Dual-Path Cuffless PPG-based Blood Pressure Estimation using Conformer & Swin Transformer

Caoyueshan Fan, Yiting Wei, *Student Member, IEEE*, Melanie Qiu, Mostafa Haghi, *Member, IEEE*, and Nima TaheriNejad, *Member, IEEE*

***Abstract*— This study introduces a novel dual-path deep learning framework using Photoplethysmogram (PPG) signals to address key challenges in continuous, non-invasive cuffless Blood Pressure (BP) monitoring. To this end, we introduce -for the first time- the use of two novel deep neural network architectures: Conformer-Transformer and 1D Swin Transformer. These architectures are adapted here to model both the morphological structure and rhythmic dynamics of PPG signals. This cross-domain transfer enables Arterial Blood Pressure (ABP) waveform reconstruction and significantly improves the accuracy and physiological consistency of Systolic Blood Pressure (SBP) and Diastolic Blood Pressure (DBP) estimation. Extensive experiments on two public datasets demonstrate that our methods consistently outperform mainstream baselines across multiple key metrics. Specifically, the Conformer-Transformer achieved the lowest Mean Absolute Error (MAE) of 2.979 mmHg for systolic and 1.603 mmHg for diastolic BP, improving upon previous studies by 9.6% and 8.4%, respectively, while delivering the best waveform reconstruction performance too. The Swin Transformer achieved a systolic MAE of 3.034 mmHg and a diastolic MAE of 1.714 mmHg. All experimental results conform to the British Hypertension Society (BHS) grade A and Association for the Advancement of Medical Instrumentation (AAMI) standards.**

***Index Terms*— Arterial blood pressure, conformer, cuffless blood pressure estimation, photoplethysmogram, swin transformer**

I. Introduction

According to the World Health Organization's Global Report on Hypertension, it now affects over 1.3 billion people and is responsible for approximately 10 million deaths annually [1]. Related complications, such as cardiovascular, cerebrovascular, ocular, metabolic, and other systemic effects, remain a leading cause of global mortality and morbidity [2]. Thus, the need for reliable and continuous Blood Pressure (BP) monitoring is critical to enable early detection, timely intervention, and effective management of hypertension-related health risks.

Compared to discrete BP estimation, continuous monitoring can contribute to long-term BP trends detection and offer deeper clinical insight. However, current mainstream methods still rely on cuff-based and intermittent devices [3]. While widely accepted, for out-of-the-lab and in-home monitoring, these tools face clear limitations, including user inconvenience, sleep disruption, movement restrictions, and discomfort [4]. As a result, both academia and industry have turned to wearable, cuffless, and non-invasive approaches for continuous BP monitoring. Among these, techniques leveraging the correlation between Pulse Transit Time (PTT) or Pulse Arrival Time (PAT) and BP have shown particular promise. These methods require the acquisition of signals from two sites to estimate vascular transit times, most commonly achieved through the simultaneous measurement of Photoplethysmogram (PPG) and Electrocardiography (ECG) [5]. These systems enable fine-grained temporal tracking of BP and show promise for early identification of pathological patterns [6]. However, obtaining stable and low-noise ECG signals in real-world settings remains a persistent technical challenge [7]. PPG offers several advantages: it is non-invasive, easy to acquire, low-cost, and well-suited to wearable form factors [8], making it a compelling single-modality signal source for continuous, cuffless BP estimation.

Traditional machine learning methods played an early role in cuffless BP estimation [9]. While handcrafted feature-based algorithms showed some promise under certain conditions, they often struggle to model the complex and variable nature of PPG signals, especially in capturing long-term dependencies and nonlinear interactions. Moreover, their generalizability across datasets remains weak. We believe that improving the accuracy of cuffless BP estimation requires not only more advanced model architectures but also a deeper understanding of the two core types of information embedded in PPG: morphological features and inter-cycle rhythm dynamics. However, most existing deep learning approaches tend to focus on only one of these aspects, and few are designed to capture both.

To address this gap, we explore the potential of cross-domain architectural transfer for physiological signal modeling. Specifically, we propose a dual-pathway modeling strategy for PPG: **Conformer-Transformer architecture:** By combining local convolutional layers and global attention mechanisms, this design can jointly extract morphological and rhythmic features. A Transformer decoder then reconstructs the ABP waveform and estimates corresponding BP. **Swin Transformer architecture:** A hierarchical window-based attention mechanism is used to extract multi-scale morphological structures from the PPG signal. Through sliding window operations, the model performs cross-cycle modeling, enabling ABP reconstruction and SBP/DBP estimation.

All Authors are with the Heidelberg University, Heidelberg, 69120 Germany (e-mail: firstname.lastname@ziti.uni-heidelberg.de). (Corresponding author: Yiting Wei)

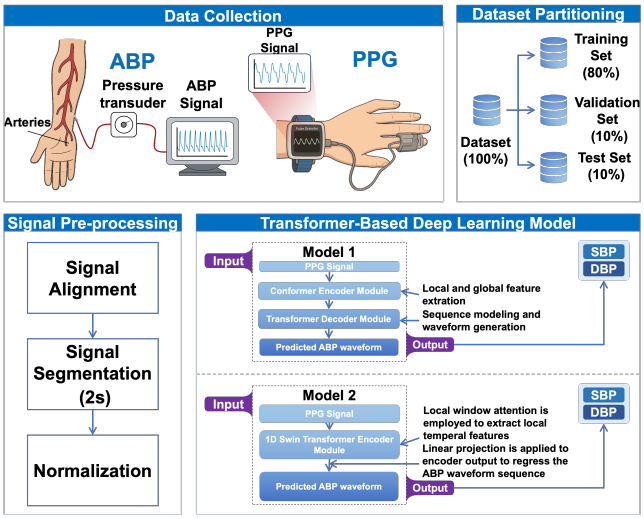


Fig. 1. Overall framework of our proposed method.

Fig. 1 introduces the overall framework of our work. The main contributions of our work are as follows: **a)** To the best of our knowledge, this is the first work to introduce Conformer and Swin Transformer architectures for cuffless BP estimation. **b)** We propose a dual-output learning objective that enables simultaneous ABP waveform reconstruction and SBP/DBP value regression, thereby enhancing the clinical interpretability of predictions. **c)** We analyze the representational demands of PPG signals in terms of both morphology and rhythm, and propose a structure-aware framework to address them jointly. **d)** We provide insight into structure-task alignment, demonstrating that the Conformer-Transformer excels in long-term BP rhythm modeling and ABP waveform reconstruction, while the Swin Transformer offers superior estimation accuracy and robustness, making it suitable for lightweight deployment.

