

Temporal Cardiovascular Dynamics for Improved PPG-Based Heart Rate Estimation

Berken Utku Demirel and Christian Holz

Abstract—The oscillations of the human heart rate are inherently complex and non-linear—they are best described by *mathematical chaos*, and they present a challenge when applied to the practical domain of cardiovascular health monitoring in everyday life. In this work, we study the non-linear chaotic behavior of heart rate through mutual information and introduce a novel approach for enhancing heart rate estimation in real-life conditions. Our proposed approach not only explains and handles the non-linear temporal complexity from a mathematical perspective but also improves the deep learning solutions when combined with them. We validate our proposed method on four established datasets from real-life scenarios and compare its performance with existing algorithms thoroughly with extensive ablation experiments. Our results demonstrate a substantial improvement, up to 40%, of the proposed approach in estimating heart rate compared to traditional methods and existing machine-learning techniques while reducing the reliance on multiple sensing modalities and eliminating the need for post-processing steps.

Index Terms—Hearth rate monitoring, photoplethysmography, deep neural networks, temporal dynamics.

I. INTRODUCTION

Healthy biological systems exhibit complex patterns of variability that can be described by *mathematical chaos* [1], [2]. A healthy heart is not a metronome; instead, its complex and constantly changing oscillations enable the cardiovascular system to rapidly adjust to sudden physical and psychological challenges to homeostasis [2]. Therefore, measuring heart rate (HR) during daily life has significant importance in monitoring individuals' health. For example, increased oscillations of HR measurements are related to cardiac conduction abnormalities, which are strongly linked to an increased risk of mortality (particularly among the elderly) [3].

Moreover, the changes in heart rate to activities, such as HR variations during running, have been shown to be predictive of cardiovascular disease (CVD) risk and are inversely associated with many other health outcomes [4]. Therefore, monitoring heart rate variations in daily life, especially during activities, not only provides valuable insights into cardiovascular health but also serves as a predictive tool for assessing the risk of cardiovascular diseases and cardiovascular mortality [5].

The electrocardiogram (ECG) is the primary tool for monitoring cardiac activity. However, it requires electrodes to be

Received 7 February 2025; revised 21 July 2025; accepted 25 September 2025. The authors are with the Department of Computer Science, ETH Zürich, Stampfenbachstrasse 48, 8092 Zurich, Switzerland (e-mail: berken.demirel@inf.ethz.ch, christian.holz@inf.ethz.ch).

connected to various parts of the body and, therefore, can only be used for a limited amount of time within medical facilities [6]. Recently, wearable devices have emerged as powerful platforms for monitoring the cardiac activity of users during their daily life through blood flow measurement sensors, i.e., photoplethysmogram (PPG). Commercial products, such as the Apple Watch [7], Google's Fitbit [8], or the Samsung Watch [9], now include these sensors to track the wearer's heart rate, for health monitoring and fitness purposes.

However, measurements on wearable devices are prone to motion artifacts (MAs), which cause variability of sensor pressure on the skin while scattering or leaking ambient light into the gap between the photodiode and the skin, resulting in a significant decrease in heart rate estimation performance [9].

For instance, when users engage in physical activities, such as walking, running, or working, the HR estimation performance in wearable devices decreases heavily. The studies that are concerned with commercial wearables state that the performance of devices extremely decreases during high-intensity activities. For example, reviewing studies on Fitbit reported severe underestimations of the HR [10]–[12]. Similarly, previous works showed that the performance of the Apple Watch for HR estimation during exercise decreases with increasing intensity and the proportion of HR values recorded by the watch decreases [13]. In other words, previous works reported a trivial underestimation of the heart rate from the Apple Watch compared to the ground truth [14].

Moreover, a recent study shows that in a commercial device, HR estimation errors reach 80–100 beats per minute (bpm) during wake periods where data recorded from sensors are contaminated by different levels of noise [9]. While the Association for the Advancement of Medical Instrumentation (AAMI) standard indicates that for cardiac monitors the allowable error rate should be less than ± 5 bpm or $\pm 10\%$, whichever is greater [15]. Although this error is defined for a single estimation instead of the mean, current state-of-the-art methods, especially those using deep learning, fall short of this performance. Moreover, estimating the HR during intensive activities is crucially important for monitoring individuals' cardiovascular health while the current solutions fail. Current solutions for estimating HR from blood volume pulse signals remain straightforward while applying machine learning as they have not dealt with the complex and non-linear dynamics of the cardiovascular system, which limits performance and overshadows the learning capability of algorithms.

In this paper, we first study the chaotic behavior of the heart rate in real-life conditions through mutual information and

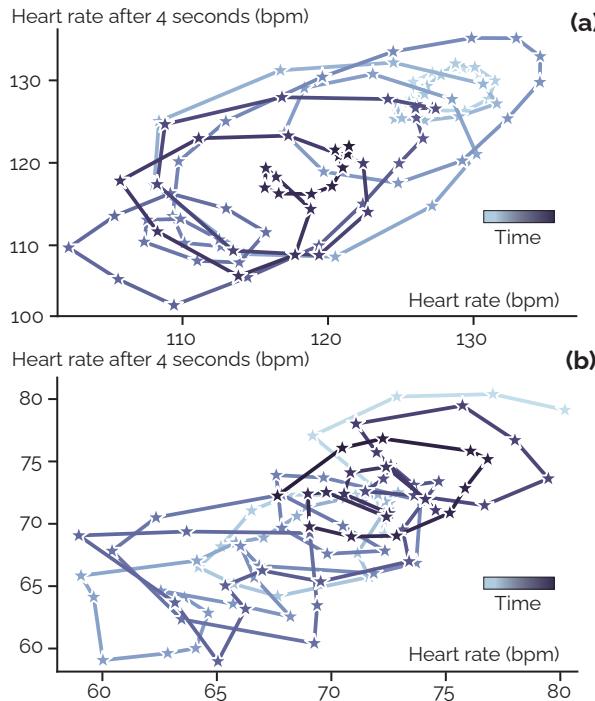


Fig. 1: Heart rate dynamics during walking (a) and reading (b). Both figures show the phase space plot of sinus rhythm for two healthy subjects over three minutes. Each point represents the current HR (x-axis) and the HR after 4 seconds (y-axis). The gradual darkening of the line color represents the time, which illustrates the chaotic fluctuations of heart rate.

