling rate of 125 Hz for each signal. Data were initially filtered by simple average filtering. Segments with abnormal BP values outside 20–200 mmHg, or irregular heart rates were removed³⁰. After excluding recordings shorter than 8 min, 1,925 recordings remained for further analysis.

Informed consent was obtained from all participants involved in the above studies, and the reuse of these datasets in this study was approved by the Ethics Committee of Shenzhen Technology University. All methods were performed in accordance with the relevant guidelines and regulations.

Data preprocessing

Signals from Dataset_Drink and Dataset_Exercise were down sampled to 125 Hz. ECG, PPG, and BP signals of the three datasets were all bandpass filtered between 0.5 and 30 Hz, 0.5–15 Hz, and 0.5–15 Hz, respectively, using Butterworth filter to remove baseline drift and high-frequency noises.

For Dataset_Drink and Dataset_Exercise, segments during Finapres calibration intervals and those contaminated by motion artifacts were removed. For Dataset_MIMIC, segments with saturated ECG or BP waveforms or motion artifacts were excluded. Due to the substantial workload involved in examining all 5-sec segments of these instances, we manually examined segments from over 200 instances, removing those contaminated by motion artifacts or saturated ECG or BP waveforms. Eventually, 163 recordings were selected for the final analysis.

The clean signals of each dataset were normalized by min-max normalization to rescale the data in the range (0, 1). The data were then partitioned into segments of 5-second length, with a 200-sample overlap. The final segment counts were 62,678, 28,814 and 15,474 segments for Dataset_Drink, Dataset_Exercise and Dataset_MIMIC, respectively. The intra-subject BP changes (Max-Min) of each dataset were also calculated and listed in Table 1.

Deep learning models

Model Architecture. The proposed *UTransBPNet* deep model for cuffless BP estimation is shown in Fig. 2. The model combines an improved Unet encoder-decoder structure, transformer layer and cross-attention mechanisms. The novelty of *UTransBPNet* lies in its hybrid architecture, which enhances Unet with both Squeeze-and-Excitation (SE) modules for channel-wise attention and a transformer encoder for capturing long-range dependencies. Additionally, cross-attention modules between adjacent Unet layers enable hierarchical feature refinement, improving the balance between local and global feature representation. This design collectively enhances the model's ability to extract discriminative physiological features, leading to more accurate BP estimation from input signals. The details of the model structure are described as follows:

Improved Unet encoder-decoder

Unet was adopted to learn short-range temporal features from input signals. The Unet encoder involves three down-sampling steps, which is comprised of four Conv_SE blocks and two mean-pooling layers. Each Conv_SE contains three Conv_block and SE block. The detailed structure of Conv_Block and SE_Block are shown in Fig. 2. The decoder mirrors the encoder with three up-sampling steps, ensuring the output BP waveform maintains the same temporal resolution as the input signal. To emphasize relevant features and suppress noises, Unet was improved by incorporating SE into its convolution layers. Specifically, the squeeze operation (global average

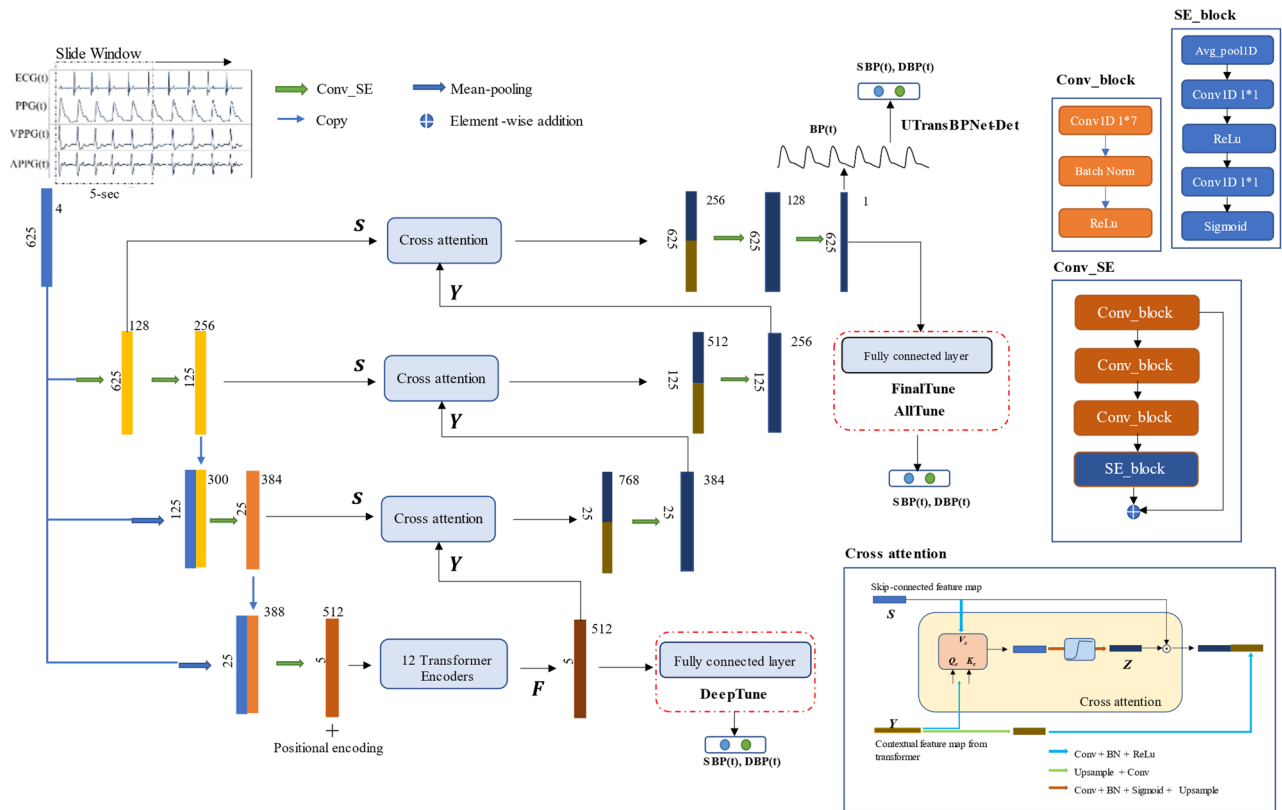


Fig. 2. The model structure of the proposed *UTransBPNet* for cuffless BP estimation.

pooling) condenses the spatial dimensions into a single descriptor per channel, capturing global contextual information. The excitation operation then dynamically learns channel-wise importance weights through convolutional layers with nonlinear activation functions (ReLU and sigmoid). These learned attention weights are subsequently applied to the original feature map via channel-wise multiplication, selectively enhancing relevant channels while suppressing less informative ones.