II. Related Work

The following representative works on cuffless BP monitoring based only on PPG signals: El-Hajj and Kyriacou used a BiLSTM-Attention model capturing temporal patterns, it is effective in short-term but weak in rhythm continuity [10]. Hasanzadeh et al. built a feature-based ML framework, which is accurate but dependent on handcrafted features and lacks scalability [11]. Haddad et al. designed a lightweight deep model for edge use, but with limited rhythm and continuity modeling [12]. Wang et al. applied visibility graphs with transfer learning, which comes with extra complexity and risk of information loss [13]. Kim et al. proposed DeepCNAP for ABP regression that is highly accurate but sensitive to shifts and waveform diversity [14]. Qiu et al. created a hybrid regression–piecewise model with balanced accuracy and efficiency, but weaker for dynamic rhythms [15]. Ma et al. introduced KD-Informer with long-term attention for PPG-to-ABP, improving fidelity but needing simplification for deployment [16]. Leitner et al. used transfer learning for personalized BP, promising in small samples but dependent on pretraining data [17]. Panwar et al. proposed PP-Net, a compact CRNN for joint SBP, DBP and Heart Rate (HR) estimation from PPG, with good deployability but sensitive to noisy input [18].

Despite progress, prior approaches share several limitations: they often model either temporal dynamics or handcrafted morphology alone. Lightweight designs improve deployability but weaken the capture of long-term dependency. Visibility graphs and transfer learning add complexity and risk information loss. Many deep models achieve short-term accuracy yet struggle with variability, noise, and generalization. Here, we propose a dual-path framework to jointly capture fine-grained PPG morphology and long-range rhythmic dynamics, thereby improving waveform fidelity, BP estimation accuracy, and robustness in a physiologically consistent manner.

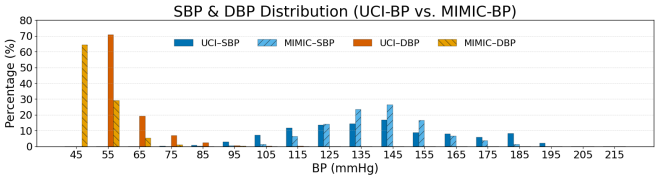


Fig. 2. Distribution of used datasets for SBP and DBP.

III. Materials and Methods

A. Datasets

1) UCI-BP Dataset: It is sourced from the University of California, Irvine (UCI) Machine Learning Repository, and originates from a curated subset of the MIMIC-II Waveform Database [19]. It consists of 12,000 recordings, each ranging from 8 to 592 seconds. All records include synchronized PPG, ECG, and ABP signals sampled at 125 Hz. The ABP signals are acquired via invasive radial arterial catheterization, which is widely regarded as the clinical gold standard for continuous and high-precision BP monitoring.

2) MIMIC-BP Dataset: MIMIC-BP is derived from the MIMIC-III Waveform Database Matched Subset and is curated for cuffless BP estimation [20]. It includes data from 1,524 Intensive Care Unit (ICU) patients, with each subject contributing 30 segments of 30-second synchronized recordings. Each segment contains ABP, PPG, ECG, and Respiration (RESP) signals, all sampled at 125 Hz, yielding over 380 hours of physiological data in total. The latest release expands subject diversity and measurement conditions, providing a high-temporal-resolution benchmark. In all experiments, we use PPG-only input; for both datasets, BP labels are derived from invasively measured ABP waveforms, ensuring accuracy and real-time reliability.

Fig. 2 shows the distributions of the two selected dataset.

B. Data Preprocessing

1) Dataset Preprocessing: Each dataset was split into training, validation, and test sets with an 8:1:1 ratio at the recording level, ensuring that no segment from the same sequence appeared across different splits. For UCI-BP, all recordings (8–592 s) were retained without any length-based exclusion.

2) Signal Preprocessing: The UCI-BP dataset was provided with preprocessing by the authors [19], including smoothing, removal of blocks with abnormal BP or HR, elimination of unresolved discontinuities, and autocorrelation filtering for pulse-to-pulse variability. For MIMIC-BP, we applied additional denoising: PPG signals were filtered with a 3rd-order Butterworth band-pass (0.5–8 Hz) and smoothed with a moving average, while ABP signals were filtered with a 2nd-order Butterworth (0.4–12 Hz) and the same smoothing. A

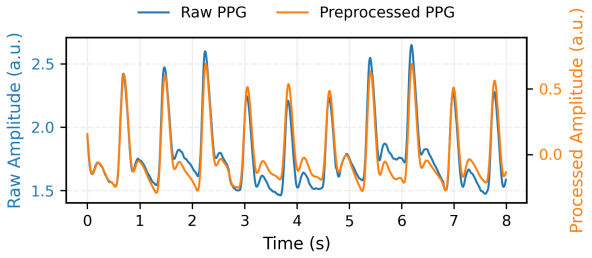


Fig. 3. Raw and preprocessed PPG of a MIMIC-BP sample.

low-pass branch was further designed to extract ABP baseline components. The comparison of PPG signals before and after filtering is shown in Fig. 3.

3) Alignment: For training and validation sets, ABP and PPG signals were aligned using cross-correlation by shifting ABP to maximize correlation with PPG. No alignment was applied to the test set to preserve independence.

4) Segmentation: Aligned PPG and ABP signals were segmented into fixed 2-second non-overlapping windows at 125 Hz. Fragments shorter than one window were discarded, as a 2-second window typically covers at least one cardiac cycle and preserves key waveform features for BP estimation [21].

5) Normalization: For PPG inputs, z-score normalization was applied independently in training, validation, and test sets using each segment's mean and standard deviation, avoiding data leakage. For ABP in the training set, per-segment z-score normalization stabilized training and unified the dynamic range. During inference, predicted ABP waveforms were rescaled to original amplitude, while ABP signals in the test set remained unprocessed as reference labels.

