

TransfoRhythm: A Transformer Architecture Conductive to Blood Pressure Estimation via Solo PPG Signal Capturing

Amir Arjomand¹, Amin Boudesh², Farnoush Bayatmakou², Kenneth B. Kent^{1,*}, and Arash Mohammadi^{2,*}

¹Department of Computer Science, University of New Brunswick, Fredericton, Canada

²Concordia Institute for Information Systems Engineering, Concordia University, Montreal, QC, Canada

ABSTRACT

Recent statistics indicate that approximately 1.3 billion individuals worldwide suffer from hypertension, a leading cause of premature death globally. Blood pressure (BP) serves as a critical health indicator for accurate and timely diagnosis and/or treatment of hypertension. Traditional BP measurement methods rely on cuff-based approaches, which lack real-time, continuous, and reliable BP estimates, crucial for timely diagnosis/treatment of hypertension. Driven by recent advancements in Artificial Intelligence (AI) and Deep Neural Networks (DNNs), there has been a surge of interest in developing data-driven and cuff-less BP estimation solutions. In this context, current literature predominantly focuses on coupling Electrocardiography (ECG) and Photoplethysmography (PPG) sensors, though this approach is constrained by reliance on multiple sensor types. An alternative, utilizing standalone PPG signals, presents challenges due to the absence of auxiliary sensors (ECG), requiring the use of morphological features while addressing motion artifacts and high-frequency noise. To address these issues, the paper introduces the TransfoRhythm framework, a Transformer-based DNN architecture built upon the recently released physiological database, MIMIC-IV. Leveraging Multi-Head Attention (MHA) mechanism, TransfoRhythm identifies dependencies and similarities across data segments, forming a robust framework for cuff-less BP estimation solely using PPG signals. To our knowledge, this paper represents the first study to apply the MIMIC IV dataset for cuff-less BP estimation, and TransfoRhythm is the first MHA-based model trained via MIMIC IV for BP prediction. Performance evaluation through comprehensive experiments demonstrates TransfoRhythm's superiority over its state-of-the-art counterparts. Specifically, TransfoRhythm achieves highly accurate results with Root Mean Square Error (RMSE) of [1.84, 1.42] and Mean Absolute Error (MAE) of [1.50, 1.17] for systolic and diastolic blood pressures, respectively.

1 Introduction

High blood pressure is a serious symptom of hypertensive disease that can lead to disability and even death¹. Based on an American Heart Association estimation, about 41.4% of US people will be struggling with hypertension by 2030². There are various approaches to measure Blood Pressure (BP), among which, cuff-based measurement³ (where a cuff is employed to temporarily stop the blood flow to obtain the extremum BP values) is the common traditional method used both in hospitals and for in-home care. Such an approach, however, carries the following drawbacks: (i) Cuff devices are not readily available in low-resource settings, leading to the lack of awareness and control of hypertensive conditions; (ii) Repeated cuff inflation and deflation can disrupt gold-standard hypertension diagnosis and hypotension surveillance, resulting in under-utilization and infrequent measurements, and; (iii) Widely used cuff-based devices do not provide continuous BP measurement, limiting immediate detection of hypotension and real-time therapy titration⁴. Alternative methods of BP measurement, such as cuff-less monitoring, have been explored to provide continuous monitoring and reduce the potential for measurement errors caused by cuff-based methods. Cuffless BP measurement can facilitate regular monitoring, enable long-term control, and provide precise evaluation and diagnosis of hypertension. Additionally, it can allow for hypotension management and therapy through seamless continuous monitoring⁴. Another advantage of continuous BP monitoring, which is provided by cuffless methods, is the potential to monitor nocturnal BP patterns. Nocturnal blood pressure can be a more informative risk factor compared to daytime blood pressure in predicting the incidence of Cardiovascular Disease (CVD). Both hypertensive and general populations are at risk of developing CVD if there is a significant surge or dip (more than 10 – 20%) in BP during nighttime, which may result in stroke, heart failure, and even death⁵. Cuffless approaches use indirect methods such as pulse wave velocity, pulse transit time, pulse wave analysis, volume clamping, and applanation tonometry to estimate blood pressure⁶.

The history of cuff-less BP estimation arises from arterial stiffness measurement through Pulse Wave Velocity (PWV)⁷⁻¹⁰. In 1981, Geddes and Voelz tried, for the first time, to analyze Pulse Transit Time (PTT) and BP on ten dogs and reported a

meaningful correlation between PTT and diastolic BP¹¹. Recent cuffless measurement techniques, typically, employ biological signals such as Photoplethysmography (PPG), Ballistocardiography (BCG), and Electrocardiography (ECG) to estimate BP. PPG is obtained based on the principle that the absorption of light by blood changes as it passes through the blood vessels. A PPG sensor placed on the skin measures the changes in light absorption caused by the pulsatile blood flow, which is then used to estimate BP. BCG is obtained based on the principle that the heart's contraction generates a force that is transmitted through the body, causing a measurable movement. A BCG sensor placed on a person's seat measures the movements caused by the heart's contraction and can be used for BP estimation. ECG measures heart's electrical activities, which can be used to derive various parameters related to the heart's function, including BP¹². Capitalizing on recent advancements of Machine Learning (ML) models, there has been a surge of interest in ML-based cuff-less measurement of BP from ECG and PPG signals. Generally speaking, ML models, in particular, Deep Learning (DL) architectures, showed significant potentials in extracting discriminative features from ECG/PPG signals, estimating volumetric change of blood flow, and predicting relative systolic and diastolic BP, referred to as SBP and DBP. For instance, Baker *et. al*¹³ developed an integrated Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) model, fed by PPG and ECG signals. The model was trained based on data from 84 individuals. The reported Mean Absolute Error (MAE) values are 4.41 and 2.91 for the SBP and DBP. Another relevant study¹ developed a Temporal Convolutional Network (TCN) architecture using multi-lead ECG and PPG sensors. The model is trained based on data of 293 individuals obtained from the MIMIC [I, III] datasets, and showed compelling results. The Root Mean Square Error (RMSE) associated with the SBP is reported as 3.03, while that of the DBP is reported as 1.58. Huang *et. al*¹⁴ proposed a modified architecture of MLP-Mixer¹⁵ with a novel pre-processing technique (referred to as the MFMC). The resulting SBP-RMSE is 5.10 and DBP-RMSE is 3.13 on 3,000 records of the MIMIC II¹⁶ dataset.

Although BP estimation via PPG and ECG sensors is promising, it suffers from some critical drawbacks such as reliance on two types of sensors, high power consumption, need for sensor re-calibration, and dependence on keeping a fixed distance between sensors. These issues can be handled by relying on stand-alone PPG signals captured from the skin surface. The PPG technique is an optical solution designed to measure volumetric variations of blood circulation and can provide invaluable information about the underlying cardiovascular system. Its principal mechanism is based on the emission and reflection of light in the near-infrared wavelength (800nm ~ 2500nm) from arteries' blood flow. Due to the easement use of PPG sensors and its potential for BP estimation, it has been extensively utilized for development of advanced processing/learning models¹⁷⁻²⁰. In this context, when it comes to feature extraction from PPG signals, two types of data representations are commonly used (as the appropriate feed for downstream processing/learning modules), i.e., raw PPG signals (time series input)^{13,14,21,22}, and feature-based vectors²³⁻²⁷. Such representations can be extracted from the original PPG signal, multiple-order signal derivative, and Pulse Wave Decomposition (PWD) of the PPG. In the modeling step, to achieve accurate BP estimation, recent studies have employed various methodologies ranging from traditional ML models, such as Support Vector Regression (SVR) and Random Forest²⁷⁻²⁹, to Deep Neural Network (DNN) architectures^{1,30} such as LSTM networks, and Gated Recurrent Unit (GRU) models with attention mechanisms^{22,26,31}. Despite these studies, standalone PPG-based BP estimation is still a critically challenging task as by removing the auxiliary sensor (ECG), one needs to particularly consider morphological features, and pay attention to vulnerability to motion artifacts and high-frequency noise³².