Transformer Module

Since conventional Unet struggles with long-range feature dependencies, we further introduced a transformer encoder at the end of the Unet encoder to enhance its ability to capture global contextual information. The 12-layer transformer encoder was positioned at the bottom of the Unet stack, to capture long-range relationships of the features from the Unet encoder by multi-head self-attention (MHSA)³¹. Additionally, positional encoding was added to the input of transformer to capture contextual dependencies within the data. The MHSA can be formulated as,

$$\text{Self-Attention}(Q_s, K_s, V_s) = \text{softmax}\left(\frac{Q_s K_s^T}{\sqrt{d_k}}\right) V_s \quad (1)$$

where Q_s , K_s and V_s refer to queries, keys and values of inputs, respectively. $\sqrt{d_k}$ denotes the dimensionality of the query/key-value sequence. Specifically, the deepest-level feature map F (including the deepest-level feature map of Unet and its learnt positional encoding) is embedded using learned embedding matrices, resulting in embedded queries Q_s , keys K_s and values V_s . Then, a dot-product operation is performed between Q_s and the transposed K_s , followed by softmax normalization to generate the contextual attention map, which reflects the similarity between each element in Q_s and the global elements of K_s . The contextual attention map is multiplied by V_s , producing a weighted average representation.

Multi-head cross attention (MHCA) mechanism

Additionally, we incorporated MHCA between adjacent Unet layers to facilitate the guidance of low-level features by high-level features, ensuring a more effective hierarchical feature refinement. These modifications are

particularly beneficial for representing input physiological signals which presents both long-range dependencies and fine-grained local features, ultimately leading to a more representative mapping relationship with BP. Specifically, the feature map Y from transformer encoder output is embedded into queries Q_c and keys K_c . The low-level skip connected feature map S is embedded into values V_c . As shown in Fig. 2, the attention weights learnt from Y are transformed into Z through a sigmoid activation function, which acts as a filter for S . By applying this filtered attention to S via a dot-product operation, irrelevant features in S are suppressed. The filtered feature map is then concatenated by the up-sampled feature map of Y .

Model training and validation

SBP and DBP Estimation. The model inputs include ECG(t), PPG(t), the first and second differential derivatives of PPG(t), i.e., VPPG(t) and APPG(t). Thus, the input shape of the model is (625, 4). The output of *UTransBPNet* is the normalized BP waveform BP(t). Then, a detection algorithm was applied to extract maximal and minimal points of the last beat of each 5-sec segment of BP(t) to obtain SBP and DBP. The detected SBP and DBP values were then de-normalized by min-max method using the maximal and minimal SBP and DBP of the training dataset. This approach was referred as *UTransBPNet-Attn-Det*.

Alternatively, a fully connected layer was added to the output of *UTransBPNet* to estimate SBP and DBP. Three different fine-tuning schemes were devised and the training process of *UTransBPNet* was implemented in two phases as described in Supplementary Table. S1. Specifically, *UTransBPNet* was first trained in *Phase I* and then the fully connected layer was finetuned in *Phase II* for SBP and DBP regression in three different ways: (1) *UTransBPNet-DeepTune* and *UTransBPNet-Crossattn-DeepTune* with and w/o cross attention. The feature map Y from the transformer output was used as inputs of the fully connected layer for SBP and DBP estimation; (2) *UTransBPNet-FinalTune* and *UTransBPNet-Crossattn-FinalTune*. The predicted BP waveform of the *UTransBPNet* was used as inputs of the fully connected layer for SBP and DBP estimation. The parameters of *UTransBPNet* θ was frozen during fine-tuning; (3) *UTransBPNet-Crossattn-AllTune*. All model parameters including θ and φ were fine-tuned. The estimated SBP and DBP were then de-normalized in the same way as *UTransBPNet-Attn-Det*.

As the proposed model did not utilize individual BP for calibration, the population averaged SBP and DBP of each dataset were adopted to build the baseline model for fair comparison as suggested in²⁶. Additionally, several widely used model architectures were implemented for performance comparison in this study, including: CNN with bidirectional LSTM and attention mechanism (*CNN-BiLSTM-Attn*), a naïve Unet architecture with the same layer configuration as *UTransBPNet*, SEUnet¹⁸, and ResUnet³². All models were trained in an end-to-end manner using SBP and DBP as ground truth.

Model Training Setup. The model was built in Pytorch. A NVIDIA Tesla V100 PCIe graphics card with 32 GB Video RAM was used for the training and testing. The batch size used for training was 32, and the learning rate was set to 0.0009. To prevent overfitting, an early stopping mechanism was implemented, whereby training would be terminated if there were no improvements for 20 epochs. The optimizer utilized in the training process was Adam. To further control overfitting, the weight decay hyperparameter was set to 0.001. Scenario-specific and cross-scenario validation were performed as follows:

Scenario-specific validation

Leave-one-subject-out cross validation was conducted separately for Dataset_Drink and Dataset_Exercise, while instance-independent ten-fold cross validation was performed for Dataset_MIMIC.

Cross-scenario tests

To assess the generalization ability of deep learning models, several cross-scenario tests were performed: (1) *Test 1*: pre-train on Dataset_Drink and test on Dataset_Exercise; (2) *Test 2*: Train on Dataset_Exercise and test on Dataset_Drink; (3) *Test 3*: Train on the combination of Dataset_Drink Dataset_Exercise and test on Dataset_MIMIC. Furthermore, to enhance model adaptability to different activity scenarios, scenario-specific data were used to finetune the model following cross-scenario pretraining. To balance between improving model performance and minimizing data collection costs in real-world applications, around 10% of scenario-specific data were used for finetuning, and the remaining 90% of the data was used for testing. Specifically, two subjects were randomly selected from the testing scenario for finetuning in Tests 1 and 2, while 10% of instances were randomly selected from Dataset_MIMIC for fine-tuning in *Test 3*. It is worth noting that, although the model was finetuned, it is not individualized; rather, the finetuning was scenario-specific.

Evaluation Metrics. The performance metrics of two international standards, the Advancement of Medical Instrumentation (AAMI) and IEEE Standard 1708a-2019, were adopted to evaluate model performance, including: the mean and standard deviation (SD) of the differences and the mean absolute differences (MAD) between the reference and estimated BP. In addition, the Pearson's correlation coefficient (PCC) between the estimated and the reference BP was also adopted as performance metric. Individual PCC were calculated for Dataset_Drink and Dataset_Exercise to evaluate model's tracking capability of intra-individual BP changes. On the other hand, as individual information were missing for Dataset_MIMIC and only very small variations existed in each recording, PCC were calculated within each fold.