C. Model Architecture

PPG is a nonlinear and non-stationary pulsatile waveform [22]. Accurate BP estimation therefore requires not only capturing the morphological features but also modeling the rhythm dynamics across cycles. We argue that effective PPG modeling should go beyond predicting BP values alone. It should uncover the internal structure of the signal and extract multi-scale and multi-modal representations that reflect underlying physiological changes. Hence, we explore two structure-aware deep neural architectures:

1) Conformer-Transformer Model for ABP Waveform Estimation: To effectively capture both multi-scale morphological patterns and temporal dependencies embedded in PPG signals, we propose a deep regression architecture based on a Conformer-Transformer structure for continuous ABP waveform estimation. This model consists of two main components: a Conformer-based encoder [23] and a Transformer-based decoder. The complete model structure is shown in Fig. 4.

a) Encoder, Conformer Blocks: The Conformer block integrates Multi-Head Self-Attention (MHSA) and convolutional modules, sandwiched between two Macaron-style feed-forward networks (FFNs). This enables joint modeling of global rhythmic dependencies and local waveform features.

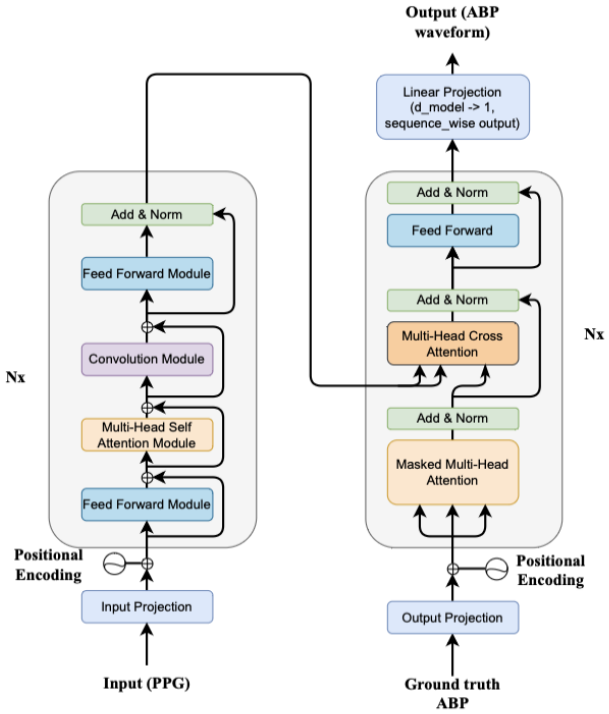


Fig. 4. Architecture of the Conformer-Transformer model.

The output of each Conformer block y_i is:

$$\tilde{x}_i = x_i + \frac{1}{2}\mathrm{FFN}(x_i) \tag{1}$$

$$x'_i = \tilde{x}_i + \mathrm{MHSA}(\tilde{x}_i) \tag{2}$$

$$x''_i = x'_i + \mathrm{Conv}(x'_i) \tag{3}$$

$$y_i = \mathrm{LayerNorm}\left(x''_i + \frac{1}{2}\mathrm{FFN}(x''_i)\right) \tag{4}$$

where, MHSA captures global periodic structures with relative position bias; The convolution module includes GLU, 1D depthwise convolution, batch normalization, and Swish; Two FFNs are placed before and after attention–convolution blocks for balanced nonlinearity. This enables the encoder to extract both local waveform morphology and long-range dynamics.

b) Decoder, Transformer Decoder: The Transformer decoder is composed of a stack of layers, each containing Masked Multi-Head Self-Attention (auto-regressive modeling), Encoder-Decoder Cross Attention (context fusion from encoder outputs), and Position-wise Feed Forward Network (nonlinear transformation). Each sublayer is wrapped with residual connection and LayerNorm, i.e.,

$$\mathrm{Output} = \mathrm{LayerNorm}(x + \mathrm{Sublayer}(x)). \tag{5}$$

The decoder first performs masked attention over previously generated outputs, then applies cross-attention to integrate the encoder-derived PPG context. Finally, each time step's hidden vector is passed through a sequence-wise linear projection to yield the predicted ABP waveform.

c) Input and Output Definition: PPG sequence of length T, denoted as $X \in \mathbb{R}^T$ is the input and Predicted ABP waveform $\hat{Y} \in \mathbb{R}^T$ is the output. The model is trained via the mean

TABLE I
SUMMARY OF MODEL STRUCTURES AND HYPERPARAMETERS

Hyperparameter	Value/Setting (Conformer)	Value/Setting (Swin1D)
Optimizer / Learning Rate	AdamW, lr = 3×10^{-4}	AdamW, lr = 3×10^{-4}
Batch Size / Epochs	batch = 128, epochs = 50	batch = 128, epochs = 50
Loss Function	Huber loss (δ = 1.0)	Huber loss (δ = 1.0)
Model Scale	Encoder: 4 Conformer blocks (d = 128, heads = 8, kernel = 31); Decoder: 4 Transformer blocks	1D Swin Transformer, 4 layers, d = 128, heads = 8
Input Window Size	2-s segments @ 125 Hz	2-s segments @ 125 Hz

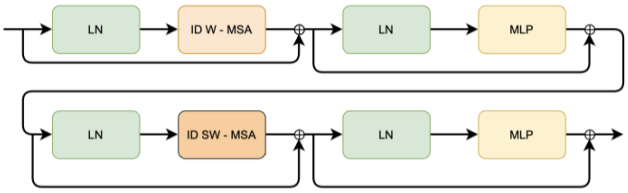


Fig. 5. Structure of Swin1D Block.

squared error (MSE) loss:

$$\mathcal{L}_{\text{MSE}} = \frac{1}{T}\sum_{i=1}^{T}(Y_i - \hat{Y}_i)^2, \tag{6}$$

where $\mathbf{Y}$ and $\hat{\mathbf{Y}}$ are the reference ABP and the predicted waveform, respectively. Table I shows an overview of the model structure and hyperparameters.