introduce a novel approach for enhancing heart rate estimation with deep learning solutions. Our proposed method overcomes previous limitations of motion artifacts by learning the chaotic behavior of cardiovascular dynamics together with the sensor data and utilizes this for predictions during inference without requiring any additional post-processing or sensor modalities. The contributions can be summarized as follows:

- We present an information-theoretical approach for understanding the non-linear chaotic behavior of cardiovascular dynamics. To the best of our knowledge, we are the first to investigate the non-linear temporal relations in the heart rate with a mathematical approach to identify underlying patterns within this complex behavior, shedding light on previously unrecognized order.
- Building on this, we integrate our findings into deep learning solutions to improve the heart rate estimations, different from intuition-based traditional methods and trial-error-based learning solutions.
- We extensively evaluated our method on established HR estimation datasets that span a wide range of physical activities, performed under close to real-life conditions. Our method outperforms the prior works in HR estimation. Unlike previous methods, our method produces estimations with high robustness over time while decreasing the number of required sensing modalities and discarding the need for post-processing.

II. RELATED WORKS

A considerable amount of literature has been published on estimating heart rate from wearable devices based on PPG signals. Most of these studies utilized power spectral density (PSD) estimation, a quantitative method for describing a signal in terms of its underlying oscillations, enabling one to observe how the frequency changes over time. Assuming the heart rate will not change abruptly and be stationary in a short window, the PSD reveals information about heartbeat frequency. While this approach works for less contaminated PPG, motion artifacts hinder the measurement of heart rate in spectral density. As a result, different approaches have been proposed to obtain the PSD of PPG signals with accelerometers to differentiate the frequency of heartbeats from motion artifacts [16]–[21].

These approaches have significant drawbacks. First, the accelerometer signals do not always have the true motion artifacts (for example, finger tapping, wrist-twisting, and fist clenching/unfolding) [22], which limits the performance of the solutions heavily. Second, the dominant frequencies from the accelerometer signals may overlap with the true frequencies of the HR, which makes signals impossible to differentiate by investigating the frequency domain. Finally, since these solutions use multiple modalities [23] to increase performance, they are more susceptible to sensor faults, such as communication or hardware failures. Also, it is known that low-cost IMUs are vulnerable to various faults because of temperature variation, performance degradation, component damage [24], [25]. The inaccurate information provided by the IMU will seriously affect the performance of these algorithms. Deep-learning solutions have been proposed to make the system more robust by only utilizing PPG signals [16], [26]–[28]. However, as learning models have no information about the degree of motion artifact contamination in PPG signals, they can estimate HR with large errors. Or even though the IMU is fed together with PPG, the models cannot learn the relation between motion artifact and PPG signals and estimate the HR incorrectly [29], [30]. To prevent this, several works have proposed applying additional post-processing that considers the estimated values from preceding segments to detect and fix the abrupt changes in the model's output [16], [29]. Although current solutions have their own specific values, they remain straightforward and have not dealt with the rich dynamics of the cardiovascular system, which limits performance and overshadows the learning capability of algorithms. In this work, we introduce a method that requires a single modality, PPG, to estimate the heart rate while concerning the temporal dynamics without requiring any post-processing.

III. MOTIVATION

Several studies have attempted to model cardiovascular dynamics in heart rate estimation using linear methods such as mean filters [16], [29], curve tracking [31] or linear regression models [32]. Although the linear relation between temporal heartbeats can be useful to predict the current HR, the oscillations of a heart are complex and non-linear [2] while exhibiting a chaotic non-Markovian dynamical system [1].

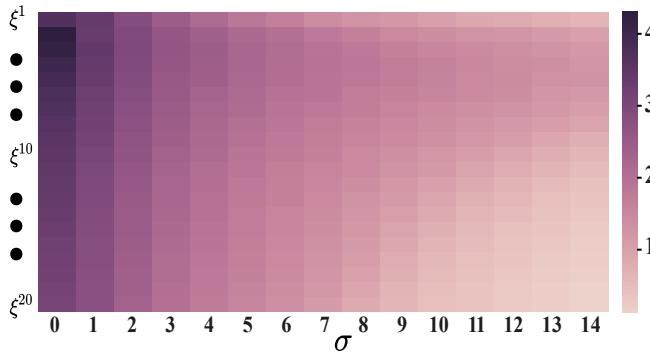


Fig. 2: The heatmap of estimated mutual information between random variables ξ (previous heart rate values) and Y (the current heart rate). We investigated the effect of time (y-axis) and noise (x-axis) for ξ with $\xi = \xi + \sigma \odot \epsilon$ where $\epsilon \sim \mathcal{N}(0, I)$. And, $\xi^N = [y_{t-1}, y_{t-2}, \dots, y_{t-N}]$.

We first study the non-linear temporal relations of cardiovascular dynamics. To investigate the relationship between heartbeat values over time, we calculated the Mutual Information (MI) between them. Unlike linear methods such as correlation and mean filters, mutual information, which is a Shannon entropy-based measure of dependence [33] as defined in Equation 1 for variables X and Z , captures non-linear statistical dependencies between variables and thus can act as a measure of true dependence [34].

$$I(X; Z) = \int_{\mathcal{X} \times \mathcal{Z}} \log \frac{d\mathbb{P}_{XZ}}{d\mathbb{P}_X \otimes \mathbb{P}_Z} d\mathbb{P}_{XZ}, \quad (1)$$

where \mathbb{P}_{XZ} is the joint probability distribution, $\mathbb{P}_X = \int_Z d\mathbb{P}_{XZ}$ and $\mathbb{P}_Z = \int_X d\mathbb{P}_{XZ}$ are the marginals. The mutual information between X and Z can be understood as the decrease of the uncertainty in X given Z . We, therefore, calculated the mutual information between previous (past) heartbeats and the current one as in Equation 2.

$$I(Y; \xi) := H(Y) - H(Y | \xi), \quad (2)$$

where H , Y , and ξ are the Shannon entropy, current heart rate, and past HR values, respectively. $H(Y | \xi)$ is the conditional entropy of Y given ξ . $I(Y; \xi)$ measures how much information on average previous heart rates convey about the current value without seeking any linear relations.