To address the above mentioned challenges, the paper introduces the TransfoRhythm framework, an attention-based DNN architecture built upon the recently released physiological database, MIMIC-IV dataset³³. To our knowledge, the paper represents the first study to apply the MIMIC IV dataset for BP prediction. At the core of the TransfoRhythm framework is a Multi-Head Attention (MHA) mechanism, making it the first MHA-based DNN model trained via MIMIC IV for the application at hand. More specifically, to the best of our knowledge, only handful of recent studies^{34,35,37} tapped into potentials of attention-mechanism for BP estimation via PPG signals. In brief, Reference³⁵ developed a MHA model using the UCL dataset³⁶. This study implemented two different models, a transformer-based architecture, and a Frequency-Domain Linear/Non-linear Regression model (FD-L/NL), with the latter providing better performance. Similarly, Reference³⁷ proposed a transformer-based model referred to as AriaNet, while authors in Reference³⁴ focused on self-supervised transfer learning based on transformers. Different from these recent studies, the TransfoRhythm is built on the regressive time series transformer network, which is crafted to learn from time series data while preserving sequence order and allowing parallel computation during training and inference. To expand the feature vectors' dimensionality, an embedding feature, similar to those used in Natural Language Processing (NLP) tasks, is employed. Positional encoding is then added to the extracted features to create the input data matrix, which is fed into a model block via the MHA mechanism. The output is refined in a position-wise block, which is then processed by time compressor and flattening blocks to shorten the time frame and calculate estimated systolic and diastolic BP values. The obtained results validate the effectiveness of such a modeling framework and highlight its potential for clinical implementation.

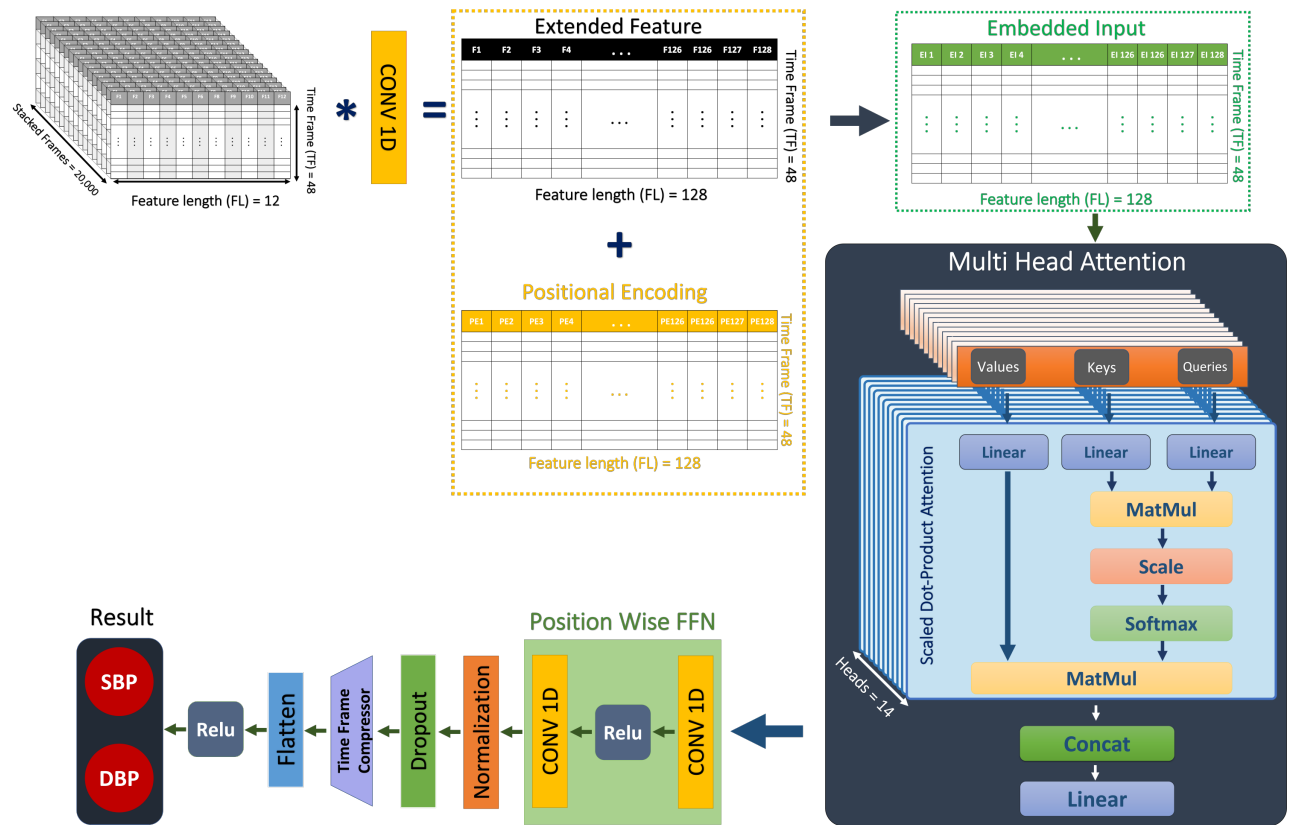


Figure 1. This block diagram depicts the streamlined process for estimating systolic and diastolic blood pressures. Features are extracted from the dataset (top left), combined with positional encoding, and form an embedded input matrix. This matrix is then inputted into the model block, which integrates a multi-head attention mechanism. The output of the model is further refined in the position-wise block. Finally, the Time compressor and flattening blocks are applied to reduce the time frame and yield the estimated systolic and diastolic blood pressure values.

2 Methodology

This section focuses on material and methodology, whereas in the first part (the dataset formation subsection) we describe the preprocessing and feature extraction approaches. In the second part, the network architecture subsection provides an overview of the utilized modeling process. In brief and according to the workflow of our study, shown in Fig. 1, after dataset formation and preprocessing, the overall dataset is compiled based on 12 final features. Next, the positional encoding values are added to the flattened values of the dataset, resulting in the construction of the embedded input matrix, which is used to feed the model. In the model part, a multi-head attention algorithm is employed. The position-wise process is applied to the output of the multi-head attention block and finally, by reducing the time frame and flattening the position-wise output, the values related to systolic and diastolic blood pressures are obtained.

2.1 Dataset Formation

The Medical Information Mart for Intensive Care (MIMIC)-IV Waveform Dataset is employed, which is available on the Physio-Net repository³⁸ to analyze the version of waveform recordings from bedside monitoring devices in modern Intensive Care Units (ICUs). MIMIC-IV includes ECG, PPG, and invasive ABP signals and is the latest and most recently released Dataset of its kind following a joint effort from Beth Israel Deaconess Medical Center (BIDMC) and Massachusetts Institute of Technology (MIT). The recording process was conducted at a sampling rate of 62.4 Hz. To ensure patient confidentiality, all 198 individuals in the dataset were completely de-identified. Fig. 2 illustrates the duration of the recorded signals and their corresponding iterations. The majority of patients were monitored for less than 1.5 hours, while some patients had multiple recordings lasting over 15 hours. Although previous studies have employed MIMIC-II and MIMIC-III datasets for blood pressure estimation, to the best of our knowledge, this study stands out as the first utilization of the released MIMIC-IV v2.0 dataset for this purpose.

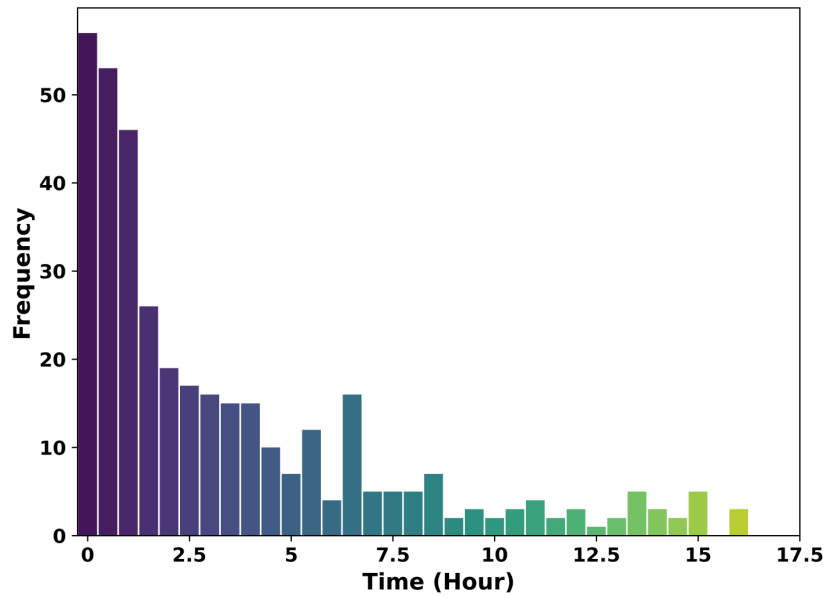


Figure 2. Time distribution of records depicted in the graph. The x – axis represents the length of records, ranging from a few minutes to a maximum of 17 hours. The darker shades of blue indicate a higher number of observations. The majority of patients have records with a time duration of less than 1.5 hours.