2) 1D Swin Transformer-Based Model for ABP Estimation: Our second model adopts a 1D Swin Transformer architecture [24], which enables end-to-end regression from raw PPG signals to ABP waveforms. We restructured the Swin Transformer from its original 2D-vision design into a one-dimensional hierarchical architecture tailored for PPG-based blood pressure estimation. This adaptation enables effective multi-scale temporal modeling, where short windows capture single-beat details and long windows integrate cross-beat dynamics, enhanced by a shifted-window mechanism for reliable estimation. Each Swin1D Block consists of two sub-blocks, and its structure is shown in Fig. 5:

- 1D Window-based Multi-Head Self-Attention (W-MSA): Performs self-attention within non-overlapping windows to capture localized features.
- 1D Shifted Window Multi-Head Self-Attention (SW-MSA): Shifts the window partition to enable cross-window interaction and enhance global receptive fields.

Both sub-blocks follow the standard transformer module pattern: LayerNorm (LN); Attention module (W-MSA or SW-MSA); Residual connection; LN; Two-layer Multi-Layer Perceptron (MLP) module; Residual connection. The final sequence representation $\mathbf{Z} \in \mathbb{R}^{T\times d}$ from the last Swin1D Block is projected through a Linear Projection layer to generate the output ABP waveform $\hat{\mathbf{Y}} \in \mathbb{R}^{T\times 1}$, matching the original input length.

This architecture allows the model to jointly capture the morphological characteristics and rhythmic dynamics, providing a dual-perspective modeling strategy for the complex structure of PPG signals. Table I shows a summary of the model structure and hyperparameters.

D. Performance Evaluation

To comprehensively assess model performance, we employed multiple standard evaluation metrics, including the Pearson correlation coefficient (R), Mean Error (ME), Mean Absolute Error (MAE), Standard Deviation (SD), and Root Mean Square Error (RMSE). To ensure clinical relevance and compliance, we further evaluate our models based on two authoritative standards:

- Association for the Advancement of Medical Instrumentation (AAMI) Standard [25]: Requires that the Mean Error (ME) should be less than or equal to ±5 mmHg and the Standard Deviation (SD) should be less than or equal to 8 mmHg, computed over at least 255 measurements.
- British Hypertension Society (BHS) Standard [26]: Classifies performance into four grades (A–D) based on the cumulative percentage of samples whose absolute errors fall within 5, 10, and 15 mmHg, as shown in Table II.

TABLE II
BHS GRADING CRITERIA

	≤5 mmHg	≤10 mmHg	≤15 mmHg	Grade
BHS	60%	85%	95%	A
	50%	75%	90%	B
	40%	65%	85%	C

IV. RESULTS AND DISCUSSION

A. Overall Performance

To evaluate the BP estimation performance of the proposed models, we conducted systematic experiments on two authoritative public datasets (UCI-BP and MIMIC-BP), including comparisons with classical approaches and recent literature. The results, summarized in Tables III – VIII, cover key metrics, performance before and after denoising and perturbation, ABP reconstruction quality, and model complexity. A comprehensive review of these results shows the following:

Superior Prediction Accuracy and Robustness: Both models outperform published and classical methods across metrics. On the UCI-BP dataset, the Conformer-Transformer achieved the lowest SBP/DBP MAE (2.979/1.603 mmHg), improving 9.6%/8.4% over [14], with the best waveform reconstruction (MAE: 3.005 mmHg, R: 0.978). On the MIMIC-BP dataset, it reached 3.414/1.774 mmHg, while the Swin Transformer showed comparable DBP accuracy and even lower ME, indicating robustness to noise. As shown in Tables V, performance remained strong on raw PPG signals and further improved after denoising (e.g., Conformer-Transformer reduced SBP RMSE from 5.722 mmHg to 4.484 mmHg). These results confirm strong accuracy and robustness, especially for systolic prediction.

Generalization and Standards: Despite dataset differences, both models consistently achieved stable and superior performance across test sets (Conformer-Transformer: R = 0.986, MAE = 2.979 mmHg; Swin Transformer: R = 0.975, MAE = 3.526 mmHg), and consistently outperformed existing methods on both SBP and DBP. All results satisfied the BHS A-grade and AAMI standards, confirming their clinical applicability.

Interpretability and Physiological Consistency: The models achieved superior ABP waveform reconstruction compared to baseline methods, enhancing interpretability and clinical consistency. They focused on physiologically meaningful

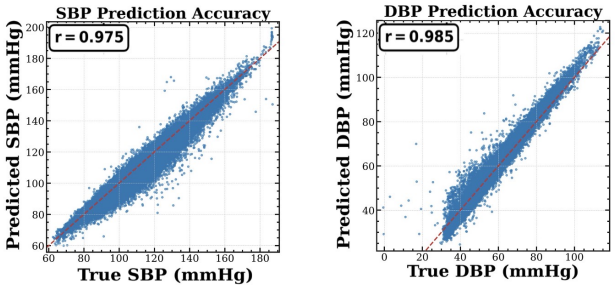


Fig. 6. MIMIC-BP regression plots using Swin Transformer.

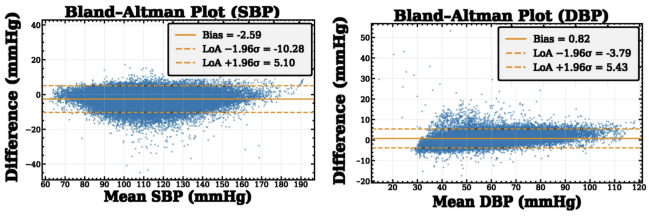


Fig. 7. Swin Transformer Bland-Altman plots on MIMIC-BP.

landmarks of the PPG waveform, but performance declined when rhythmic continuity was disrupted, indicating reliance on rhythm and confirming the physiological plausibility of the learned features.

Efficiency and Complementary Strengths: The Swin Transformer provides advantages in lightweight design, faster convergence, and lower latency, making it suitable for real-time deployment. In contrast, the Conformer-Transformer, while more computationally demanding, achieves higher precision by capturing long-term dependencies and fine-grained morphological features. Together, the two models offer complementary strengths and provide practical solutions for diverse real-world applications.

B. Correlation and Bias Analysis

As illustrated in Fig. 6, the results of the Swin Transformer on the MIMIC-BP dataset demonstrate a high degree of consistency between predicted and ground-truth values. Specifically, the Pearson correlation coefficients for both SBP and DBP exceed 0.95, indicating that the model effectively captures both the overall trends and fluctuations in actual BP values. Fig. 7 further supports this finding by examining the distribution of prediction bias. The majority of data points lie within an acceptable deviation range of ±5 mmHg, and ME is −2.589 mmHg for SBP and 0.821 mmHg for DBP, respectively. This suggests that the model does not exhibit any systematic tendency toward overestimation or underestimation. It is worth noting that similar patterns were consistently observed across all experimental settings.