Figure 2 shows that knowing previous HR values reduces the uncertainty of the current HR. Moreover, it also indicates that even the past values are observed with an error, they carry information about the future. Also, the noise affects the MI in different degrees depending on how many past values are included in ξ . The experiment for additive noise is especially important as it explains the effect from a statistical point of the possible realizations of past values on future estimations. We note that small decreases in MI with longer windows come from finite-sample effects and estimator bias/variance, not a violation of the data processing inequality.

A. The non-linearity

Previous studies have often relied on leveraging prior HR values to address errors in predictions. However, many of

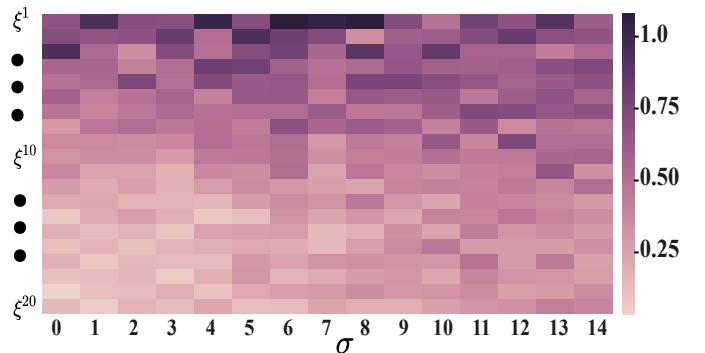


Fig. 3: The heatmap of calculated Pearson correlation between random variables ξ (previous heart rate values) and Y (the current heart rate). We investigated the effect of time (y-axis) and noise (x-axis) for ξ with $\xi = \xi + \sigma \odot \epsilon$ where $\epsilon \sim \mathcal{N}(0, I)$. And, $\xi^N = [y_t, y_{t-1}, y_{t-2}, \dots, y_{t-N}]$.

these techniques were characterized by their straightforward and predominantly linear nature. Here, we seek to find any linear relations in HR using the Pearson correlation coefficient (PCC). PCC calculates the linear correlation between two variables and ignores other types of relationships since it is defined as with covariance itself.

When we calculate the PCC, we follow the same setup with MI estimation and take one variable as time and sum the absolute value of PCC instead of ± 1 since the overall trend is important rather than increase and decrease. Figure 3 shows the PCC values where no clear and distinct pattern is observed at all. Conversely, there are some cases in which the addition of noise increases the correlation between past and future HRs. These experiments show that HR changes are non-linear. Thus, we propose a method to capture non-linear variations for improved HR estimation, unlike the linear methods used in previous works [16], [29], [32].

IV. OBJECTIVE

Considering our findings, we aim to train a learner f_θ which seeks to learn the behavior of heart rate variability with volumetric changes in blood together for estimating the current heart rate value. Therefore, we only require volumetric blood volume change signals to estimate the heart rate, unlike the related works which require additional modalities [31], [35].

A. Framework

The overall proposed framework can be explained as follows: Given an observation of the blood volume signal \mathbf{x}_t that is a single dimensional vector drawn from the distribution \mathcal{X} at time t and composed of a sequence of voltage values with length T , (x_1, x_2, \dots, x_T) , we aim to map this sequence to the heart rate value of this segment y_t .

Previous deep learning approaches attempted to estimate the conditional probability $p(y_t | \mathbf{x}_t)$ by training different mapping functions, using Equation 3, that perform $f_\theta : \mathcal{X} \rightarrow \mathcal{Y}$ where the \mathcal{Y} is a one-dimensional space consisting of real-valued heart rate $y \in \mathcal{Y}$, and f_θ is a parametric family of mappings

(e.g., CNNs), under a loss function L .

$$\min_{\theta} \sum_t L(f_{\theta}(\mathbf{x}_t), y_t) \quad (3)$$

As the previous methods optimized the models to estimate y_t using only the current signal \mathbf{x}_t , they lack the ability to capture heart rate variability. As a result, when inputs deviate from distributions seen during training due to noise, the models produce large errors [16], [29], decreasing the performance.

To solve this problem and make the trained model suitable for real-life conditions, we train the model with the current signal and HR variations together. We optimize the model to estimate the heart rate distribution given the volumetric changes in blood and HR variations, i.e., $p(y_t|\mathbf{x}_t, \xi_t)$ where y_t is the heart rate of the current segment and ξ_t is the vector composed of previously estimated HR values.

We optimized the training task for a set of input-target pairs using the loss function shown in Equation 4 where the network function f_{θ} , is parameterized by θ :

$$\mathcal{L}_{\text{MAE}}(\theta) = \mathbb{E}[||y_t - f_{\theta}(\mathbf{x}_t, y_{t-1}, \dots, y_{t-K})||_1] \quad (4)$$

The loss function employed to train the learner is closely similar to transfer entropy, with the goal of maximizing it. Transfer entropy, $T_{Y \rightarrow X}$, is a non-parametric statistic measuring the rate of information flow between two random variables [36].

$$\sum p(y_t, y_{t-L}, x_{t-L}) \log \left(\frac{p(y_t|y_{t-L}, x_{t-L})}{p(y_t|y_{t-L})} \right) \quad (5)$$

In our case, we train the model to learn the information flow from the past to the future while incorporating the present. Hence, we characterize this as the conditional mutual information between variables, as defined in Equation 6.

$$T_{Y \rightarrow X} = I(X_t; Y_t | X_{t-1:t-L}) \quad (6)$$

Learning the function f_{θ} by minimizing Equation 7—also known as the empirical risk minimization—maximizes the conditional mutual information between the past, present, and future states of the dynamical complex system.

$$\min_{\theta} \sum_t L(f_{\theta}(\mathbf{x}_t, \xi_t), y_t) \quad (7)$$

In contrast to common dynamical systems [37], [38], cardiovascular dynamics are not Markovian due to their chaotic nature, such as we have no information when someone starts exercising and how it will affect the heart. We, therefore, incorporate our findings from MI analysis into a common network and show how a small modification can improve performance significantly. This non-linear relationship, as revealed through MI analysis (shown in Figure 2, derived from Equations 1 and 2), is instrumental in enhancing the model's performance.