Statistical Test. The paired Student's *t*-test and Pitman-Morgan test were conducted to compare MAD and SD between models. A statistically significant result is indicated by an asterisk (*) ($P < 0.05$), denoting that the SD or MAD of *UTransBPNet-Crossattn-AllTune* was significantly lower than that of other models. We also tested if there was significant performance improvement by introducing cross attention to *UTransBPNet*, with (') indicating a significant contribution.

Results

Scenario-Specific validation

Supplementary Table.S2 demonstrates the contribution of cross attention and different finetune schemes to UTransBPNet. Adding a fully connected layer with fine-tuning generally achieves higher accuracy than with the detection scheme. In addition, adding cross-attention to the AllTune and FinalTune methods further enhances performance by reducing MAD and increasing PCC, except in the DeepTune method, indicating that cross attention contributed significantly to the model performance. *UTransBPNet-Crossattn-AllTune* performs comparably to *UTransBPNet-Crossattn-FinalTune* on Dataset_Drink and Dataset_Exercise. However, for Dataset_MIMIC, *UTransBPNet-Crossattn-AllTune* achieves notably lower MAD (4.38 vs. 6.54 mmHg for SBP and 2.25 vs. 3.22 mmHg for DBP) and significantly higher PCC. Overall, *UTransBPNet-Crossattn-AllTune* outperformed the other UTransBPNet models.

As shown in Table 2, *UTransBPNet-Crossattn-AllTune* outperformed state of models in previous works across all three datasets. We further provided the Bland-Altman plots under three scenarios, showing the agreement between the reference BP values and the estimations provided by *UTransBPNet-Crossattn-AllTune* as illustrated in Supplementary Fig.S1. The correlation plots of the reference and predicted SBP and DBP by *UTransBPNet-Crossattn-AllTune* are shown in Supplementary Fig.S2. The results suggest that, the predicted BP of Dataset_Drink and Dataset_MIMIC are in good agreement with the reference BP values, while larger errors in Dataset_Exercise.

Influence of distribution shift. Figure 3 (a) shows the estimation results for two typical subjects from Dataset_Drink, one with small estimation errors and the other with large estimation errors. In the case of *Subject_21*, the model accurately estimates BP and tracks the changes, with MAD of 3.99 and 3.41 mmHg for *UTransBPNet-Crossattn-AllTune* and *UTransBPNet-Crossattn-FinalTune*, respectively. However, for *Subject_11*, the MAD exceeds 20 mmHg. Figure 3 (b) illustrates the BP distributions of the training and testing sets of the two subjects, highlighting the influence of distribution shifts on estimation accuracy. When the BP distribution of the training set fully covers that of the testing set (as with *Subject_21*), the model provides accurate estimations. In contrast, for *Subject_11*, where the BP range of the training set (90–160 mmHg) does not cover the testing set (110–190 mmHg), the estimation performance significantly degrades.

Influence of distribution imbalance. Additionally, the MAD distributions across different BP intervals for five UTransBPNet models in Drink and Exercise scenarios, are illustrated in Fig. 4. Model performance shows a notable dependency on BP distribution: as the proportion of BP measurements decreases, model accuracy tends to decline, a trend more pronounced for SBP than DBP. Among the five models, *UTransBPNet-Crossattn-AllTune* and *UTransBPNet-Crossattn-FinalTune* exhibit the most favorable performance in BP ranges with more BP readings. However, when dealing with ranges that have insufficient number of BP readings, no model consistently performed the best.

		Mean \pm SD, mmHg		MAD, mmHg		PCC	
Model Structure	Individual/General	SBP	DBP	SBP	DBP	SBP	DBP
Dataset_Drink							
Baseline model	Population	0.00 \pm 15.09*	0.00 \pm 9.90*	12.42*	7.87	NA	NA
CNN + Bi-GRU (2021) ¹⁴	Population	2.06 \pm 16.64	0.97 \pm 10.51	13.43	8.48	0.45 \pm 0.33	0.49 \pm 0.22
CNN-BiLSTM-Attn (2021) ¹⁹	Population	-2.03 \pm 15.32*	-0.75 \pm 10.09*	12.41*	8.29*	0.53 \pm 0.19	0.51 \pm 0.20
Unet	Population	-1.80 \pm 14.50*	-0.11 \pm 9.62*	11.87*	7.83*	0.58 \pm 0.20	0.61 \pm 0.16
SEUnet (2024) ¹⁸	Population	-2.34 \pm 14.53	-1.36 \pm 9.51	11.68	7.8	0.58 \pm 0.24	0.61 \pm 0.17
ResUnet (2025) ³²	Population	-3.49 \pm 15.09*	-1.04 \pm 10.20*	12.61*	8.20*	0.56 \pm 0.21	0.60 \pm 0.16
UTransBPNet-Crossattn-AllTune (ours)	Population	-0.74\pm8.72	-0.23\pm6.31	6.59	4.99	0.59 \pm 0.19	0.60 \pm 0.17
Dataset_Exercise							
Baseline model	Population	0.00 \pm 18.37*	0.00 \pm 12.05*	14.59*	9.78*	NA	NA
CNN-BiLSTM-Attn (2021) ¹⁹	Population	-2.43 \pm 14.31*	-3.28 \pm 11.66*	10.85*	8.84*	0.74 \pm 0.14	0.55 \pm 0.24
Unet	Population	-2.43 \pm 14.78*	-2.33 \pm 11.10*	11.20*	8.50*	0.69 \pm 0.21	0.57 \pm 0.25
SEUnet (2024) ¹⁸	Population	-2.54 \pm 13.51	-2.80 \pm 10.64	10.32	8.11	0.75 \pm 0.13	0.63 \pm 0.23
ResUnet (2025) ³²	Population	-3.59 \pm 15.36*	-2.33 \pm 10.46*	12.52*	8.33*	0.71 \pm 0.18	0.62 \pm 0.20
UTransBPNet-Crossattn-AllTune (ours)	Population	-0.18\pm10.52	-0.77\pm8.11	8.51	6.22	0.82 \pm 0.11	0.72 \pm 0.18
Dataset_MIMIC							
Baseline model	Population	0.00 \pm 20.07*	0.00 \pm 8.19*	16.66*	6.39*	NA	NA
AdaBoost regression (2017) ⁷	Population	NA \pm 6.11	NA \pm 10.09	11.17	5.35	0.59	0.48
ABP-Net (2021) ⁸	Population	1.64 \pm 7.42	-0.28 \pm 5.81	6.41	3.73	NA	NA
KD-informer (Transfer learning, 2022) ⁹	Population	0.02 \pm 5.93	0.02 \pm 3.87	4.01	2.97	NA	NA
KD-informer (Tuning on MIMIC, 2022) ⁹		0.03 \pm 6.38	0.02 \pm 4.49	4.30	3.13	NA	NA
UTransBPNet-Crossattn-AllTune (ours)	Population	0.40 \pm 6.21	-0.11 \pm 3.10	4.38	2.25	0.78 \pm 0.07	0.69 \pm 0.09

Table 2. Comparing the results of UTransBPNet with other models across the three datasets. (*) indicates that the SD or MAD of *UTransBPNet-Crossattn-AllTune* was significantly lower than other tested models.