C. Error Distribution and Statistical Stability Analysis

Fig. 8 presents the histograms of prediction error distributions for the two proposed models across both datasets, encompassing four experimental configurations in total. Overall, the error distributions for both SBP and DBP predictions exhibit a bell-shaped, centralized pattern, indicating that the majority of prediction errors fall within the ±5 mmHg range. On

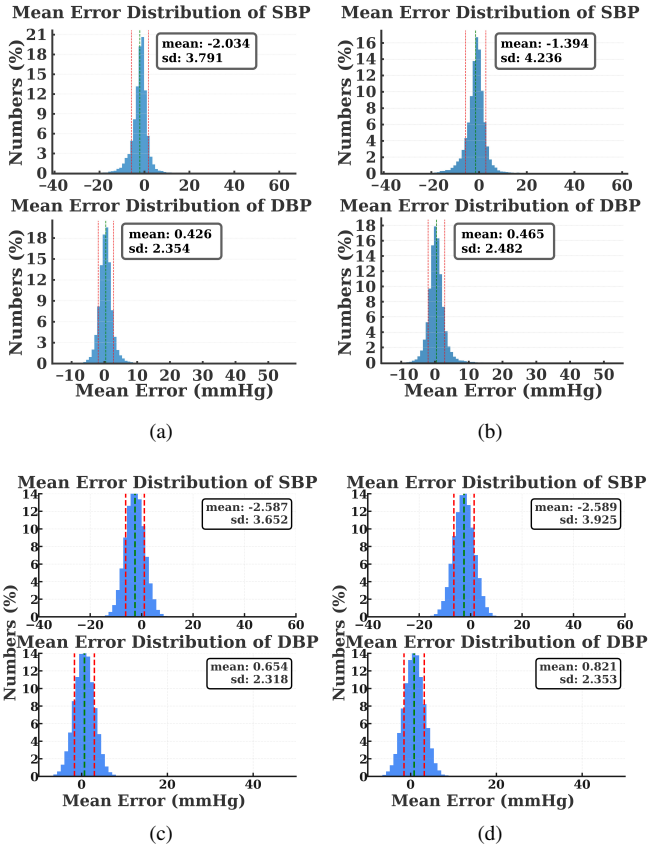


Fig. 8. The mean error distribution of two datasets' SBP and DBP for (a) UCI-BP using Conformer-Transformer. (b) UCI-BP using Swin Transformer. (c) MIMIC-BP using Conformer-Transformer. (d) MIMIC-BP using Swin Transformer.

the UCI-BP dataset, the Conformer-Transformer demonstrates a smoother error distribution, with a slightly lower SD for SBP compared to the Swin Transformer. This suggests a stronger capacity for modeling complex rhythmic patterns in populations with higher signal variability. On the MIMIC-BP dataset, although the Conformer-Transformer achieves slightly lower mean errors, the Swin Transformer exhibits comparable performance in DBP prediction and achieves competitive standard deviations. This suggests that the Swin Transformer provides enhanced robustness to local fluctuations and noise, aligning well with its lightweight, window-based attention design. These findings validate the complementary strengths of the two model architectures across varying populations and data conditions, highlighting the robustness and consistency of the proposed framework in diverse real-world scenarios.

D. ABP Waveform Reconstruction Performance

Fig. 9 illustrates the ABP waveform prediction results of the Conformer-Transformer and Swin Transformer on the UCI-BP dataset. The following observations can be made: Both models closely match the key morphological features of the waveform, such as systolic peaks, diastolic troughs, and dicrotic notches—across multiple cardiac cycles; Even in segments with substantial rhythmic disturbances, the models maintain stable tracking of the primary waveform trend. In addition, we quantitatively evaluated the reconstructed ABP waveforms using multiple performance metrics, as summarized in Table VI. The results indicate that both models consistently

TABLE III
PERFORMANCE COMPARISON AND AAMI COMPLIANCE EVALUATION OF PUBLISHED METHODS FOR UCI-BP AND MIMIC-BP

Dataset	Method	Model	SBP (mmHg)						DBP (mmHg)					
			R	ME	MAE	SD	RMSE	AAMI	R	ME	MAE	SD	RMSE	AAMI
UCI-BP	[13]	Visibility Graph+CNN+Ridge Regression	0.880	0.000	6.170	8.460	8.460	Fail	0.840	0.040	3.660	5.360	5.360	Pass
	[14]	CNN+Transformer	0.964	1.230	3.400	5.400	5.490	Pass	0.949	-0.530	1.750	2.810	2.820	Pass
	[27]	RDAE	–	1.648	5.424	6.640	–	Pass	–	-1.280	3.144	3.740	–	Pass
	[28]	LSTM-based Autoencoder	–	4.050	4.050	4.600	5.250	Pass	–	2.410	2.410	3.110	3.170	Pass
	Ours	Conformer-Transformer	**0.986**	-2.034	**2.979**	**3.791**	**4.303**	Pass	**0.979**	0.426	**1.603**	**2.354**	**2.362**	Pass
	Ours	Swin Transformer	0.982	-1.394	3.034	4.236	4.459	Pass	0.978	0.465	1.714	2.482	2.525	Pass
MIMIC-BP		CNN	0.957	-6.686	7.015	5.319	8.544	Fail	0.962	6.494	6.544	3.103	7.198	Fail
		InceptionTime	0.939	-8.553	8.985	6.270	10.605	Fail	0.949	0.980	3.205	4.072	4.188	Pass
		TCN	0.942	-7.410	7.861	6.171	9.643	Fail	0.971	**0.052**	2.369	3.167	9.643	Pass
	Ours	Conformer-Transformer	**0.978**	**-2.587**	**3.414**	**3.652**	**4.484**	Pass	**0.985**	0.654	**1.774**	**2.318**	**2.419**	Pass
	Ours	Swin Transformer	0.975	-2.589	3.526	3.925	4.707	Pass	**0.985**	0.821	1.843	2.353	2.496	Pass