B. Architecture

We modify the encoder-decoder architecture to learn heart rate variations together with blood volume changes. Specifically, we used linear layers with Long-Short Term Memory (LSTM) to encode the previous HR estimations and CNNs with LSTMs to extract features from the current segment, which is common in literature for sequential data [27], as

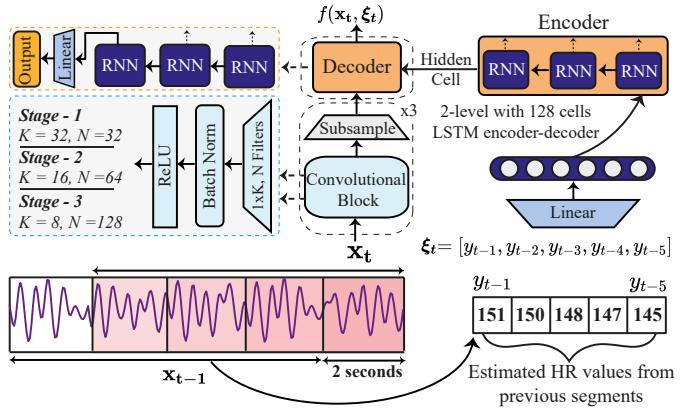


Fig. 4: Modified encoder-decoder architecture. The encoder part incorporates the HR variations into the model. The decoder outputs the HR by using both the current signal and encoded HR. Black arrows represent the input-output flow while double arrows show the content of a block.

shown in Figure 4. We first expand the HR values using a linear transformation with fifty output features. Then, extracted features are fed to the encoder. Our feature extractor includes convolutional blocks that are composed of 1D convolutions followed by batch normalization [39] and ReLU. We added a max-pooling layer with a size of two after each convolutional block. The decoder takes the extracted features from the output of the last pooling layer and conditions the encoder output. To estimate the HR, we used the last hidden state of the decoder with a linear layer. We also used a data augmentation to increase the diversity of heart rate variations. We, therefore, added a random noise sampled from the normal distribution to the HR, which is given in Equation 8. At the end of augmentation, some segments that are chosen randomly map to the different heart rate vectors ξ , with small differences.

$$\xi_{\text{aug}} = \xi + 3 \cdot z \text{ where } z \sim \mathcal{N}(0, 1) \quad (8)$$

Although this procedure adds noise during training, we observed significant performance gains, especially on smaller datasets, as it exposes the model to HR variations not present in the training data. Importantly, the added perturbations are small relative to overall HR dynamics. For example, in a commonly used large PPG dataset [16], HR values have a standard deviation of ≈ 17 bpm. With added noise of $\sigma=3$ bpm, the resulting Signal-to-Noise ratio (SNR) is $(17/3)^2 \approx 32$, meaning most of the original HR information is preserved. We randomly selected 10% of the HR vectors in the training data (excluding validation and test) for this augmentation.

V. EXPERIMENTAL SETUP

A. Datasets

We used four popular PPG datasets: IEEE Signal Processing Cup 2015 (SPC15) [44], [45], Dalia [16], WildPPG [46], and WESAD [47], which are selected to assess performance across diverse activities and age groups.

IEEE SPC: Includes 22 recordings from participants aged 18–58 [28]. Data includes accelerometer signals, two PPG

TABLE I: Performance comparison of methods. Best results are shown in **bold**, and second-best are underlined.

Work (Year)	Input	Pre-Processing	Method (# of params)	Post-Processing	Dataset	Results (bpm)
<i>Heuristic Solutions</i>						
Troika*[40]	PPG + ACC	Signal decomposition	Signal reconstruction	N/A	SPC15-12	2.34
SpaMa**[16]	PPG + ACC	0.5- to 3-Hz filtering Downsampling Freq. Transformation	Spectral filtering using power spectrum	Interpolation	SPC15-12 SPC15-22 Dalia WESAD	13.1±20.7 9.20±11.4 15.56±7.5 11.51±3.7
SpaMaPlus [16]	PPG + ACC	0.5- to 3-Hz filtering Downsampling Freq. Transformation	Picking the maximum in a weighted spectrum using linear prediction	HR tracking	SPC15-12 SPC15-22 Dalia WESAD	4.25±5.9 12.31±15.5 11.06±4.8 9.45±2.9
Schaeck [16] [17]	PPG + ACC	0.5- to 6-Hz filtering Downsampling Freq. Transformation	Comparing spectra of PPG and accelerometer	HR tracking	SPC15-12 SPC15-22 Dalia WESAD	2.91±4.6 24.65±24 20.45±7.1 19.97±8.1
CurToSS*[31]	PPG + ACC	Sparse spectrum reconstruction	Curve tracing	N/A	SPC15-12 SPC15-22 Dalia WESAD	2.2±– 4.5±– 5.0±2.8 6.4±1.8
<i>Deep Learning Solutions</i>						
CNN [16]	PPG + ACC	0- to 4-Hz filtering Freq. Transformation	CNN ensemble (8.5M)	N/A	SPC15-12 SPC15-22	4.0±5.4 16.51±16.1
PPGNet [26]	PPG	0.4- to 18-Hz filtering z-score normalization	CNN + LSTM (765k)	N/A	SPC15-12 SPC15-22	<u>3.36</u> 12.48
CorNet [41] [28]	PPG	0.4- to 18-Hz filtering z-score normalization	CNN + LSTM (257k)	N/A	SPC15-12 SPC15-22	4.67 <u>5.55</u>
BinCorNet [28]	PPG	0.4- to 18-Hz filtering z-score normalization	CNN + LSTM	N/A	SPC15-12 SPC15-22	6.78 7.32
CardioGAN [35]	PPG	1- to 8-Hz filtering z-score normalization	GANs (5.1M)	N/A	Dalia WESAD	8.4 8.6
Only-LSTM [27]	PPG	N/A	LSTM (680k)	N/A	SPC15-12	4.47±3.68
TimePPG [29]	PPG + ACC	0.5- to 4-Hz filtering	TCN (900k)	HR clipping	Dalia	4.88±3.23
LSTM [42]	PPG + ACC	N/A	LSTM	Filtering	Dalia	7.44±3.26
KID** [23]	PPG + ACC	Adaptive linear filtering	CNN + Attention	N/A	Dalia	3.79
WildPPG	PPG	0.5- to 4-Hz filtering	1D ResNet	N/A	WildPPG	<u>8.62</u> ±0.06
Ours (2025)	PPG	0.5- to 4-Hz filtering z-score normalization	Encoder-Decoder + (430k) HR variations	N/A	SPC15-12 SPC15-22 Dalia WildPPG WESAD	2.24 ±1.03 (<u>+33%</u>) 4.30 ±2.97 (<u>+22%</u>) 4.16±1.74 (<u>-10%</u>) 7.73 ±0.05 (<u>+12%</u>) 4.75 ±2.25 (<u>+45%</u>)