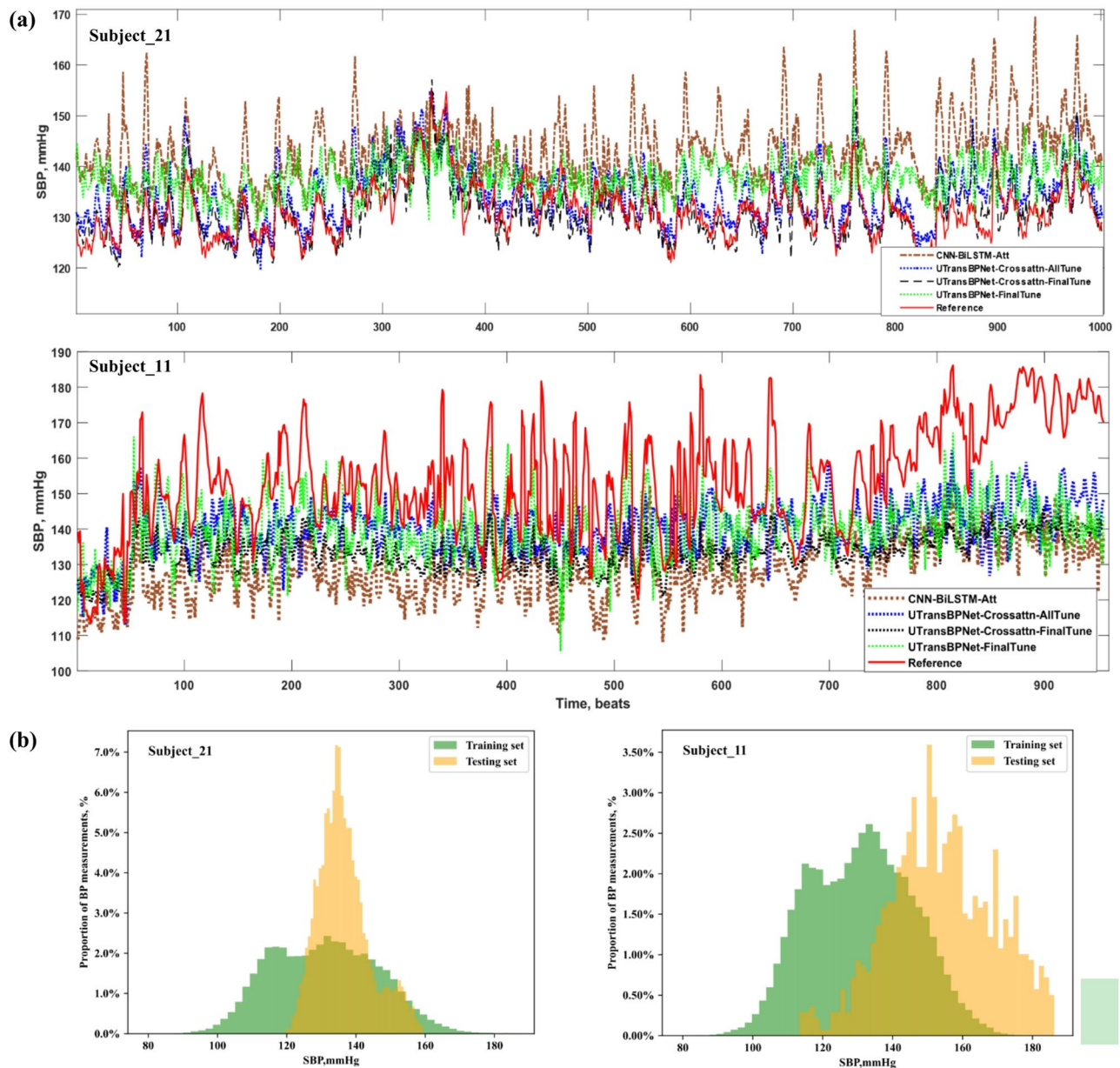


Fig. 3. (a) The SBP estimation results of two representative subjects in Dataset_Drink by four different models. The MADs are 10.43, 8.31, 3.41, 3.99 mmHg by CNN-BiLSTM-Attention, UTransBPNet-FinalTune, UTransBPNet-Crossattn-FinalTune and UTransBPNet-Crossattn-AllTune for Subject_21, and 25.08, 16.88, 20.61 and 15.36 mmHg for Subject_11, respectively. (b) The histograms of SBP of the training and testing sets for the two subjects.

Cross-Scenario tests

The statistical results of the cross-scenario tests are summarized in Table 3, with representative estimation results shown in Supplementary Fig.S2. Without finetuning, the model struggles to adapt to cross-scenario data, especially when the BP range of the training set is narrower and does not fully encompass that of the testing set (*Test 1*). Even when the BP range of the training set covers the testing set (*Tests 2 and 3*), the cross-scenario results remain unsatisfactory in two ways, as illustrated in Supplementary Fig.S2 (a): (1) large bias, reflected by high MAD values, and (2) poor tracking capability, indicated by extremely low PCC values.

After finetuning with scenario-specific data, the MADs and PCCs of all tests improve significantly. Notably, the accuracy of *Test 3* surpasses that under scenario-specific conditions, with MADs of 4.18 mmHg for SBP and 2.15 mmHg for DBP compared to 4.38 and 2.25 mmHg, respectively. Representative results are illustrated in Supplementary Fig.S2 (b). However, the accuracies of *Tests 1 and 2* remain substantially lower than those under scenario-specific conditions. These results indicate the limited generalization capability of UTransBPNet across dynamic scenarios.