TABLE IV
COMPARISON WITH THE BHS STANDARD ON TWO DATASETS

Dataset	Method	Model	SBP (%)				DBP (%)			
			≤ 5 mmHg	≤ 10 mmHg	≤ 15 mmHg	BHS-Grade	≤ 5 mmHg	≤ 10 mmHg	≤ 15 mmHg	BHS-Grade
UCI-BP	[13]	Visibility Graph+CNN+Ridge Regression	53.46	81.15	92.43	B	75.72	95.04	98.56	A
	[14]	CNN+Transformer	80.69	94.56	97.57	A	94.07	98.70	99.65	A
	[27]	RDAE	58.50	85.60	95.00	B	81.50	96.40	99.00	A
	[28]	LSTM-based Autoencoder	70.60	94.10	98.60	A	91.10	99.10	99.80	A
	Ours	Conformer-Transformer	**83.10**	**96.50**	**99.10**	A	**96.30**	**99.50**	**99.90**	A
	Ours	Swin Transformer	82.40	95.90	98.70	A	95.70	**99.50**	99.80	A
MIMIC-BP		CNN	39.10	75.70	94.40	D	29.10	88.60	99.60	D
		InceptionTime	26.60	61.40	85.60	D	80.90	97.20	99.60	A
		TCN	35.40	71.00	88.60	D	90.20	99.10	99.80	A
	Ours	Conformer-Transformer	**77.10**	**96.60**	**99.50**	A	**96.40**	**99.60**	**99.90**	A
	Ours	Swin Transformer	76.00	95.60	99.30	A	95.90	**99.60**	**99.90**	A

TABLE V
PERFORMANCE COMPARISON ON THE MIMIC-BP DATASET BEFORE AND AFTER SIGNAL PREPROCESSING

Method	Model	SBP (mmHg)		DBP (mmHg)	
		MAE	RMSE	MAE	RMSE
Raw signal	Conformer-Transformer	4.399	5.722	1.909	2.542
Denoised	Conformer-Transformer	3.414↓	4.484↓	1.774↓	2.419↓
Raw signal	Swin Transformer	4.471	5.865	1.940	2.548
Denoised	Swin Transformer	3.526↓	4.707↓	1.843↓	2.496↓

TABLE VI
COMPARISON OF PREDICTED ABP WAVEFORM

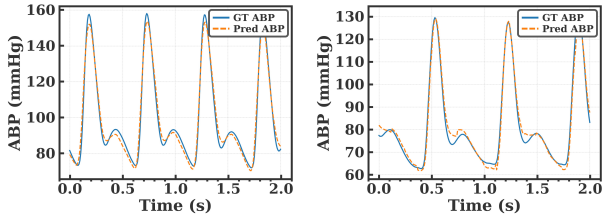
Dataset	Method	Model	MAE (mmHg)	SD (mmHg)	RMSE (mmHg)	*R*
UCI-BP		CNN	4.919	7.571	6.554	0.955
		LSTM	4.884	7.800	6.801	0.952
		Transformer	4.395	7.099	5.904	0.960
	Ours	Conformer-Transformer	**3.005**	**5.356**	**4.303**	**0.978**
	Ours	Swin Transformer	3.258	5.706	4.613	0.975
MIMIC-BP		CNN	5.084	7.309	6.608	0.942
		InceptionTime	6.194	8.509	7.997	0.924
		TCN	5.478	7.768	7.631	0.936
	Ours	Conformer-Transformer	**3.349**	**5.497**	**4.588**	**0.968**
	Ours	Swin Transformer	3.497	5.676	4.773	0.965

outperform baseline models on both the UCI-BP and MIMIC-BP datasets. Notably, the Conformer architecture demonstrates superior performance in segments with pronounced systolic pressure variations, owing to its capacity for modeling long-range temporal dependencies. Since most of the other published works using the same dataset did not perform ABP reconstruction [13], [29], [30], we decided to use serveral representative models as comparison.

It is worth highlighting that the proposed approach not only reconstructs the full ABP waveform sequence but also derives SBP and DBP values directly from it. This stands in contrast to traditional models that regress these values independently, and offers enhanced clinical interpretability along with improved support for multi-task learning.

E. Model Characteristics and Performance Differences

1) Architectural Differences and Performance Characteristics: Based on cross-validation, the Conformer-Transformer achieves superior SBP prediction and ABP waveform reconstruction, highlighting its strength in modeling long-term rhythmic dependencies by combining convolution for local morphology with attention for global temporal structure. The Swin Transformer performs comparably in DBP prediction



(a) Conformer-Transformer (b) Swin Transformer

Fig. 9. Comparison between the predicted and ground truth ABP waveforms on the UCI-BP dataset.

while offering faster inference and lower complexity, making it suitable for resource-constrained environments and more robust to local perturbations. Overall, the two models show complementary strengths—Conformer-Transformer toward rhythm modeling and Swin toward local feature perception—providing a foundation for future hybrid designs targeting different tasks such as long-term forecasting or real-time tracking.

2) Model Complexity: We evaluated the training and inference efficiency of the proposed models on an NVIDIA GeForce RTX 4070 Ti SUPER, 16 GB VRAM. For a fair comparison, we included two representative works [18], [31] and evaluated their inference efficiency on their reported hardware platforms. The results are summarized in Table VII.

In terms of architectural differences, PP-Net is based on an LRCN, focusing on temporal modeling but with limited capacity to capture complex morphological features. IMCA-PPG leverages ResNet-50 combined with multi-head cross-modal attention to enhance feature interactions, but at the cost of significant computational overhead. By contrast, our two models strike a better balance between efficiency and accuracy: the Conformer benefits from rhythm–morphology fusion, whereas the Swin Transformer exploits multi-scale temporal representation. Overall, the Swin Transformer is more suitable for real-time deployment scenarios (e.g., wearable devices), whereas the Conformer-Transformer is more appropriate for tasks requiring higher precision and robustness.

3) Model Loss Curves: The training and validation loss curves for both Conformer-Transformer and Swin Transformer models on the MIMIC-BP dataset are shown in Fig. 10. Both models exhibit a sharp decline in loss during the initial epochs, followed by a stable convergence, reflecting effective optimization. The validation curves remain slightly above the training curves, suggesting good generalization without overfitting. Notably, the Conformer-Transformer converges faster and reaches a lower final validation loss, while the Swin Transformer converges more gradually, highlighting their trade-off between training efficiency and robustness. It should be noted that the loss values are computed directly from the raw model outputs, rather than from the rescaled blood pressure predictions.