* Dataset-specific threshold optimization is performed, requiring prior knowledge of HR distribution and activity intensity.

** Increasing dataset size with generating synthetic PPG waveforms to form the high HR signals using speed up [43].

signals, and ECG sampled at 125 Hz, with sensors worn on the wrist. Ground truth ECG was recorded via a chest sensor.

Dalia: Recorded from 15 participants, with each session lasting 2 hours [16]. ECG and PPG signals were collected during daily activities such as sitting, walking, and cycling.

WildPPG: Contains 13.5 hours of recordings from 16 participants during outdoor activities [46]. It captures heart rate data under challenging conditions with extreme noisy environments, including temperature variations.

WESAD: Includes 15 participants (12 males, 3 females, mean age 27.5) [47], with over one hour of recordings during tasks like solving arithmetic problems and watching videos.

B. Evaluation

Our evaluation uses leave-one-session-out (LOSO) cross-validation (CV). In LOSO, each test session is excluded from training, allowing assessment under realistic conditions with subject variations. Then, a 3-fold CV is used on the training set to select the model with the lowest validation error. During testing, we generated the ξ using model predictions without any information about the test subject. For the first four segments, HR was estimated using Fourier transformation by selecting the harmonic with the highest spectral energy. We evaluated performance using the most common metric in prior work, the mean absolute error (MAE) [16].

C. Training Architecture

The designed architecture takes as input only the pre-processed PPG and no other patient- or PPG-related features. We used the Adam optimizer [48] with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and a mini-batch size of 32. The learning rate was initialized to $5e-4$ with a weight decay of $1e-6$ and reduced by 0.1 when the validation loss stopped improving for 10 consecutive epochs. The training continues until 30 successive epochs without validation performance improvements. The best model is chosen as the lowest mean absolute error on the validation.

VI. EXPERIMENTAL RESULTS AND ANALYSIS

We evaluated our method with fixed architecture and hyperparameters on four datasets. Table I presents the results, along with the percentage performance gain of our method compared to the best learning-based solution for each dataset. We compared with the learning-based algorithms because traditional rule-based methods have optimized their parameters and thresholds according to the dataset statistics, which require prior knowledge of HR behavior with activities.

Our method outperforms all related works while using fewer sensor modalities. Many published approaches rely on additional inputs, such as power spectra, raw signals, or their combination with multiple sensor modalities, which makes their models more complex. Yet, their performance remains lower than ours, showing that our method is more effective for real-world settings where data is heavily contaminated with noise. We also examined model performance across demographic and activity-related factors. While no clear differences were observed across age or sex groups, our HR-integrated architecture showed greater improvements during high-motion activities. In the Dalia, which includes activity labels, our method reduced error by $\approx 10\text{--}15\%$ in *stairs* and *table soccer* segments, where error is typically high [16].

Table I shows that the performance of related works, which evaluated their method on multiple datasets, vary significantly between datasets since the Dalia and SPC15-22 datasets are seen as the most challenging to estimate HR due to including a wide range of activities performed under close to real-life conditions [16]. While these datasets are challenging due to low signal quality, our method reduces overall error and either outperforms or closely matches prior methods (e.g., second-best on Dalia). The results indicate that the performance of models should be evaluated across datasets with the same hyperparameters to investigate if the models generalize and perform well enough under numerous types of activities with different noise levels. Otherwise, the architectures and hyperparameters (post-processing with threshold values) can be optimized for a dataset while performing poorly for the rest.

Overall, our method improves HR estimation by 30–45% compared to prior solutions. This shows that learning HR variations together with volumetric flow changes help the model to accurately estimate heart rate under real-life settings.

A. Ablation Study

To investigate the effectiveness of our method, we perform ablations by removing each of the components individually.

TABLE II: Ablation studies across datasets. Top: model improvements with HR dynamics and augmentation. Bottom: comparison of different ξ lengths for past HR inputs.

Method / Datasets	SPC15-12	SPC15-22	Dalia	WESAD	WildPPG
<i>Model Variants</i>					
Traditional, $p(y_t \mathbf{x}_t)$	7.80 \pm 4.71	11.11 \pm 6.21	8.10 \pm 5.45	6.13 \pm 3.52	8.62 \pm 0.06
+ HR variations	5.85 \pm 3.31	6.50 \pm 5.15	6.12 \pm 3.20	4.88 \pm 2.45	7.92 \pm 0.02
+ Data Aug. (Proposed)	2.24 \pm 1.03	4.30 \pm 2.97	4.16 \pm 1.74	4.75 \pm 2.25	7.73 \pm 0.05
<i>Effect of ξ Length (Past HR Window)</i>					
$\xi_3 = y_{t-1}, \dots, y_{t-3}$	3.37 \pm 1.69	5.23 \pm 3.14	5.68 \pm 2.65	5.34 \pm 2.57	8.28 \pm 0.10
$\xi_5 = y_{t-1}, \dots, y_{t-5}$	2.24 \pm 1.03	4.30 \pm 2.97	4.16 \pm 1.74	4.75 \pm 2.25	7.73 \pm 0.05
$\xi_7 = y_{t-1}, \dots, y_{t-7}$	3.35 \pm 1.13	5.19 \pm 3.11	5.62 \pm 2.28	5.14 \pm 2.35	7.79 \pm 0.17

For the *Traditional* case, we only feed the signal similar to previous solutions. For the second part, we used the complete model, which learns HR variations, without augmentations.