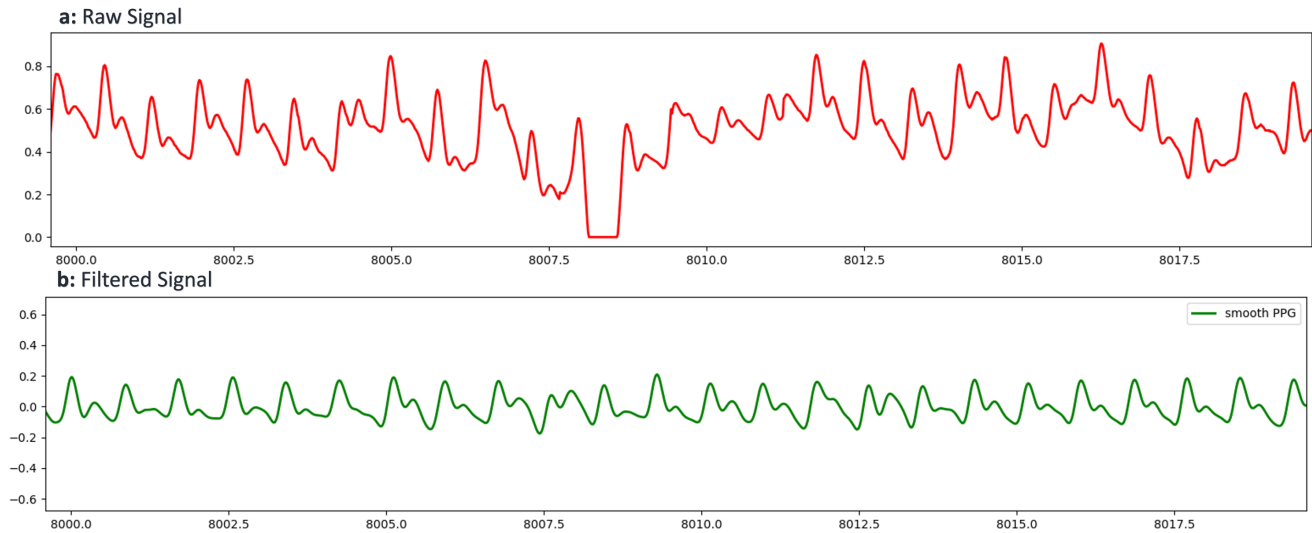


Figure 3. A comparison of the signal before and after preprocessing. The photo demonstrates the exceptional noise filtering achieved, resulting in a clean and smooth signal. Furthermore, the peaks and feet are accurately preserved, showcasing the effectiveness of the preprocessing techniques employed.

2.1.1 Preprocessing

The preprocessing stage is a crucial step to enhance the quality of the underlying signals and remove potential noise and/or artifact sources that may have been introduced during data acquisition. In this study, in the preprocessing phase, data cleaning and signal filtering were performed. In the data cleaning step, signals with frequent outrange amplitudes and records that did not meet the minimum time duration (less than 15 minutes) as well as certain frames of long records that contained disqualified signals, such as flattened-line or extremely noisy segments were detected, and removed from the records. After the data cleaning

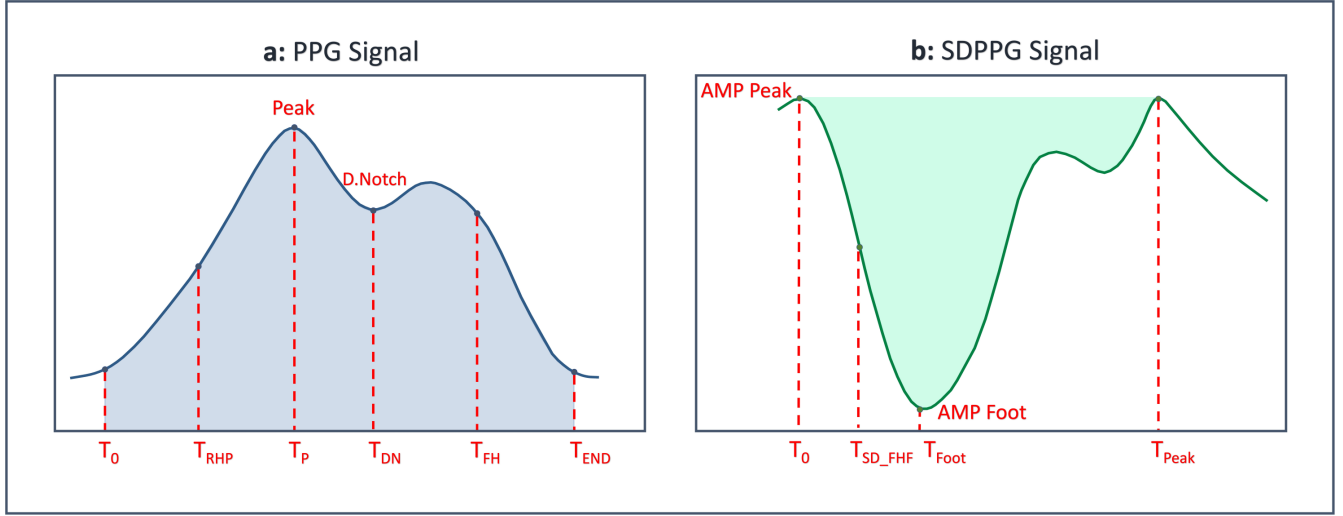


Figure 4. Characterization of features on the signal. In the PPG signal, the peak is defined as the maximum value of a PPG wave, while d.notch is defined as the minimum value between two consecutive peaks. In the SD PPG signal, amp peak refers to the maximum amplitude, while amp foot refers to the minimum amplitude between two consecutive waves.

step, a total of 360 purified records were identified. In the second step of the preprocessing phase, i.e., the signal filtering step, a 5th order Butterworth filter in the frequency range of 0.7 to 10 Hz and a 5th order Moving Average Filter (MAF) were applied to remove overshoots and fluctuations. More specifically, the Butterworth Infinite Impulse Response (IIR) bandpass filter was employed to remove wandering baseline and high-frequency noise, such as power-line interference (50 60 Hz). The MAF involves taking the average of a subset of data points and using it to smooth out the underlying signals. The combination of the Butterworth filter and MAF resulted in a highly filtered and smooth PPG signal. Some statistics associated with the result of the preprocessing phase are shown in Fig. 3.

2.1.2 Feature Extraction

The first step in the feature extraction phase involves identifying and separating the cycles of each frame to obtain meaningful data. In this regard, we employ a robust algorithm based on a one-dimensional array that locates peak and foot (valley) points of the cycles in the PPG and the corresponding second derivative of the PPG (SDPPG) signal inside each frame. This algorithm is parameterized by an adjustable peak-to-peak distance, which allows the detection of varying heights and amplitudes of the peaks. Subsequently, according to these identified foot and peak points, PPG and its corresponding SDPPG cycles were extracted. Additionally, before starting feature extraction, each frame was evaluated to be used or to be skipped based on an adaptive threshold for peak detection, which sets the threshold value based on the average amplitude of the signal. This approach ensures that the algorithm is sensitive enough to detect peaks of varying heights and amplitudes, while also being robust to noise and other signal distortions. Based on the correlation matrix between the features, a set of 12 features was selected. The definition and description of these incorporated features are shown in Table 1, furthermore, visual illustrations of PPG and SDPPG signals representing selected features' specifications are shown in Fig. 4. Furthermore, to address potential bias and ensure a well-balanced dataset for detecting hypertension and hypotension events, a dynamic random frame generator is utilized. This generator automatically increases the number of observations with systolic blood pressure (SBP) exceeding 140 mmHg or falling below 80 mmHg. This step is crucial in achieving dataset balance and preventing overfitting of the model.