4) Interpretability through Rhythmicity and Morphological Feature Learning: We evaluate interpretability using the Landmark Overlap Score (LOS), which quantifies whether salient points fall on physiologically meaningful landmarks in the PPG waveform. Saliency sequences were derived using Integrated Gradients (IG) and Gradient-weighted Class Activation Mapping (Grad-CAM), and the top 20% points were selected to form Ω_s. Three landmarks (foot onset, systolic peak, dicrotic notch) were detected on the waveform, with ±100 ms windows assigned as Ω_L. The LOS is defined as $LOS = \frac{|\Omega_s \cap \Omega_L|}{|\Omega_s|}$.

Under random attribution, salient points are expected to be uniformly distributed, with the expected LOS corresponding to the proportion of landmark windows. A 2-second segment at 125 Hz contains approximately 250 samples; each ±100 ms window spans about 25 points, and three landmarks together cover around 75 points (≈ 30%). Allowing for potential window overlap, the effective coverage is estimated at 25–30%, which defines the random baseline of the LOS.

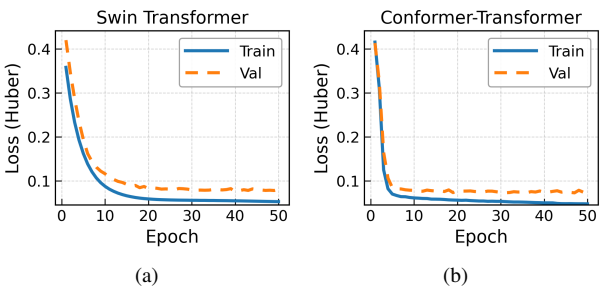
In experiments, the Conformer-Transformer achieved IG=0.329 and Grad-CAM=0.388, while the Swin Transformer achieved IG=0.395 and Grad-CAM=0.382, all above the baseline, as shown in Fig. 11. This confirms that both models attend to physiologically meaningful landmarks rather than arbitrary positions.

To assess whether the models rely on rhythmic continuity across segments, we designed a rhythm perturbation experiment. Specifically, while preserving the within-segment morphology and the original temporal order of target labels, we disrupted only the cross-segment rhythmic continuity on the input side by randomly shuffling signal segments. The results showed a significant degradation in BP estimation accuracy after shuffling, as shown in Table VIII, indicating that the models indeed depend strongly on rhythmic context rather than relying solely on isolated segment morphology. Here, DCRC refers to "Disrupting Cross-segment Rhythmic Continuity."

V. Conclusion

We proposed a dual-path deep learning framework for cuffless BP estimation using only PPG signals. By reconstructing ABP waveforms and predicting personalized SBP/DBP, our method enables interpretable and real-time monitoring. We successfully adapted Conformer-Transformer and 1D Swin Transformer—originally developed for speech and image recognition—to physiological signals. The Conformer-Transformer excels at modeling long-range rhythm and waveform structure, whereas the Swin Transformer offers robustness and low-latency inference suitable for wearable use.

On two public datasets, the Conformer-Transformer achieved the lowest MAE (2.979 mmHg systolic, 1.603 mmHg diastolic), improving on prior studies by 9.6% and 8.4% and delivering the best waveform reconstruction. The Swin Transformer obtained 3.034 mmHg systolic and 1.714 mmHg diastolic MAE with an average latency of 0.13 ms/sample, outperform-



(a) (b)

Fig. 10. Learning curves on the MIMIC-BP dataset.

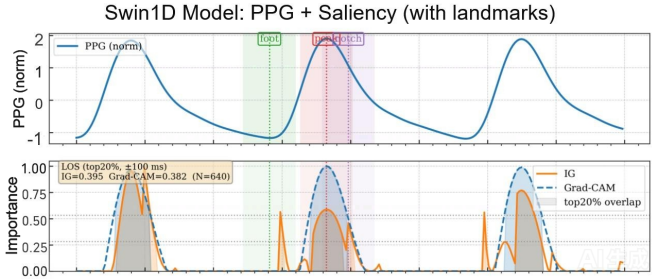


Fig. 11. Morphological saliency analysis of PPG with landmarks using the Swin Transformer on the MIMIC-BP dataset.

TABLE VII
MODEL COMPLEXITY AND INFERENCE EFFICIENCY ACROSS PLATFORMS

Model	Platform	Architecture	Model Output	Inference Time (ms/sample)	Sample Length	Inference Time / Unit (ms/s)
PP-Net [18]	NVIDIA Quadro P4000 (8 GB)	LRCN	SBP, DBP, HR	1.00	8 s	0.125
IMCA-PPG [31]	NVIDIA A40 (48 GB)	ResNet-50 + MHCA	SBP, DBP	30.85	15 s	2.056
Conformer-Transformer (Our)	NVIDIA GeForce RTX 4070 Ti SUPER (16 GB)	Conformer	SBP, DBP, ABP waveform	0.47	2 s	0.235
Swin Transformer (Our)	NVIDIA GeForce RTX 4070 Ti SUPER (16 GB)	1D Swin Transformer	SBP, DBP, ABP waveform	0.13	2 s	0.065

TABLE VIII
PERFORMANCE ON MIMIC-BP UNDER TWO SIGNAL CONDITIONS (FILTERED VS. DCRC)

Model	Signal	SBP (mmHg)		DBP (mmHg)	
	Condition	MAE	RMSE	MAE	RMSE
Conformer-Transformer	Original	3.414	4.484	1.774	2.419
	DCRC	7.941	10.369	3.952	5.263
Swin Transformer	Original	3.526	4.707	1.843	2.496
	DCRC	7.852	10.349	4.005	5.318

ing existing methods. All results meet BHS grade A and AAMI standards.

These findings show that BP estimation is not merely regression but a joint challenge of temporal modeling and morphological analysis. The complementary strengths of both models suggest potential for hybrid or adaptive approaches. Future work will explore architectural integration, multimodal fusion, and clinical translation to wearables, as well as personalized strategies such as subject-specific adaptation, federated learning, and physiology-aware fine-tuning.