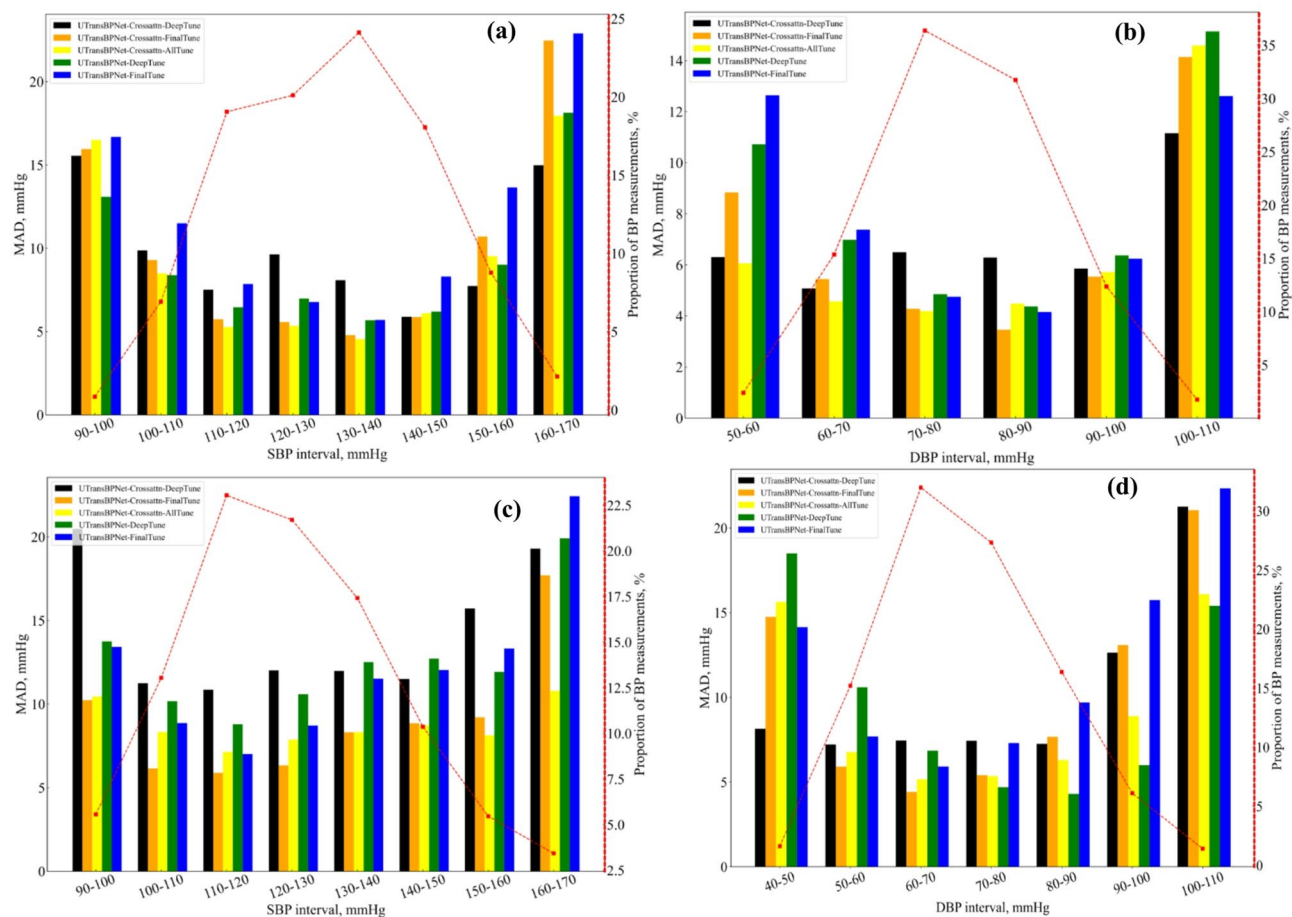


Fig. 4. MAD distributions at different BP intervals for five different models under (a-b) Drink and (c-d) Exercise scenarios.

Cross-scenario Tests (Training → Testing)	SBP		DBP	
	MAD, mmHg	Mean PCC	MAD, mmHg	Mean PCC
w/o Scenario-specific Finetuning				
1. Drinking → Exercise	24.13	0.12	15.65	0.02
2. Exercise → Drinking	13.70	0.01	9.52	0.16
3. Drinking + Exercise → MIMIC	16.98	0.02	12.89	0.01
with Scenario-specific Finetuning				
1. Drinking → Exercise	13.16	0.55	8.91	0.56
2. Exercise → Drinking	11.79	0.46	7.62	0.50
3. Drinking and Exercise → MIMIC	4.18	0.76	2.15	0.68

Table 3. Performance of *UTransBPNet_Crossattn_AllTune* under cross-scenario tests.

Influence of dataset complexity. To further explore influential factors of model performance, the correlation between the model’s MAD and three metrics of the test dataset, including the averaged individual BP changes, the mean and standard deviation of BP, were calculated for scenario-specific and cross-scenario conditions as illustrated in Fig. 5. A strong association was observed between model performance and the extent of activity-induced individual BP variations. In contrast, only a weak association was found between model performance and overall BP statistics of the datasets. This finding suggests the large impact of individual BP changes on model performance which may be ignored in previous studies.