2.2 Network Architecture

The proposed TransfoRhythm framework is developed based on the regressive time series transformer network, which is designed to learn from time series data while maintaining sequence order and enabling parallel computation during the training and inference phases. To enhance the dimensionality of the feature vectors, an embedding feature, previously implemented in Natural Language Processing (NLP) tasks³⁹ is utilized. The initial structure of the features is defined by Stacked Frames (SF) based on shuffled frames, with 20,000 frames (N_i) and a time frame sequence of 48, which is given by

$$SF = [F_T^{L,1}, F_T^{L,2}, \dots, F_T^{L,n}], \quad (1)$$

where n refers to the number of frames, $F_T^{L,i}$ represents an individual frame that spans a shuffled frame with the input feature length of L ($L_{in} = 12$). Furthermore, T refers to the time frame sequence and is equal to 48. To extract the extended features, a

Table 1. Definition of the selected features.

Feature	Symbol	Description
PPG_Cycle_Duration	TD1 = Tend - Tstart	Total Duration
PPG_Rise_Half_Peak	Trhp	The duration between start to the half peak amplitude (rise mode)
PPG_Peak To_ Notch	TD2 = Tdn-Tp	The duration between peak to dicrotic notch
PPG_Rise_Peak	Tp	The duration between the start to the peak
PPG_Fall_Half	Tfh	The duration between peak to half of its amplitude (fall mode)
PPG_Fall_Peak	TD3 = Tend - Tp	The duration between peak to foot (end)
PPG_BPM_Frame	PBF	Cycles number per frame
PPG_Integration	PPGI = (Green Area)	Under Curve(cycle) Area
SDPPG_Fall_Half_Foot	SDPPG_Tfhf	The duration between the start to the half of its amplitude (fall mode)
SDPPG_Extermum_Amp	SDPPG_AMP	SDPPG Maximum Amplitude
SDPPG_Integration	SDPPGI = (Blue Area)	SDPPG Upper Curve Area
SDPPG_Foot_Peak	TD4 = Tpeak - Tfoot	SDPPG Duration between Foot to Peak

1D convolution layer is applied to the stacked frames as follows

$$ST_{Extended}(N_i, L_{out}) = Bias(L_{out}) + \sum_{k=0}^{L_{in}} weight(L_{out}, k) input(N_i, k), \quad (2)$$

where $ST_{Extended} \in \mathbb{R}^{N_i \times L_{out} \times T}$, $weight \in \mathbb{R}^{L_{out} \times L_{in} \times KS}$, and $input \in \mathbb{R}^{N_i \times L_{in} \times T}$. Eq. 2 expands the feature length from 12 (L_{in}) to 128 (L_{out}) with a kernel size (KS) of 1.

2.2.1 Multi-Head Attention Mechanism

The self-attention mechanism is employed to capture the sequence dependencies and features carried on each frame. This mechanism employs three main matrices, i.e., Key, Queue, and Value, which are obtained by multiplying the input embedding matrix with learnable weights W_K , W_Q , and W_V , respectively. The resulting matrices are then used to compute a set of attention scores. To capture additional information about in-frame sequence dependency, multiple learnable sets (Heads) are stacked and concatenated, in the current study, we utilize a multi-head attention mechanism with 14 heads for BP estimation. Additionally, the model incorporates a linear layer in the last section of multi-head attention to introduce non-linearity. This approach enables the model to capture complex relationships and dependencies among input features, resulting in better BP predictions.

Despite the self-attention mechanism being a key component of the transformer architecture, it does not incorporate information about the relative positions of tokens in the input sequence. To address this limitation, a technique called Positional Encoding (PE) was introduced that encodes the position of each element in the input sequence by adding a sinusoidal function of different frequencies to the input embedding³². The PE array is added, in an element-wise fashion, to the $ST_{Extended}$ array, resulting in an Embedded input with a feature length of 48 and a time frame sequence length of 128. The incorporated approach enables the transformer architecture to capture both the content and positional information of the input sequence.

Position-Wise Feed-Forward Neural Network: The next component of the attention module is the position-wise feed-forward layer that employs a 1D convolution layer with a ReLU activation function to help the model learn better representations for each position in the input sequence. This layer enables the model to capture context-specific information and understand the relationship between different positions in the sequence⁴⁰. Furthermore, generalization techniques such as the normalization layer and dropouts are implemented to improve the model's performance and avoid overfitting.

Time Frame Compressor and Flattening: In the initial stages of experimentation, we encountered computational challenges related to the extensive computational overhead on local hardware. We identified that the final flattening layer, responsible for converting the high-dimensional output into a one-dimensional structure, posed a bottleneck. To overcome this, we employed a Time Frame Compressor unit that effectively reduces dimensionality while preserving temporal signals. Specifically, this non-learnable transformer shortens the temporal order into a more condensed representation. Thereafter, in the last stage of the model, we employ a flattened layer with non-learnable weights to transform the temporally compressed signals to systolic and diastolic blood pressure values through a ReLU function.

Table 2. Values of the considered hyperparameters

Hyperparameters	Type and Value
Dropout	0.15
Optimization Algorithm	Adam
Learning Rate	1E-4 (with dynamic wight decay)
Activation Function	ReLU
Loss Function	MSE
Batch Size	128
Epoch	400

2.3 Evaluation Metrics

To evaluate the effectiveness of the proposed TransfoRhythm framework, we have identified three key evaluation metrics, i.e., R-squared (R^2), Mean Absolute Error (MAE), and Root Mean Square Error ($RMSE$), given by

$$R^2 = \frac{\sum_{i=1}^n (T_i - P)^2}{\sum_{i=1}^m (T_i - \bar{T})^2}, \quad (3)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |T_i - P_i|, \quad (4)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (T_i - P_i)^2}. \quad (5)$$

where T_i refers to the target and P_i is the predicted value. In brief, Metric R^2 serves as an indicator of the extent to which the variance in the dependent variable can be attributed to the independent variables. The other two metrics (MAE and $RMSE$) quantify the magnitude and/or percentage of errors between actual and predicted values.

3 Results

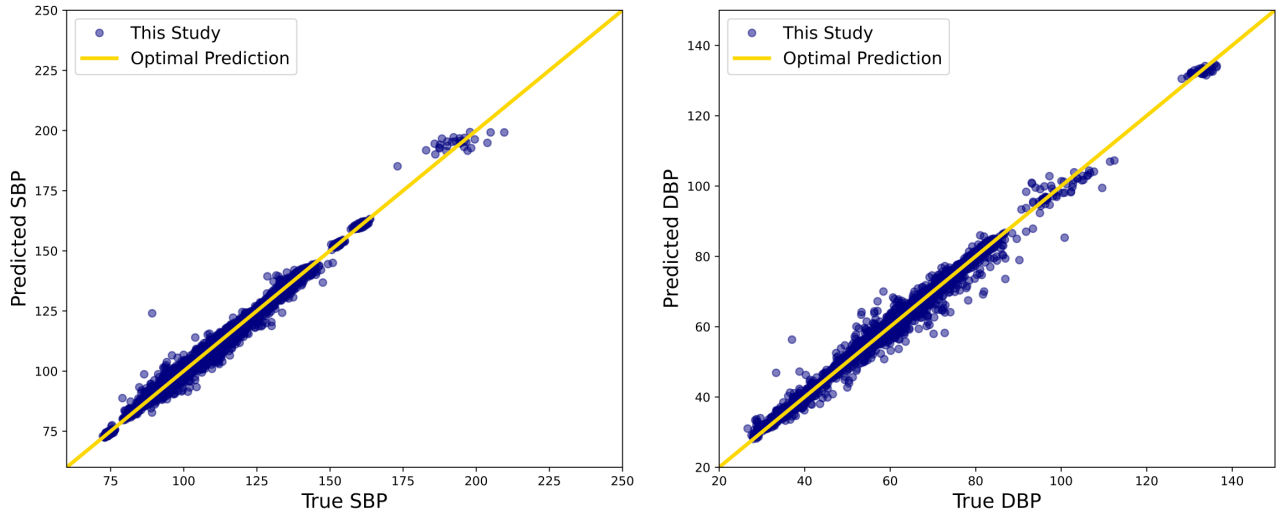
To evaluate different aspects and potentials of the proposed TransfoRhythm framework, we conducted a series of experiments involving various hyperparameters, as well as modifications to the model architecture. The final set of hyperparameters used in our experiments is presented in Table 2. More specifically, to evaluate performance of the proposed model, we employed a 5-fold cross-validation approach. This involved dividing the data into five subsets, with 80% allocated for training and 20% for the validation dataset. The utilization of the cross-validation method provided validation on the entire dataset and effectively mitigated the risk of overfitting, ensuring the model's generalizability and reliability. The proposed TransfoRhythm framework was subjected to 500 epochs of training for each fold on a high-end local server equipped with an Intel 12700 CPU, 16 gigabytes of RAM, and an Nvidia 3060 GPU with 12 gigabytes of VRAM. Although the final model was selected based on the best metric of the cross-validation process, the reported results were obtained by calculating the mean values of the performance metrics across the five folds. This highlights the model's capability to generalize its performance across the entire dataset. We obtained mean R^2 values of [0.991, 0.991], mean MAE values of [1.50, 1.17], and mean RMSE values of [1.84, 1.42] for the SBP and DBP, respectively. Further details can be found in Table 5.