Table II shows that incorporating our method improves estimation. Notably, augmentation reduces error for the smaller datasets, SPC15 pairs, more than in the bigger dataset, Dalia. This evidence supports our motivation for introducing HR variations to the model. Interestingly, increasing temporal dependencies with longer ξ reduces performance, despite providing more information for learning non-linear HR variations.

In other words, the model performs better when past HRs are included to a certain extent. This is an interesting finding, revealing that incorporating past information into the model is effective up to a limit. And, inputting all data into the model in hopes of uncovering relationships decreases the performance. To further understand the model's temporal behavior, we also analyzed its performance during sudden HR changes. In the Dalia dataset, the largest HR jump between consecutive segments was 16.46 bpm. For such high-change intervals, the model's average error is 4.52 bpm—around 10% higher than the dataset average. This suggests that while past information is helpful, handling abrupt HR transitions may require modifications, such as dynamic modeling or specialized loss terms.

Additionally, we performed ablations whether learning HR dynamics is architecture-agnostic and improves performance across models with minimal changes. Specifically, we used a two-layer MLP to extract features from past HR values and concatenated them with PPG features before the final prediction layer. Results in Table III show that our method improves performance across all models. We, also, show the Bland-Altman plot with correlation for SPC and Dalia in Figure 5 to investigate how learning the HR variability affects estimations. When the model takes only the current segment as input, it produces large errors. However, if the model learns the HR variations besides the current segment, the error decreases.

VII. DISCUSSION AND LIMITATIONS

We propose a HR estimation algorithm that leverages temporal dynamics to enhance cardiac monitoring with wearable devices, particularly in real-life conditions where noise

TABLE III: Architecture ablations on WildPPG. Values in parentheses indicate baselines without learning HR dynamics.

Dataset / Models	CNN [16]	LSTM [27]	CNN+Attention [23]	Ours
WildPPG	7.80 (8.87)	7.98 (9.03)	7.59 (8.53)	7.73 (8.62)

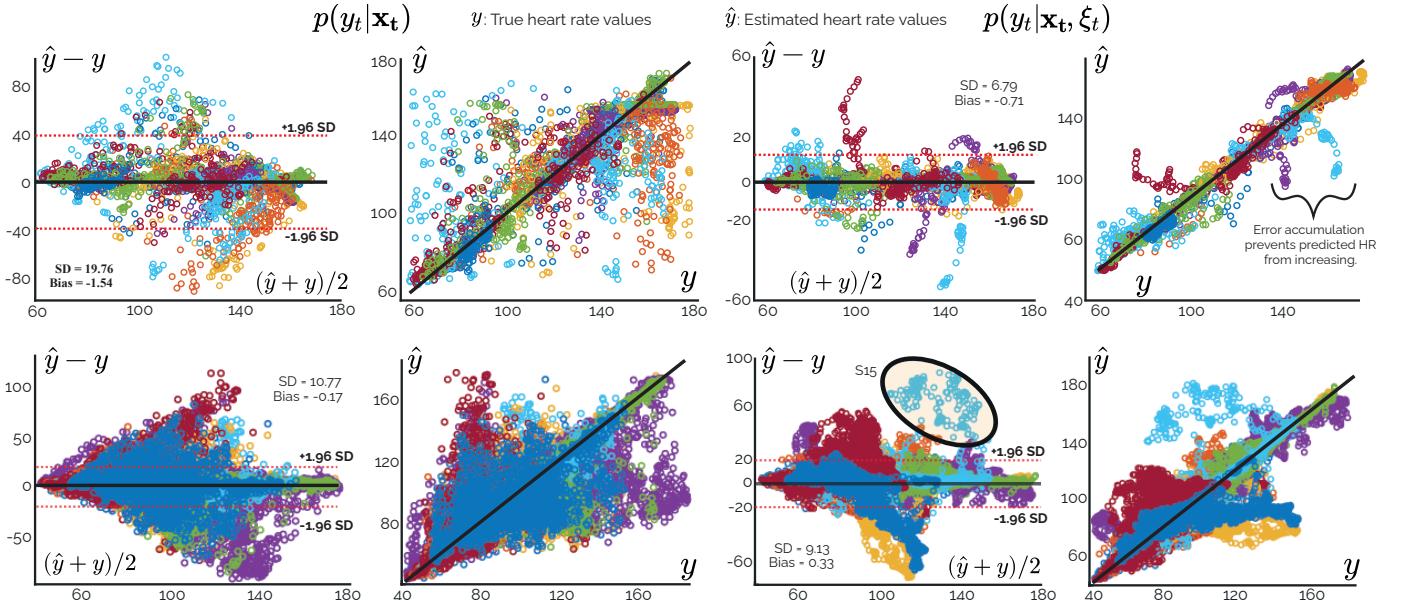


Fig. 5: Bland-Altman and correlation plots. Top row: SPC15-22, bottom row: Dalia. Left column: without learning the heart rate variations, right column: learning variations with ξ vector for the HR estimation. Colors represent subjects.

is prevalent. Our method surpasses current state-of-the-art approaches by reducing estimation errors by up to 40%, demonstrating improvements in robustness and accuracy.