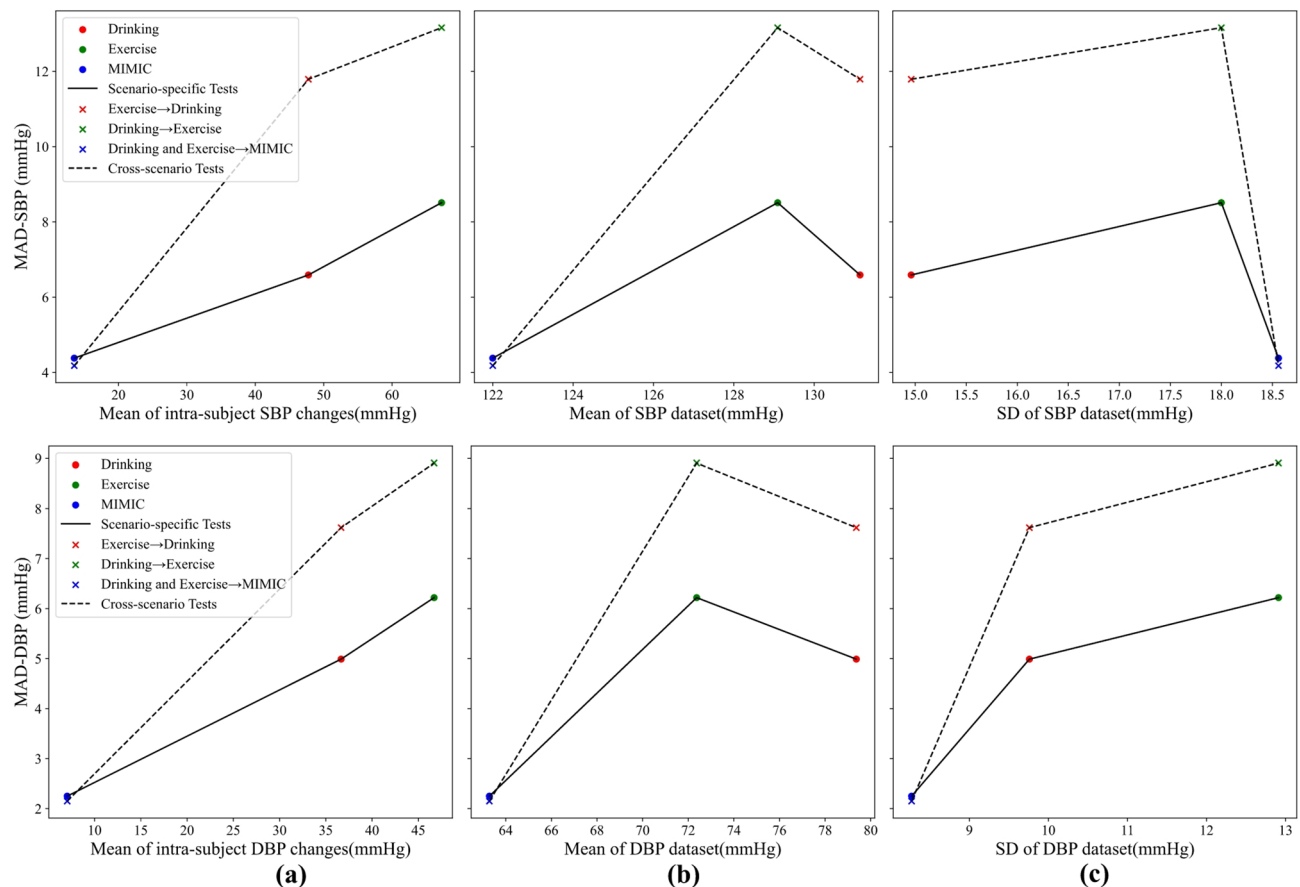


Fig. 5. Correlation between model performance and three metrics from the test datasets across both scenario-specific and cross-scenario tests for SBP and DBP: **(a)** average individual BP changes, **(b)** mean BP, and **(c)** BP standard deviation.

Discussion

Despite ongoing efforts in cuffless BP estimation modeling, accurately estimating BP remains challenging in conditions with substantial intra-individual BP variations induced by activities or interventions^{14,19}. This study introduces *UTransBPNet*, a population-based and calibration-free deep learning model, and rigorously validates its performance across several dynamic datasets in both scenario-specific and cross-scenario experimental settings.

Optimized short- and long-range feature representation. Compared to existing models such as *CNN-BiLSTM-Attn*, the proposed *UTransBPNet* leverages the advantages of transformer in long-range feature representation and the improved Unet in short-range feature representation, yielding improved performance for estimating and tracking BP variations under dynamic conditions. As illustrated in the two typical examples shown in Fig. 3, *UTransBPNet* captures both short- and long-range BP variation patterns that closely align with the reference BP, whereas *CNN-BiLSTM-Attn* displays suboptimal results with over-amplified short-range fluctuations. This highlights *UTransBPNet*'s superior capability in representing both feature ranges.

Moreover, introducing a cross-attention mechanism in *UTransBPNet* further reduces exaggerated high-frequency variations. Direct skip connections from the Unet encoder's short-range features to corresponding layers in the decoder may inadvertently introduce noisy short-range features, but the cross-attention mechanism, guided by the transformer's contextual feature map, optimizes these features to enhance BP tracking.

Optimal Finetuning Scheme. Additionally, the fine-tuning scheme of *UTransBPNet* consistently outperformed the detection scheme in estimating SBP and DBP from BP waveforms. Among the three fine-tuning schemes, AllTune, which uses features from the final layer and updates all model parameters, yielded the best performance. Several prior studies have also explored different feature maps from different Unet layers for SBP and DBP prediction. For example, Mahmud et al. used features from the deepest layer³³, while Yu et al. employed features from the final layer³⁴. However, these studies did not make direct comparisons. Our findings suggest that using features from the final layer and updating all model parameters may offer a more comprehensive representation for accurate SBP and DBP estimation.

Validation under Dynamic Conditions. Previous studies have validated models under dynamic conditions, but these often require individual calibration. For example, one study developed a nonlinear autoregressive exogenous model for BP estimation and tested it under daily activities³⁵. Although this model achieved satisfactory results with MADs of 6.79 and 5.31 mmHg for SBP and DBP, respectively, it required individual

data for calibration. In our previous work¹⁴, a CNN + Bi-GRU model was validated on Dataset_Drink, yielding poor results without individual fine-tuning, with MADs of 13.43 mmHg for SBP and 8.48 mmHg for DBP. These metrics only improved to 9.49 and 5.54 mmHg, respectively, when 10% individual data was used for fine-tuning. In addition, compared to *CNN-BiLSTM-Attn* and *CNN-BiGRU*, the proposed model achieved optimal individual PCCs, demonstrating robust tracking capabilities for intra-individual BP variations during activities.

For Dataset_MIMIC, a recent model, KD-informer, achieved state-of-the-art results²⁵. A key factor in KD-informer's success is the use of hand-crafted PPG morphological features, which improved performance on a private dataset from -0.031 ± 6.315 to 0.011 ± 4.453 mmHg for SBP and from 0.013 ± 6.237 to 0.046 ± 7.652 mmHg for DBP²⁵. In contrast, our model eliminates the need for labor-intensive feature extraction from PPG signals. Additionally, we excluded very short segments (< 8 min) from the original MIMIC dataset, ensuring sufficient BP fluctuations within each instance and allowing us to assess BP variation tracking over longer time periods. While *UTransBPNet* contains significantly more parameters than KD-informer (32.55 M vs. 0.81 M), future work should aim to reduce model size for greater computational efficiency.