3.1 Validation against AAMI and BHS Standards

To ensure that the proposed TransfoRhythm architecture meets the international standards for medical device regulation, we evaluated our findings using the standards of the Association for the Advancement of Medical Instrumentation (AAMI) and The British Hypertension Society (BHS). To comply with AAMI standards, certain statistical conditions must be met, including a sample size of at least 85 individuals aged 12 years or older, a Mean Error (ME) of less than 5 mmHg, and a Standard Deviation (SD) of the error less than 8 mmHg. Additionally, BHS grades accuracy based on the cumulative percentage of error. Table 3 provides a comparison between our results and these standard metrics. As presented in Table 3 the proposed model in this study has successfully met the AAMI standard. Moreover, the model has also achieved "Grade A" in the BHS standard. These results confirm the reliability and outstanding performance of the proposed model based on the prediction error values.

Table 3. Evaluation of the Proposed Model with AAMI and BHS Standards.

		Records	Mean Error (mmHg)	STD (mmHg)	Cumulative error percentage		
					<5 mmHg	<10 mmHg	<15 mmHg
TransfoRhythm Architecture	SBP	360	0.138	1.93	98%	100%	100%
	DBP	360	-0.166	1.58	98%	100%	100%
AAMI Standard Requirement		>85	<5	<8	-	-	-
BHS Standard Requirement	Grade A	-	-	-	60%	85%	95%
	Grade B	-	-	-	50%	75%	90%
	Grade C	-	-	-	40%	65%	85%

**Figure 5.** Comparison of true values and prediction. The blue dots represent the predicted values, while the gold line represents the optimal prediction. (a) shows the results for SBP, and; (b) shows the results for DBP.

3.2 Error Investigation

Comparison between the predicted and true values is an essential step in evaluating the performance of the TransfoRhythm model as it provides insight into how well the model can be generalized to unseen data and also helps to identify the functional range of the model. Fig. 5 represents a comparison of true and predicted SBP and DBP values. Fig. 5 effectively showcases the model’s ability to estimate these blood pressure measurements and provides a clear understanding of its predictive accuracy.

3.3 Statistical Assessment via Bland-Altman

The Bland-Altman diagram is a statistical method for assessing the agreement between two measurement methods by plotting their difference from their mean. In Fig. 6, we present the Bland-Altman plot for the TransfoRhythm model. The plot shows a mean difference of 0.12 mmHg with a 95% limit of agreement ranging from -3.21 to 3.46 mmHg for SBP and a mean difference of -0.31 mmHg with a 95% limit of agreement ranging from -5.37 to 4.74 mmHg for DBP.

3.4 Comparisons with Benchmark Models

This section focuses on validating our results by comparing the performance of our proposed model with benchmark algorithms from previous studies. Given the lack of prior work utilizing the MIMIC IV dataset for BP prediction, we implemented benchmark models from other studies that analyzed earlier versions of the MIMIC dataset. These benchmark models are employed to establish a baseline performance for blood pressure estimation in the context of the MIMIC IV dataset. The selection of benchmark models was based on specific criteria. Firstly, we considered models that exclusively utilized PPG signals. Furthermore, due to the impracticality of conducting extensive feature engineering for each model, we specifically chose models that were trained on raw data. Additionally, we selected models with open-access codes with satisfying results in their respective papers. The benchmark models enclosed various architecture families. From CNN-based networks, we

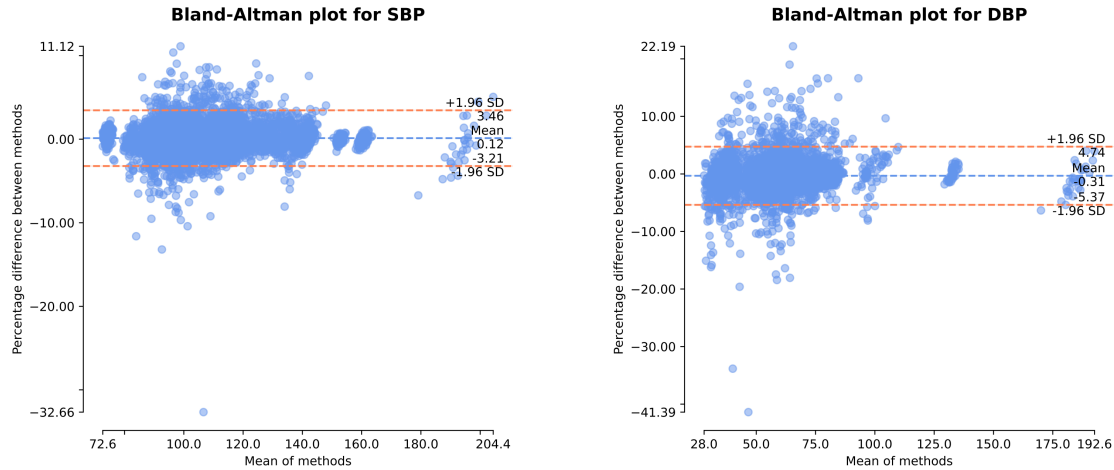


Figure 6. Bland–Altman plots for the SBP and DBP. The Bland–Altman plot illustrates the agreement between true and predicted values. The plot depicts the 95% statistical limits of agreement, corresponding to 1.96 standard deviations (SD) of the mean. The mean difference or bias observed is 0.12 and -0.31 for SBP and DBP, respectively.

Table 4. Comparison of the proposed model with benchmark models on Mimic IV dataset

Architecture Type	Model	MAE SBP	MAE DBP	RMSE SBP	RMSE DBP
CNN	ResNet_1D	2.28	1.06	5.55	3.21
	AlexNet_1D	3.09	1.59	7.98	4.60
	U-Net	2.76	1.56	7.47	4.70
RNN	Bi-LSTM	3.86	1.9	7.98	4.19
Hybrid	Hybrid CNN-RNN Network	2.79	1.96	5.62	3.91
MHA	TransfoRhythm	1.50	1.17	1.84	1.42

selected ResNet1D and AlexNet1D, which were originally implemented and evaluated on MIMIC II and MIMIC-III datasets⁴¹. Additionally, we included a U-Net model that had been evaluated on the MIMIC III dataset⁴². In terms of Recurrent Neural Networks (RNN), we defined a BiLSTM model for our benchmark set, which aimed to address the absence of suitable RNN-based models meeting our selection criteria. The BiLSTM model begins with a 1D Convolutional layer for feature extraction, followed by a Max pooling layer to enhance pattern detection. Four Bidirectional LSTM layers then capture contextual understanding by considering past and future states. To prevent overfitting, four Dropout layers are placed after the LSTM layers. Finally, two dense layers are used to perform the final regression task and map features to the desired output format. By containing the BiLSTM model, we aimed to gain insights into RNN-based approaches for blood pressure prediction and expand the scope of evaluation by comparing our model performance with other benchmark models. Lastly, we incorporated a hybrid architecture combining CNN and RNN layers, which utilized the raw PPG signal along with its first and second derivatives of the MIMIC III dataset as input features⁴³. It is essential to note that the pre-processing methods employed in the benchmark studies were customized for their specific MIMIC dataset version. As the structure of the MIMIC IV dataset differs from the earlier versions, we utilized our pre-processing pipelines. Additionally, since the referenced studies employed a 10 – second window size, we maintained consistency by adopting the same window length for the data.