a) Comparison with previous methods: Unlike traditional methods that rely heavily on heuristic-based signal processing techniques, our approach integrates non-linear temporal dynamics. While previous works [16] focused on hand-crafted features or trial-and-error optimization in deep learning models, our method models the chaotic behavior of cardiovascular dynamics, uncovering temporal patterns. This leads to more stable HR estimations, even under motion-corrupted conditions where conventional algorithms fail. Moreover, our method reduces dependency on multiple sensing modalities, outperforming multi-sensor approaches while using fewer modalities. While our method ranks second on the Dalia, trailing KID—uses adaptive filtering with inertial data and attention-based models—by 10% [23]), our approach is complementary. In addition, our model is relatively lightweight and avoids reliance on inertial signals or attention mechanisms, making it more suitable for deployment on memory- and energy-constrained wearable devices, which is an important design factor for HR estimation [49]. In future work, combining our dynamics module with other methods, including data augmentation [50], could further improve the performance.

b) Clinical Relevance: Accurate HR estimation is critical in ambulatory settings, where continuous monitoring is essential for managing cardiovascular diseases [5], stress detection, and fitness tracking [9]. By minimizing errors and reducing reliance on sensors, our method offers a solution for wearable devices. The ability to maintain high accuracy in dynamic environments enhances the reliability of HR monitoring for early detection of abnormal heart rate patterns, contributing to preventive healthcare.

c) Limitations and Future Work: Despite its strengths, our approach has limitations. First, it relies on accurate initial HR

values to predict subsequent readings, making it sensitive to noise during the initial segments. In noisy scenarios, heuristic methods like Fourier transforms may struggle to extract reliable initial HR, potentially affecting the performance.

Second, our model requires tuning of the temporal window size to effectively utilize historical data. While we identified an optimal range, incorporating excessive historical information can degrade performance and increase computational complexity. Future research can focus on adaptive mechanisms that adjust window sizes based on real-time signal quality or detecting and handling the motion artifacts [51].

Lastly, while our method shows strong performance on datasets, further validation in clinical settings with patients is necessary to confirm its generalizability. Testing on datasets that include abnormal heart rate would help evaluate the model's sensitivity to irregular and sudden changes in HR.

VIII. CONCLUSION

Our work is the first to study the behavior of HR variations through mutual information and have introduced an approach for improving HR estimations. Our method overcomes previous limitations by learning the HR dynamics together with blood volume changes to prevent large errors made by previously proposed learning methods. As a result, our method works without requiring post-processing or optimizing hyperparameters, e.g., threshold values. These findings underscore the efficacy of incorporating past information into the model, up to a certain threshold, while random inclusion of historical data decreases the performance while increasing computations. To the best of our knowledge, our method achieves the best performance, decreasing the error rates up to 40%, on HR estimation compared to the existing solutions.

REFERENCES

[1] A. Goldberger, "Is the normal heartbeat chaotic or homeostatic?" *Physiology*, 1991.

[2] F. Shaffer and J. P. Ginsberg, "An overview of heart rate variability metrics and norms," *Frontiers in Public Health*, vol. 5, 2017.

[3] T. F. of the European Society of Cardiology the North American Society of Pacing Electrophysiology, "Heart rate variability," *Circulation*, vol. 93, no. 5, pp. 1043–1065, 1996.

[4] K. Mandsager et al., "Association of Cardiorespiratory Fitness With Long-term Mortality Among Adults Undergoing Exercise Treadmill Testing," *JAMA Network Open*, 10 2018.

[5] L. Sandvik et al., "Heart rate increase and maximal heart rate during exercise as predictors of cardiovascular mortality: a 16-year follow-up study of 1960 healthy men," *Coronary Artery Disease*, 1995.

[6] M. Azram et al., "Clinical validation and evaluation of a novel six-lead handheld electrocardiogram recorder compared to the 12-lead electrocardiogram in unselected cardiology patients (EVALECG Cardio)," *European Heart Journal - Digital Health*, 2021.

[7] G. M. Marcus, "The Apple Watch can detect atrial fibrillation: so what now?" *Nature Reviews Cardiology*, vol. 17, Mar. 2020.

[8] S. A. Lubitz, A. Z. Faranesh, C. Selvaggi, S. J. Atlas, D. D. McManus, D. E. Singer, S. Pagoto, M. V. McConnell, A. Pantelopoulos, and A. S. Foulkes, "Detection of atrial fibrillation in a large population using wearable devices: The fitbit heart study," *Circulation*, 2022.

[9] F. Sarhaddi et al., "A comprehensive accuracy assessment of samsung smartwatch heart rate and heart rate variability," *PLOS ONE*, 2022.

[10] S. Benedetto and et al., "Assessment of the fitbit charge 2 for monitoring heart rate," *PLOS ONE*, 2018.

[11] M. Nissen and et al., "Heart rate measurement accuracy of fitbit charge 4 and samsung galaxy watch active2: Device evaluation study," *JMIR Form Res*, 2022.

[12] S. Suman M and et al., "Photoplethysmogram signal analysis for the heart rate estimation using shannon entropy," in *21st IEEE CSPA*, 2025.

[13] A. Khushhal and et al., "Validity and Reliability of the Apple Watch for Measuring Heart Rate During Exercise," *Sports Medicine International Open*, Oct. 2017.

[14] M. P. Wallen and et al., "Accuracy of heart rate watches: Implications for weight management," *PLOS ONE*, 2016.

[15] "ANSI/AAMI EC13-2002 - Cardiac Monitors, Heart Rate Meters, and Alarms." [Online]. Available: <https://webstore.ansi.org/standards/aami/ansiaamiec132002>

[16] A. Reiss and et al., "Deep ppg: Large-scale heart rate estimation with convolutional neural networks," *Sensors*, 2019.

[17] T. Schäck and et al., "Computationally efficient heart rate estimation during physical exercise using photoplethysmographic signals," in *2017 25th European Signal Processing Conference*, 2017.

[18] S. M. A. Salehizadeh and et al., "A novel time-varying spectral filtering algorithm for reconstruction of motion artifact corrupted heart rate signals during intense physical activities using a wearable photoplethysmogram sensor," *Sensors*, 2016.

[19] Q. Xie and et al., "Combining adaptive filter and phase vocoder for heart rate monitoring using photoplethysmography during physical exercise," in *EMBC*, 2018.