ACKNOWLEDGMENT

We gratefully acknowledge Hector Stiftung for partially funding of this work.

REFERENCES

[1] A. K. Keates *et al.* Cardiovascular disease in africa: epidemiological profile and challenges. *Nature Reviews Cardiology*, 14(5):273–293, 2017.

[2] B. Zhou *et al.* Global epidemiology, health burden and effective interventions for elevated blood pressure and hypertension. *Nature Reviews Cardiology*, 18(11):785–802, 2021.

[3] T. Panula *et al.* Advances in non-invasive blood pressure measurement techniques. *IEEE Reviews in Biomedical Engineering*, 16:424–438, 2022.

[4] G. Parati *et al.* Home blood pressure monitoring: methodology, clinical relevance and practical application: a 2021 position paper by the working group on blood pressure monitoring and cardiovascular variability of the european society of hypertension. *Journal of Hypertension*, 39(9):1742–1767, 2021.

[5] R. Mieloszyk *et al.* A comparison of wearable tonometry, photoplethysmography, and electrocardiography for cuffless measurement of blood pressure in an ambulatory setting. *IEEE Journal of Biomedical and Health Informatics*, 26(7):2864–2875, 2022.

[6] A. B. Sheikh *et al.* Blood pressure variability in clinical practice: past, present and the future. *Journal of the American Heart Association*, 12(9):e029297, 2023.

[7] H. Kim *et al.* A configurable and low-power mixed signal soc for portable ecg monitoring applications. *IEEE Transactions on Biomedical Circuits and Systems*, 8(2):257–267, 2013.

[8] D. Castaneda *et al.* A review on wearable photoplethysmography sensors and their potential future applications in health care. *International Journal of Biosensors & Bioelectronics*, 4(4):195, 2018.

[9] C. El-Hajj and P. A. Kyriacou. A review of machine learning techniques in photoplethysmography for the non-invasive cuff-less measurement of blood pressure. *Biomedical Signal Processing and Control*, 58:101870, 2020.

[10] C. El-Hajj and P. A. Kyriacou. Cuffless blood pressure estimation from ppg signals and its derivatives using deep learning models. *Biomedical Signal Processing and Control*, 70:102984, 2021.

[11] N. Hasanzadeh *et al.* Blood pressure estimation using photoplethysmogram signal and its morphological features. *IEEE Sensors Journal*, 20(8):4300–4310, 2019.

[12] S. Haddad *et al.* Continuous ppg-based blood pressure monitoring using multi-linear regression. *IEEE Journal of Biomedical and Health Informatics*, 26(5):2096–2105, 2021.

[13] W. Wang *et al.* Cuff-less blood pressure estimation from photoplethysmography via visibility graph and transfer learning. *IEEE Journal of Biomedical and Health Informatics*, 26(5):2075–2085, 2021.

[14] D.-K. Kim *et al.* Deepcnap: A deep learning approach for continuous noninvasive arterial blood pressure monitoring using photoplethysmography. *IEEE Journal of Biomedical and Health Informatics*, 26(8):3697–3707, 2022.

[15] Z. Qiu *et al.* Joint regression network and window function-based piecewise neural network for cuffless continuous blood pressure estimation only using single photoplethesmogram. *IEEE Transactions on Consumer Electronics*, 68(3):236–260, 2022.

[16] C. Ma *et al.* Kd-informer: A cuff-less continuous blood pressure waveform estimation approach based on single photoplethysmography. *IEEE Journal of Biomedical and Health Informatics*, 27(5):2219–2230, 2022.

[17] J. Leitner *et al.* Personalized blood pressure estimation using photoplethysmography: A transfer learning approach. *IEEE Journal of Biomedical and Health Informatics*, 26(1):218–228, 2021.

[18] M. Panwar *et al.* Pp-net: A deep learning framework for ppg-based blood pressure and heart rate estimation. *IEEE Sensors Journal*, 20(17):10000–10011, 2020.

[19] M. Kachuee *et al.* Cuff-less high-accuracy calibration-free blood pressure estimation using pulse transit time. In *2015 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1006–1009. IEEE, 2015.

[20] I. Sanches *et al.* Mimic-bp: A curated dataset for blood pressure estimation. 11(1):1233, 2024.

[21] C. Ma *et al.* Diffcnbp: Lightweight diffusion model for iomt-based continuous cuffless blood pressure waveform monitoring using ppg. *IEEE Internet of Things Journal*, 2024.

[22] H. Mohammed *et al.* Meta-analysis of pulse transition features in non-invasive blood pressure estimation systems: Bridging physiology and engineering perspectives. *IEEE Transactions on Biomedical Circuits and Systems*, 17(6):1257–1281, 2023.

[23] A. Gulati *et al.* Conformer: Convolution-augmented transformer for speech recognition. *arXiv preprint arXiv:2005.08100*, 2020.

[24] Z. Liu *et al.* Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 10012–10022, 2021.

[25] ANSI/AAMI/ISO. Non-invasive sphygmomanometers – part 2: Clinical investigation of automated measurement type. Standard ANSI/AAMI/ISO 81060-2:2013, Webstore.ansi.org (accessed 6 September 2016), 2013.

[26] E. O'Brien *et al.* The british hypertension society protocol for the evaluation of automated and semi-automated blood pressure measuring devices with special reference to ambulatory systems. *Journal of Hypertension*, 8(7):607–619, 1990.

[27] K. Qin *et al.* Deep generative model with domain adversarial training for predicting arterial blood pressure waveform from photoplethysmogram signal. *Biomedical Signal Processing and Control*, 70:102972, 2021.

[28] L. N. Harfiya *et al.* Continuous blood pressure estimation using exclusively photopletysmography by lstm-based signal-to-signal translation. *Sensors*, 21(9):2952, 2021.

[29] P. Guo *et al.* Cuffless blood pressure estimation model based on prior information of physiological data and lstm. *IEEE Sensors Journal*, 2024.

[30] D. Wang *et al.* Photoplethysmography-based blood pressure estimation combining filter-wrapper collaborated feature selection with lasso-lstm model. *IEEE Transactions on Instrumentation and Measurement*, 70:1–14, 2021.

[31] V. S. Roha *et al.* Evolving blood pressure estimation: From feature analysis to image-based deep learning models. *Journal of Medical Systems*, 49(1):97, 2025.