Table 4 presents the quantitative evaluation metrics obtained by implementing the benchmark models on the MIMIC IV dataset. Each benchmark model was trained using different sets of hyperparameters, and the most accurate model was identified and reported. The results demonstrate the superior performance of our proposed model compared to the benchmark algorithms across all the evaluation metrics. This notable performance confirms the efficacy of our approach in accurately estimating blood pressure solely from PPG signals.

4 Discussion and Conclusion

Long-term BP measurement could not be carried out using cuff-based solutions, because of its discrete nature and the inaccuracy caused by factors such as the cuff size, placement, and the patient’s movement. Consequently, there has been a surge of recent

interest in using DL techniques to estimate BP based on ECG and/or PPG data⁴⁴. In this context, the TransfoRhythm framework is developed in this study for continuous BP estimation. The TransfoRhythm takes as input 12 features extracted from the MIMIC IV raw and derivatives PPG waveforms. It is worth noting that despite many studies on the MIMIC datasets, to the best of our knowledge, this work is the first investigation of BP prediction implemented on the MIMIC IV waveform version dataset. At the core of the TransfoRhythm framework is a Multi-Head Attention (MHA)-based mechanism, which is a neural network architecture with several advantages over its traditional DL methods for time series analysis⁴⁵. Several studies have shown that MHA can outperform conventional DL methods in various time series prediction tasks. For example, MHA-based networks have been shown to outperform RNNs in predicting fault diagnosis⁴⁶, Energy Consumption assessment⁴⁷, and speech enhancement⁴⁸. Additionally, MHA has been shown to improve prediction accuracy in stock price prediction⁴⁹, and traffic flow prediction⁵⁰. When it comes to BP estimation, conventional DL methods are, typically, developed based on RNN and/or CNN architectures. While these methods have shown promising results, they have limitations, such as difficulty in capturing long-term dependencies or invariance to temporal shifts in the input⁵¹. The primary advantage of MHA is its ability to attend to various input sequence segments, enabling the model to learn complicated relationships between various observations in the sequence and improve representations of the input. It, therefore, can capture both short-term and long-term dependencies in time series data, where detailed relations between distinct observations might be difficult to capture by conventional DL methods⁵². Attention-based networks have also the potential to enhance forecasting accuracy and improve the robustness of time series analysis by handling noise and perturbations⁵³. Another merit of the MHA is its ability to handle variable-length inputs without using reduction methods, making it possible to learn more accurate representations of the input. In conventional DL methods, padding or truncation is often used to ensure all input sequences have the same length. This could lead to information loss and reduced model performance⁴⁷. Finally, the Attention mechanism can also be used to improve the interpretability of the model in diverse academic contexts⁵⁴. Attention scores can provide insight into the specific segments of the input sequence that the model focuses on when making predictions⁵⁵. This can be extremely practical in BP studies, in which identifying important observations can provide valuable insight.

Our results, obtained through different evaluation criteria, demonstrate the supremacy of the proposed TransfoRhythm framework for estimating BP. In terms of MAE and STD, the model obtained an interval of [1.50, 1.84] mmHg for DBP and [1.17, 1.42] mmHg for SBP. These findings fully satisfy the AAMI SP10 and BHS standard requirements. The lower MAE and RMSE values achieved by the proposed model highlight its potential for accurate and reliable blood pressure estimation in real-world applications. To further evaluate the performance of the proposed model, the following two assessment stages are considered.

- Benchmarks on MIMIC IV Dataset:** In the first stage, the model's performance on the MIMIC IV dataset is compared to that of benchmark models trained on the same dataset. The results are shown in Table 4. This comparative evaluation ensures a fair assessment of the model's performance and provides confidence in its accuracy and reliability for BP estimation. Among the evaluated models, the proposed TransfoRhythm outperforms its counterparts, which could be attributed to its internal MHA mechanism. By selectively attending to important features, the MHA mechanism allows the TransfoRhythm to effectively capture the relevant temporal dependencies and spatial relationships present in the PPG signals. Furthermore, it allows the model to focus on informative patterns, leading to more accurate BP estimations. In the benchmark models, the ResNet1D model demonstrates the best performance (second best compared to the TransfoRhythm). ResNet1D is a variant of the Residual Neural Network (ResNet) that utilizes one-dimensional convolutional layers to capture spatial features from the PPG signals. By employing residual connections, ResNet1D can effectively address the vanishing gradient problem and facilitate the flow of gradients during training. Following ResNet1D, the hybrid⁴³ and the *U-Net* models exhibit competitive performance. The hybrid architecture combines the strengths of both CNN and RNN allowing it to capture both spatial and temporal dependencies in the PPG signals. The U-Net architecture is modified to effectively capture temporal dependencies and extract meaningful features from sequential data. The AlexNet1D model performs slightly worse than ResNet1D and the hybrid model, achieving an SBP MAE of 3.09, a DBP MAE of 1.59, an SBP RMSE of 7.98, and a DBP RMSE of 4.6. AlexNet1D employs 1D convolutional layers followed by fully connected layers to extract features from the PPG signals. Another evaluated benchmark model utilizes Bi-LSTM layers to capture temporal dependencies in the PPG signals. Bi-LSTM processes the input sequence in both forward and backward directions, effectively encoding past and future information. Although Bi-LSTM performs reasonably well, it does not achieve the same level of accuracy as our proposed model, ResNet1D, the hybrid model, *U-Net*, and AlexNet1D models.
- Comparisons with the State-of-The-Art Models:** In the second stage of performance evaluation, the model's functionality is investigated by comparing it to related state-of-the-art studies. Table 5 illustrates the comparisons, which include model architecture, input signal type, dataset, feature injection model, and metrics. The main objective is to provide a comprehensive evaluation of the model's performance, independent of dataset consideration. This analysis allows for a

Table 5. Comparison of studies in BP estimation in terms of methodology, database, features, and metrics (NR means no result reported)

Model Architecture	Signal	Datasets	Model Injection Data Type	MAE SBP	MAE DBP	RMSE SBP	RMSE DBP	STD SBP	STD DBP
MLP1stm-BP (Based on MLP Mixer) ¹⁴	ECG + PPG	MIMIC II	Raw Waveform (multi-filter to multi-channel)	3.52	2.13	5.1	3.13	5.09	3.07
Temporal Convolution Network (TCN) ¹	ECG + PPG	MIMIC I + MIMIC III	Feature Extraction	2.59	1.33	3.03	1.58	3.58	1.97
CNN + LSTM ¹³	ECG + PPG	MIMIC III	Raw Waveform Concatenation	4.41	2.91	NR	NR	6.11	4.23
NABNet (BiConvLSTM) ²²	ECG + PPG	MIMIC III	Raw Waveform (multi-channel)	2.63	1.09	NR	NR	2.96	1.05
Bi-LSTM ⁵⁶	ECG + PPG	MIMIC II	Feature Extraction	2.51	1.38	3.22	1.78	NR	NR
Deep Autoencoder ²¹	PPG	MIMIC II	Raw Waveform Concatenation	5.42	3.14	NR	NR	6.64	3.74
1D_CNN + 2D_CNN ⁴⁴	PPG	MIMIC II	Raw Waveform + Image	3.05	1.58	NR	NR	5.26	2.6
CNN (Transfer Learning) ³²	PPG	MIMIC II	Image (Transformed PPG)	6.17	3.66	8.46	5.36	8.46	5.36
Bi-GRU + Attention ³¹	PPG	MIMIC II	Feature Extraction (22 features)	2.58	1.26	NR	NR	3.35	1.63
BiLSTM + LSTM + Attention ²⁶	PPG	MIMIC II	Feature Extraction	4.51	2.6	NR	NR	7.81	4.41
GRU ²⁴	PPG	MIMIC II	Feature Extraction (7 features)	3.25	1.43	NR	NR	4.76	1.77
DNN ²⁵	PPG	MIMIC II	Feature Extraction (32 features)	3.21	2.23	4.64	3.3	NR	NR
U-Net ⁴²	PPG	MIMIC III	Raw Waveform	5.73	3.45	NR	NR	9.16	6.14
Feed Forward ANN ⁵⁷	PPG	MIMIC II + UCI	Feature Extraction (21 features)	7.41	3.32	NR	NR	10.40	4.89
CNN+LSTM ⁵⁸	PPG	MIMIC III	Feature Extraction	3.79	2.89	NR	NR	7.89	5.34
ArterialNet Transformer ³⁷	PPG	MIMIC III	raw waveform	4.15	3.17	5.26	4.01	1.32	1.37
Transformer ³⁵	PPG	UCI	raw waveform	19.0	9.0	NR	NR	NR	NR
FD L/NL regression ³⁵	PPG	UCI	Frequency-domain waveform	11.87	8.01	NR	NR	NR	NR
STP (Self-supervised transformer) ³⁴	PPG	MIMIC III	raw waveform	3.37	2.48	NR	NR	4.21	2.76
TransfoRhythm (This Study)	PPG	MIMIC IV (360 record)	Feature Extraction (12 features)	1.50	1.17	1.84	1.42	1.84	1.42