[20] J. Guo and et al., "An effective photoplethysmography heart rate estimation framework integrating two-level denoising method and heart rate tracking algorithm guided by finite state machine," *IEEE Journal of Biomedical and Health Informatics*, 2022.

[21] S. S. Chowdhury and et al., "Real-time robust heart rate estimation from wrist-type ppg signals using multiple reference adaptive noise cancellation," *IEEE Journal of Biomedical and Health Informatics*, 2018.

[22] H. Chung, H. Lee, and J. Lee, "Finite state machine framework for instantaneous heart rate validation using wearable photoplethysmography during intensive exercise," *IEEE Journal of Biomedical and Health Informatics*, 2019.

[23] C. Kechris, J. Dan, J. Miranda, and D. Atienza, "Kid-ppg: Knowledge informed deep learning for extracting heart rate from a smartwatch," *IEEE Transactions on Biomedical Engineering*, 2025.

[24] Y. Zhong and et al., "Sensor Fault Detection and Diagnosis for an Unmanned Quadrotor Helicopter," *Journal of Intelligent & Robotic Systems*, 2019.

[25] Y. Zhong and et al., "Sensor fault diagnosis for unmanned quadrotor helicopter via adaptive two-stage extended kalman filter," in *International Conference on Sensing, Diagnostics, Prognostics, and Control*, 2017.

[26] A. Shyam and et al., "Ppgnet: Deep network for device independent heart rate estimation from photoplethysmogram," in *41st EMBC*, 2019.

[27] L. G. Rocha and et al., "Lstm-only model for low-complexity hr estimation from wrist ppg," in *2021 EMBC*, 2021.

[28] L. G. Rocha and et al, "Binary cornet: Accelerator for hr estimation from wrist-ppg," *IEEE Transactions on Biomedical Circuits and Systems*, 2020.

[29] A. Burrello, D. J. Pagliari, P. M. Rapa, M. Semilia, M. Risso, T. Polonelli, M. Poncino, L. Benini, and S. Benatti, "Embedding temporal convolutional networks for energy-efficient ppg-based heart rate monitoring," *ACM Trans. Comput. Healthcare*, 2022.

[30] P. Kasnesis, L. Toumanidis, A. Burrello, C. Chatzigeorgiou, and C. Z. Patrikakis, "Feature-level cross-attentional ppg and motion signal fusion for heart rate estimation," in *2023 IEEE 47th Annual Computers, Software, and Applications Conference (COMPSAC)*, 2023.

[31] M. Zhou and N. Selvaraj, "Heart rate monitoring using sparse spectral curve tracing," in *2020 EMBC*, 2020.

[32] A. Temko, "Accurate heart rate monitoring during physical exercises using ppg," *IEEE Transactions on Biomedical Engineering*, 2017.

[33] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.

[34] J. B. Kinney and et al., "Equitability, mutual information, and the maximal information coefficient," *Proceedings of the National Academy of Sciences*, 2014.

[35] P. Sarkar and A. Etemad, "Cardiogan: Attentive generative adversarial network with dual discriminators for synthesis of ecg from ppg," *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021.

[36] K. Hlaváčková-Schindler, M. Paluš, M. Vejmelka, and J. Bhattacharya, "Causality detection based on information-theoretic approaches in time series analysis," *Physics Reports*, 2007.

[37] H. Zhu, C. Balsells-Rodas, and Y. Li, "Markovian gaussian process variational autoencoders," in *Proceedings of the 40th International Conference on Machine Learning*, ser. ICML'23. JMLR.org, 2023.

[38] H. Wu, A. Mardt, L. Pasquali, and F. Noe, "Deep generative markov state models," in *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, ser. NIPS'18, 2018.

[39] S. Ioffe and et al., "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proceedings of the 32nd International Conference on Machine Learning*. PMLR, 2015.

[40] Z. Zhang and et al, "Troika: A general framework for heart rate monitoring using wrist-type photoplethysmographic signals during intensive physical exercise," *IEEE Transactions on Biomedical Engineering*, 2015.

[41] D. Biswas and et al., "Cornet: Deep learning framework for ppg-based heart rate estimation and biometric identification in ambulant environment," *IEEE Transactions on Biomedical Circuits and Systems*, 2019.

[42] M. Wójcikowski, "Real-time ppg signal conditioning with long short-term memory (lstm) network for wearable devices," *Sensors*, 2022.

[43] T. Ko, V. Peddinti, D. Povey, and S. Khudanpur, "Audio augmentation for speech recognition," in *Interspeech 2015*, 2015, pp. 3586–3589.

[44] Z. Zhang, "Photoplethysmography-based heart rate monitoring in physical activities via joint sparse spectrum reconstruction," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 8, pp. 1902–1910, 2015.

[45] "IEEE Signal Processing Cup 2015," Dec. 2020. [Online]. Available: <https://signalprocessingsociety.org/community-involvement/ieee-signal-processing-cup-2015>

[46] M. Meier and et al., "WildPPG: A real-world PPG dataset of long continuous recordings," in *The Thirty-eight Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2024.

[47] P. Schmidt and et al., "Introducing wesad, a multimodal dataset for wearable stress and affect detection," in *Proceedings of the 20th ACM International Conference on Multimodal Interaction*, 2018.

[48] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2017.

[49] A. Burrello, D. J. Pagliari, M. Bianco, E. Macii, L. Benini, M. Poncino, and S. Benatti, "Improving ppg-based heart-rate monitoring with synthetically generated data," in *2022 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, 2022, pp. 153–157.

[50] L. Benfenati, S. Belloni, A. Burrello, P. Kasnesis, X. Wang, L. Benini, M. Poncino, E. Macii, and D. J. Pagliari, "Enhanceppg: Improving ppg-based heart rate estimation with self-supervision and augmentation," *CoRR*, vol. abs/2412.17860, 2024.

[51] Z. Guo, C. Ding, X. Hu, and C. Rudin, "A supervised machine learning semantic segmentation approach for detecting artifacts in plethysmography signals from wearables," *Physiological Measurement*, dec 2021.