Influences of Dataset Characteristics. Despite its importance for data-driven approaches, the impact of dataset characteristics on the generalization capability of BP estimation models has been explored in only a few studies³⁶. Our findings identify several influential factors that significantly affect model performance: (1) *Distribution shift*. As shown in Fig. 3, the model's performance degrades significantly when test samples fall outside the BP range covered by the training dataset. This finding highlights the importance of ensuring that training datasets adequately represent the full spectrum of physiological variability encountered in real-world scenarios. Expanding the training data distribution or incorporating strategies that enhance model's ability to generalize to out-of-distribution data can improve its robustness across different subjects. (2) *Distribution imbalance*. As shown in Fig. 4, the model tends to perform less accurately at the extreme ends of the BP range. To address this, future models should be trained on datasets with sufficient samples across the entire BP spectrum to ensure robust performance, especially at these extremes. (3) *Dataset complexity*. Intra-subject BP variability also impacts model performance, as demonstrated in Fig. 5. Varying degrees of BP deviation from individual's baseline may involve different physiological regulatory mechanisms, leading to a more complex relationship and nonlinear relationship between the input signals and BP. As a result, datasets with larger individual BP fluctuations, presents greater challenges for accurate estimation. Given these findings, we strongly recommend that new data-driven models undergo rigorous evaluation across diverse data distributions and complexities. This will ensure their robustness and reliability in real-world applications.

These factors can help explain the degraded performance in Dataset_Exercise, which did not meet the performance requirements set by the AAMI and IEEE standards. Specifically, Dataset_Exercise has a broader BP range and exhibits a long-tail distribution, particularly at the extreme BP values. These extremes are represented by very few samples, making the model more susceptible to distribution shifts between training and testing subsets. Additionally, this dataset has a smaller sample size compared to Dataset_Drink (28,814 vs. 62,678 segments), further exacerbating the effects of data imbalance. Moreover, intra-subject BP variability in Dataset_Exercise is substantial, making accurate estimation significantly more challenging than the other two datasets. Despite these difficulties, our model, *UTransBPNet*-Crossattn-AllTune, achieved mean absolute errors (MAEs) of 8.51 mmHg for SBP and 6.22 mmHg for DBP, which are considerably lower than the population average baseline (14.59 and 9.78 mmHg, respectively), and also outperformed other state-of-the-art models, as shown in Table 2.

Furthermore, PPG acquisition setup including acquisition mode (transmissive or reflective) and wavelength can also lead to differences in PPG morphology, potentially challenging the model's ability to generalize across datasets. However, our findings indicate that despite the differences in acquisition setup between Dataset_MIMIC and the other two datasets, the model demonstrates strong generalization from Dataset_Drink and Dataset_Exercise to Dataset_MIMIC, as shown in Table 3. This suggests that signal normalization plays a crucial role in standardizing amplitude variations across datasets, resulting in a marginal effect on the model's overall generalization capability. On the other hand, distribution shift, distribution imbalance, and dataset complexity have a much more significant impact on the model's performance.

While the proposed model demonstrated state-of-the-art performance in scenario-specific validation, its generalization across different scenarios remains limited. Future research should focus on developing advanced transfer learning techniques to improve the generalization capability across different scenarios. Additionally, the model's large size needs significant reduction for deployment on embedded devices. Notable progress has been made in this area through knowledge distillation techniques²⁵, and our work may provide an accurate teacher model that could inform a smaller, efficient student model without sacrificing accuracy.

Conclusion

In conclusion, this study introduces *UTransBPNet*, a novel, calibration-free deep learning model for cuffless BP estimation. It combines squeeze-and-excitation enhanced Unet and transformer architectures to effectively capture both short- and long-range BP variations. Extensive validation on dynamic datasets demonstrated that *UTransBPNet* significantly outperformed traditional models in scenario-specific conditions. This study also reveals several dataset characteristics that significantly influence model performance. In particular, distribution imbalance, distribution shifts, and individual BP variability strongly influenced model accuracy. These findings emphasized the need for well-distributed, representative data as well as comprehensive validation under highly dynamic datasets to ensure reliable BP estimation in real-life scenarios.

Data availability

The datasets used and/or analyzed during the current study available from the corresponding author on reasonable request.