thorough assessment of the model's functionality and its relative standing compared to other advanced approaches in the field of BP estimation. Regarding datasets, most of the similar works employed MIMIC II or MIMIC III datasets. MIMIC-IV, which was used in this study in comparison with other available MIMIC datasets includes more PPG data. Meanwhile, MIMIC-IV includes more recent data, which may be more representative of current clinical practices,

and has undergone rigorous quality control processes to ensure that the data is accurate and reliable³⁸. Regarding the model selection, first, we should note that multiple studies applied traditional ML methods such as SVM⁵⁹, RF⁶⁰, and XGBOOST⁶¹. These are not mentioned in Table 5, as despite achieving acceptable results in some scenarios, DL methods tend to perform considerably better given the time series nature of ECG and PPG signals. Some other studies, such as³², demonstrate good performance in BP prediction using transfer learning without RNN layers. In another type of modeling, RNN networks, especially LSTM and GRU, have been used to extract two sets of features, e.g.²⁴. The importance of employing hybrid networks to extract temporal and recursive features of the signal was indicated by multiple studies^{26,29,31}. For instance, Malayeri *et al.* used a hybrid CNN-based network and gathered the recursive features of the signal with a 2D CNN structure Instead of RNN networks⁴⁴. Furthermore, a combination of RNN and attention layers was presented by^{26,31}, and the importance of attention usage was shown.

Despite the state-of-the-art performance achieved by the proposed TransfoRhythm framework, below are some restrictions and potential future research directions for this study. In terms of generalization, we used a portion of the MIMIC IV dataset for the training and validation phases, and all findings are based on this data. In other words, the TransfoRhythm is the first BP estimation model developed solely based on PPG signals from the MIMIC IV dataset. It would be interesting to compare the model with new developments in the MIMIC IV dataset. Furthermore, to improve the model's generalizability, it would be interesting to extend it by incorporating other data sources. A final fruitful future research direction is to go beyond the utilization of a pure attention model, capitalizing on the recent progression of attention-based networks.

References

1. Zabihi, S. *et al.* Bp-net: Cuff-less and non-invasive blood pressure estimation via a generic deep convolutional architecture. *Biomed. Signal Process. Control.* **78**, 103850 (2022).
2. Heidenreich, P. A. *et al.* Forecasting the future of cardiovascular disease in the united states: a policy statement from the american heart association. *Circulation* **123**, 933–944 (2011).
3. Lewis, P. S. & British and Irish Hypertension Society's Blood Pressure Measurement Working Party. Oscillometric measurement of blood pressure: a simplified explanation. a technical note on behalf of the british and irish hypertension society. *J. human hypertension* **33**, 349–351 (2019).
4. Mukkamala, R., Stergiou, G. S. & Avolio, A. P. Cuffless blood pressure measurement. *Annu. Rev. Biomed. Eng.* **24**, 203–230 (2022).
5. Kario, K. *et al.* Effect of esaxerenone on nocturnal blood pressure and natriuretic peptide in different dipping phenotypes. *Hypertens. Res.* **45**, 97–105 (2022).
6. Hu, J.-R. *et al.* Validating cuffless continuous blood pressure monitoring devices. *Cardiovasc. Digit. Heal. J.* (2023).
7. Shao, J. *et al.* An optimization study of estimating blood pressure models based on pulse arrival time for continuous monitoring. *J. Healthc. Eng.* **2020** (2020).
8. Yousefian, P. *et al.* Pulse transit time-pulse wave analysis fusion based on wearable wrist ballistocardiogram for cuff-less blood pressure trend tracking. *IEEE Access* **8**, 138077–138087 (2020).
9. Esmaili, A., Kachuee, M. & Shabany, M. Nonlinear cuffless blood pressure estimation of healthy subjects using pulse transit time and arrival time. *IEEE Transactions on Instrumentation Meas.* **66**, 3299–3308 (2017).
10. Pielmus, A.-G. *et al.* Surrogate based continuous noninvasive blood pressure measurement. *Biomed. Eng. Tech.* **66**, 231–245 (2021).
11. Geddes, L., Voelz, M., Babbs, C., Bourland, J. & Tacker, W. Pulse transit time as an indicator of arterial blood pressure. *psychophysiology* **18**, 71–74 (1981).
12. Mishra, B. & Thakkar, N. Cuffless blood pressure monitoring using ptt and pwv methods. In *2017 International Conference on Recent Innovations in Signal processing and Embedded Systems (RISE)*, 395–401 (IEEE, 2017).
13. 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 (2021).
14. Huang, B., Chen, W., Lin, C.-L., Juang, C.-F. & Wang, J. Mlp-bp: A novel framework for cuffless blood pressure measurement with ppg and ecg signals based on mlp-mixer neural networks. *Biomed. Signal Process. Control.* **73**, 103404 (2022).

15. Tolstikhin, I. O. *et al.* Mlp-mixer: An all-mlp architecture for vision. *Adv. neural information processing systems* **34**, 24261–24272 (2021).
16. Goldberger, A. L. *et al.* Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *circulation* **101**, e215–e220 (2000).
17. Le, T. *et al.* Continuous non-invasive blood pressure monitoring: a methodological review on measurement techniques. *IEEE Access* **8**, 212478–212498 (2020).
18. Welykholowa, K. *et al.* Multimodal photoplethysmography-based approaches for improved detection of hypertension. *J. Clin. Medicine* **9**, 1203 (2020).
19. Bassiouni, M. M., Hegazy, I., Rizk, N., El-Dahshan, E.-S. A. & Salem, A. M. Combination of ecg and ppg signals for smart healthcare systems: Techniques, applications, and challenges. In *2021 Tenth International Conference on Intelligent Computing and Information Systems (ICICIS)*, 448–455 (IEEE, 2021).
20. Chao, P. C.-P. *et al.* The machine learnings leading the cuffless ppg blood pressure sensors into the next stage. *IEEE Sensors J.* **21**, 12498–12510 (2021).
21. Qin, K., Huang, W. & Zhang, T. Deep generative model with domain adversarial training for predicting arterial blood pressure waveform from photoplethysmogram signal. *Biomed. Signal Process. Control.* **70**, 102972 (2021).
22. Mahmud, S. *et al.* Nabnet: a nested attention-guided biconvlstm network for a robust prediction of blood pressure components from reconstructed arterial blood pressure waveforms using ppg and ecg signals. *Biomed. Signal Process. Control.* **79**, 104247 (2023).
23. Hassani, A. & Foruzan, A. H. Improved ppg-based estimation of the blood pressure using latent space features. *Signal, Image Video Process.* **13**, 1141–1147 (2019).
24. El Hajj, C. & Kyriacou, P. A. Cuffless and continuous blood pressure estimation from ppg signals using recurrent neural networks. In *2020 42nd annual international conference of the IEEE engineering in medicine & biology society (EMBC)*, 4269–4272 (IEEE, 2020).
25. Hsu, Y.-C., Li, Y.-H., Chang, C.-C. & Harfiya, L. N. Generalized deep neural network model for cuffless blood pressure estimation with photoplethysmogram signal only. *Sensors* **20**, 5668 (2020).
26. El-Hajj, C. & Kyriacou, P. A. Cuffless blood pressure estimation from ppg signals and its derivatives using deep learning models. *Biomed. Signal Process. Control.* **70**, 102984 (2021).
27. Gupta, S., Singh, A., Sharma, A. & Tripathy, R. K. Higher order derivative-based integrated model for cuff-less blood pressure estimation and stratification using ppg signals. *IEEE Sensors J.* **22**, 22030–22039 (2022).
28. Liu, M., Po, L.-M. & Fu, H. Cuffless blood pressure estimation based on photoplethysmography signal and its second derivative. *Int. J. Comput. Theory Eng.* **9**, 202 (2017).
29. Rastegar, S., Gholam Hosseini, H. & Lowe, A. Hybrid cnn-svr blood pressure estimation model using ecg and ppg signals. *Sensors* **23**, 1259 (2023).
30. Paviglianiti, A., Randazzo, V., Villata, S., Cirrincione, G. & Pasero, E. A comparison of deep learning techniques for arterial blood pressure prediction. *Cogn. Comput.* **14**, 1689–1710 (2022).
31. El-Hajj, C. & Kyriacou, P. A. Deep learning models for cuffless blood pressure monitoring from ppg signals using attention mechanism. *Biomed. Signal Process. Control.* **65**, 102301 (2021).
32. 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. Heal. Informatics* **26**, 2075–2085 (2021).
33. Johnson, A. *et al.* Mimic-iv. *PhysioNet*. Available online at: <https://physionet.org/content/mimic-iv-ed/2.0/> (accessed August, 2023) (2020).
34. Ma, C. *et al.* Stp: Self-supervised transfer learning based on transformer for noninvasive blood pressure estimation using photoplethysmography. *Expert. Syst. with Appl.* 123809 (2024).
35. Tahir, M. A. *et al.* Cuff-less arterial blood pressure waveform synthesis from single-site ppg using transformer & frequency-domain learning. *arXiv preprint arXiv:2401.05452* (2024).
36. Reiss, A., Indlekofer, I. & Schmidt, P. PPG-DaLiA. UCI Machine Learning Repository (2019). DOI: <https://doi.org/10.24432/C53890>.
37. Huang, S., Jafari, R. & Mortazavi, B. J. Arterialnet: Arterial blood pressure reconstruction. In *2023 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI)*, 1–4 (IEEE, 2023).

38. Moody, B. *et al.* Mimic-iv waveform database (version 0.1.0), DOI: [10.13026/a2mw-f949](https://doi.org/10.13026/a2mw-f949) (2022).
39. Kalyan, K. S. & Sangeetha, S. SecNlp: A survey of embeddings in clinical natural language processing. *J. biomedical informatics* **101**, 103323 (2020).
40. Vaswani, A. *et al.* Attention is all you need. *Adv. neural information processing systems* **30** (2017).
41. Schrumpp, F., Frenzel, P., Aust, C., Osterhoff, G. & Fuchs, M. Assessment of non-invasive blood pressure prediction from ppg and rppg signals using deep learning. *Sensors* **21**, 6022 (2021).
42. Ibtehaz, N. *et al.* Ppg2abp: Translating photoplethysmogram (ppg) signals to arterial blood pressure (abp) waveforms. *Bioengineering* **9**, 692 (2022).
43. Slapničar, G., Mlakar, N. & Luštrek, M. Blood pressure estimation from photoplethysmogram using a spectro-temporal deep neural network. *Sensors* **19**, 3420 (2019).
44. Malayeri, A. B. & Khodabakhshi, M. B. Concatenated convolutional neural network model for cuffless blood pressure estimation using fuzzy recurrence properties of photoplethysmogram signals. *Sci. Reports* **12**, 6633 (2022).
45. Wu, F., Fan, A., Baevski, A., Dauphin, Y. N. & Auli, M. Pay less attention with lightweight and dynamic convolutions. *arXiv preprint arXiv:1901.10430* (2019).
46. Jiang, L., Li, X., Wu, L. & Li, Y. Bearing fault diagnosis method based on a multi-head graph attention network. *Meas. Sci. Technol.* **33**, 075012 (2022).
47. Bu, S.-J. & Cho, S.-B. Time series forecasting with multi-headed attention-based deep learning for residential energy consumption. *Energies* **13**, 4722 (2020).
48. Nicolson, A. & Paliwal, K. K. Masked multi-head self-attention for causal speech enhancement. *Speech Commun.* **125**, 80–96 (2020).
49. Chen, Z., Jiaze, E., Zhang, X., Sheng, H. & Cheng, X. Multi-task time series forecasting with shared attention. In *2020 International Conference on Data Mining Workshops (ICDMW)*, 917–925 (IEEE, 2020).
50. Reza, S., Ferreira, M. C., Machado, J. J. & Tavares, J. M. R. A multi-head attention-based transformer model for traffic flow forecasting with a comparative analysis to recurrent neural networks. *Expert. Syst. with Appl.* **202**, 117275 (2022).
51. Blohm, M., Jagfeld, G., Sood, E., Yu, X. & Vu, N. T. Comparing attention-based convolutional and recurrent neural networks: Success and limitations in machine reading comprehension. *arXiv preprint arXiv:1808.08744* (2018).
52. Li, Y., Du, M. & He, S. Attention-based sequence-to-sequence model for time series imputation. *Entropy* **24**, 1798 (2022).
53. Niu, P., Zhou, T., Wang, X., Sun, L. & Jin, R. Attention as robust representation for time series forecasting. *arXiv preprint arXiv:2402.05370* (2024).
54. Li, A., Xiao, F., Zhang, C. & Fan, C. Attention-based interpretable neural network for building cooling load prediction. *Appl. Energy* **299**, 117238 (2021).
55. Ahmed, S. *et al.* Transformers in time-series analysis: A tutorial. *arXiv preprint arXiv:2205.01138* (2022).
56. Chen, H., Lyu, L., Zeng, Z., Jin, Y. & Zhang, Y. Beat-to-beat continuous blood pressure estimation with optimal feature set of ppg and ecg signals using deep recurrent neural networks. *Vessel. Plus* **7** (2023).
57. Samimi, H. & Dajani, H. R. A ppg-based calibration-free cuffless blood pressure estimation method using cardiovascular dynamics. *Sensors* **23**, 4145 (2023).
58. Mahardika T, N. Q., Fuadah, Y. N., Jeong, D. U. & Lim, K. M. Ppg signals-based blood-pressure estimation using grid search in hyperparameter optimization of cnn-lstm. *Diagnostics* **13**, 2566 (2023).
59. Zhang, Y. & Feng, Z. A svm method for continuous blood pressure estimation from a ppg signal. In *Proceedings of the 9th international conference on machine learning and computing*, 128–132 (2017).
60. He, R. *et al.* Beat-to-beat ambulatory blood pressure estimation based on random forest. In *2016 IEEE 13th international conference on wearable and implantable body sensor networks (BSN)*, 194–198 (IEEE, 2016).
61. Che, X., Li, M., Kang, W., Lai, F. & Wang, J. Continuous blood pressure estimation from two-channel ppg parameters by xgboost. In *2019 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, 2707–2712 (IEEE, 2019).

Acknowledgments

This Project was partially supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada through the NSERC Discovery Grant RGPIN-2023-05654.

Data Availability

The utilized dataset is publicly available through the following link: <https://physionet.org/content/mimic4wdb/0.1.0/>³⁸.

Author Contributions Statement

A.A. and A.B. implemented the deep/machine learning models and performed evaluations; A.A and A.B. together with F.B. drafted the manuscript; F.B. and A.M. contributed to the analysis and interpretation and edited the manuscript; All authors reviewed the manuscript.

Additional Information

Competing Interests: Authors declare no competing interests.