Received: 3 December 2024; Accepted: 16 May 2025

Published online: 21 May 2025

References

1. Organization, W. H. *Fact sheets on Hypertension*. Available. (2021). Available: <https://www.who.int/news-room/fact-sheets/detail/hypertension>
2. Schutte, A. E., Kollias, A. & Stergiou, G. S. Blood pressure and its variability: classic and novel measurement techniques. *Nat. Rev. Cardiol.* **19** (10), 643–654 (Oct 2022).
3. Mukkamala, R., Stergiou, G. S. & Avolio, A. P. Cuffless blood pressure measurement. *Annu. Rev. Biomed. Eng.* **24** (1), 203–230 (Jun 2022).
4. Sayer, G. et al. Continuous monitoring of blood pressure using a Wrist-Worn cuffless device, (in eng). *Am. J. Hypertens.* **35** (5), 407–413 (May 2022).
5. Pellaton, C. et al. Accuracy testing of a new optical device for noninvasive Estimation of systolic and diastolic blood pressure compared to intra-arterial measurements. *Blood Press. Monit.* **25** (2), 105–109 (Apr 2020).
6. Nyvad, J., Christensen, K. L., Buus, N. H. & Reinhard, M. The cuffless somnotouch NIBP device shows poor agreement with a validated oscillometric device during 24-h ambulatory blood pressure monitoring, (in eng). *J. Clin. Hypertens. (Greenwich)*. **23** (1), 61–70 (Jan 2021).
7. Ding, X. R. & Zhang, Y. T. Photoplethysmogram intensity ratio: A potential indicator for improving the accuracy of PTT-based cuffless blood pressure estimation, (in eng). *Annu Int Conf IEEE Eng Med Biol Soc*, vol. pp. 398–401, 2015. (2015).
8. Liu, Q., Yan, B. P., Yu, C., Zhang, Y. & Poon, C. Y. C. Attenuation of systolic blood pressure and pulse transit time hysteresis during exercise and recovery in cardiovascular patients. *Biomedical Eng. IEEE Trans. On.* **61** (2), 346–352 (2013).
9. Ma, Y. et al. Relation between blood pressure and pulse wave velocity for human arteries. *Proc. Natl. Acad. Sci. U.S.A.* **115** (44), 11144–11149 (2018).
10. Pandit, J. A., Lores, E. & Battle, D. Cuffless blood pressure monitoring: promises and challenges. *Clin. J. Am. Soc. Nephrol.* **15** (10), 1531–1538 (2020).
11. El-Hajj, C. & Kyriacou, P. A. A review of machine learning techniques in photoplethysmography for the non-invasive cuff-less measurement of blood pressure. *Biomed. Signal. Process. Control.* **58**, 101870 (Apr 2020).
12. Senturk, U., Polat, K. & Yucedag, I. A non-invasive continuous cuffless blood pressure Estimation using dynamic recurrent neural networks. *Applied Acoustics*, **170**, Dec 15 2020, Art. 107534.
13. Liu, Q., Zheng, Y., Zhang, Y. & Poon, C. C. Y. Beats-to-Beats Estimation of blood pressure during supine cycling exercise using a probabilistic nonparametric method. *IEEE Access.* **9**, 115655–115663 (Aug 2021).
14. 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, in *IEEE 17th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, Athens, Greece, 2021, pp. 1–4., Athens, Greece, 2021, pp. 1–4. (2021).
15. Song, K., Chung, K. & Chang, J. H. Cuffless deep Learning-Based blood pressure Estimation for smart wristwatches. *IEEE Trans. Instrum. Meas.* **69** (7), 4292–4302 (Oct 2020).
16. Wang, W., Mohseni, P., Kilgore, K. L. & Najafizadeh, L. Cuff-Less blood pressure Estimation from photoplethysmography via visibility graph and transfer learning. *IEEE J. Biomed. Health Inform.* **26**, 5, pp. 2075–2085, May 2022.
17. Malayeri, A. B. & Khodabakhshi, M. B. Concatenated convolutional neural network model for cuffless blood pressure Estimation using fuzzy recurrence properties of photoplethysmogram signals. *Sci. Rep.* **12** (1), 6633 (Apr 2022).
18. Zhang, G., Choi, D. & Jung, J. Reconstruction of arterial blood pressure waveforms based on squeeze-and-excitation network models using electrocardiography and photoplethysmography signals. *Knowl. Based Syst.* **295**, 111798 (Jul 2024).
19. Lee, D. et al. Beat-to-beat continuous blood pressure estimation using bidirectional long short-term memory network, *Sensors*, vol. 21, no. 1, Dec (2021).
20. Eom, H. et al. End-To-End Deep Learning Architecture for Continuous Blood Pressure Estimation Using Attention Mechanism, *Sensors*, vol. 20, no. 8, p. 2338, Apr (2020).
21. Jeong, D. U. & Lim, K. M. Combined deep CNN–LSTM network-based multitasking learning architecture for noninvasive continuous blood pressure Estimation using difference in ECG–PPG features. *Sci. Rep.* **11** (1), 13539 (Jun 2021).
22. Baker, S., Xiang, W. & Atkinson, I. A hybrid neural network for continuous and non-invasive Estimation of blood pressure from Raw electrocardiogram and photoplethysmogram waveforms. *Comput. Methods Programs Biomed.* **207**, 106191 (Aug 2021).
23. Yang, S., Zhang, Y., Cho, S. Y., Correia, R. & Morgan, S. P. Non-invasive cuff-less blood pressure Estimation using a hybrid deep learning model. *Opt. Quant. Electron.* **53** (2), 93 (Jan 2021).
24. Zhang, G., Choi, D. & Jung, J. Development of continuous cuffless blood pressure prediction platform using enhanced 1-D SENet–LSTM. *Expert Syst. Appl.* **242**, 122812 (May 2024).
25. Ma, C. et al. KD-Informer: Cuff-Less continuous blood pressure waveform Estimation approach based on single photoplethysmography. *IEEE J. Biomed. Health Inf.* **27** (5), 2219–2230 (Jun 2022).
26. Mukkamala, R. et al. Evaluation of the Accuracy of Cuffless Blood Pressure Measurement Devices: Challenges and Proposals, (in eng). *Hypertension*, vol. 78, no. 5, pp. 1161–1167, Nov (2021).
27. Petit, O. et al. U-Net transformer: self and cross attention for medical image segmentation. In *Machine Learning in Medical Imaging* 267–276 (Springer International Publishing, 2021).
28. Zheng, Y., Poon, C. C. Y. & Zhang, Y. T. Investigation of temporal relationship between cardiovascular variables for cuffless blood pressure estimation, in *Proceedings of 2012 IEEE-EMBS International Conference on Biomedical and Health Informatics*, pp. 644–646. (2012).
29. Liu, Q., Yan, B. P., Yu, C. M., Zhang, Y. T. & Poon, C. C. Attenuation of systolic blood pressure and pulse transit time hysteresis during exercise and recovery in cardiovascular patients. *IEEE Trans Biomed Eng.* vol. 61, no. 2, pp. 346–52, Feb (2014).
30. Kachuee, M., Kiani, M. M., Mohammadzade, H. & Shabany, M. Cuff-less high-accuracy calibration-free blood pressure estimation using pulse transit time, in *IEEE International Symposium on Circuits and Systems (ISCAS)*, 2015, pp. 1006–1009., 2015, pp. 1006–1009. (2015).
31. Vaswani, A. et al. Attention is all you need. *Advances Neural Inform. Process. Systems*, **30**, (2017).
32. Jamil, Z., Lui, L. T. & Chan, R. H. M. Blood pressure Estimation using Self-Attention mechanism Built-In resnet on pulsedb: demographic fairness and generalization. *IEEE Sens. J.* **25** (1), 1694–1705 (2025).
33. Mahmud, S. et al. A Shallow U-Net Architecture for Reliably Predicting Blood Pressure (BP) from Photoplethysmogram (PPG) and Electrocardiogram (ECG) Signals, *Sensors*, vol. 22, no. 3, p. 919, Jan (2022).
34. Yu, M., Huang, Z., Zhu, Y., Zhou, P. & Zhu, J. Attention-based residual improved U-Net model for continuous blood pressure monitoring by using photoplethysmography signal. *Biomed. Signal. Process. Control.* **75**, 103581 (May 2022).

35. Li, Y. H., Harfiya, L. N., Purwandari, K. & Lin, Y. D. Real-Time Cuffless Continuous Blood Pressure Estimation Using Deep Learning Model, *Sensors*, vol. 20, no. 19, p. 5606, (2020).
36. Cheung, M. Y., Sabharwal, A., Coté, G. L. & Veeraraghavan, A. Wearable blood pressure monitoring devices: Understanding heterogeneity in design and evaluation. *IEEE Trans. Biomed. Eng.* **71** (12), 3569–3592 (2024).

Acknowledgements

This work was supported by Young Scientists Fund from National Natural Science Foundation of China (NSFC) [62301333], XJTU Research Development Fund [RDF-21-02-068], Research Foundation of Education Department of Guangdong Province [2022ZDJS115], National Science Foundation of China (NSFC) (62373259), and Common University Innovation Team Project of Guangdong [2021 KCXTD041].

Author contributions

Y.L.Z. wrote the main manuscript text and contributed to conceptualization. H.H. and J.G. developed the algorithm and prepared the figures and performed visualization. J.H. and S.W. conducted data preprocessing. Y.T.Z. and Q.L. contributed to conceptualization and manuscript review. All authors reviewed the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-02963-3>.

Correspondence and requests for materials should be addressed to Y.Z. or Q.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025