Learning from Bilateral PPG via Cross-Site Attention and Knowledge Distillation

Anonymous Authors[1]

Abstract

Accurate heart rate estimation from wrist-worn photoplethysmography (PPG) during sports is challenging due to motion artifacts and asymmetric signal quality across wrists. Existing methods typically rely on single-wrist sensing and hand-crafted quality measures, overlooking complementary bilateral information. We propose an end-to-end bilateral PPG framework that jointly models both wrists using dual PapaGei encoders, SQI-conditioned temporal gating, and cross-wrist self-attention to emphasize cleaner signals. Evaluated on a bilateral sports dataset spanning six activities and 30 subjects with ECG ground truth, the model achieves $2.85\,\mathrm{bpm}$ MAE and $r = 0.96$, outperforming unilateral baselines. To improve deployability, we further introduce a knowledge distillation framework that transfers bilateral motion-compensation priors into a single-wrist EfficientNet-1D student. The distilled student achieves $2.79\,\mathrm{bpm}$ and $3.72\,\mathrm{bpm}$ MAE on the non-dominant and dominant wrist, respectively, surpassing corresponding unilateral baselines while requiring only one sensor at inference. These results establish bilateral PPG as an effective training paradigm for robust single-wrist heart rate estimation under motion.

1. Introduction

Accurate heart rate (HR) monitoring during physical activity is important for sports performance assessment and cardiovascular health. Wrist-worn photoplethysmography (PPG) is widely used due to its low cost and continuous sensing capability, but its accuracy degrades under motion, where movement artifacts overlap with the cardiac signal (Charlton et al., 2022).

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

This challenge is amplified by asymmetry between the dominant and non-dominant wrists. In many sports, the dominant wrist experiences stronger motion artifacts, while the non-dominant wrist is often more stable. However, signal quality remains activity- and time-dependent: both wrists may be corrupted simultaneously, and the cleaner signal can shift over time due to movement, contact pressure, or sensor placement. This creates complementary, time-varying information that single-wrist systems cannot exploit. At the same time, requiring two synchronized sensors is impractical for everyday wearable use, motivating a two-stage strategy: learn from bilateral sensing during training, then transfer that knowledge to a single-sensor model.

We therefore propose a learnable bilateral PPG framework for motion-robust HR estimation, followed by a knowledge distillation pipeline for practical deployment. The bilateral model jointly processes synchronized wrist signals using (i) dual PapaGei encoders (Pillai et al., 2025), (ii) SQI-conditioned temporal gating, (iii) cross-wrist self-attention fusion, and (iv) a regression head. A lightweight EfficientNet-1D student is then trained to inherit the bilateral model's motion-compensation priors while using only one wrist at inference.

Evaluated on a bilateral sports dataset spanning six activities and 30 subjects with ECG ground truth, the bilateral model achieves $2.85\,\mathrm{bpm}$ MAE and $r = 0.96$. The distilled single-wrist student achieves $2.79\,\mathrm{bpm}$ and $3.72\,\mathrm{bpm}$ MAE on the non-dominant and dominant wrist, respectively, substantially outperforming unilateral baselines and showing that bilateral learning can improve practical single-wrist HR estimation under motion.

2. Related Work

Bilateral PPG sensing has been studied mainly in clinical settings, including peripheral arterial disease detection, vascular asymmetry assessment, pulse transit time analysis, and cuffless blood pressure estimation (Wu et al., 2015; Du & Stephanus, 2016; Chen et al., 2024; Chan et al., 2019).

Multi-channel and multi-site PPG have also been explored for robust heart rate estimation. Prior work includes adaptive motion artifact suppression and frequency tracking for wrist PPG (Fallet & Vesin, 2017), spectral fusion of concurrent

PPG channels during exercise (Schäck et al., 2017), and template-based multi-site fusion that dynamically weights signals from different body locations (Meier & Holz, 2024). However, these methods largely rely on hand-crafted quality metrics or require all sensing sites at inference.

Knowledge distillation (Hinton et al., 2015) is widely used to compress large models for efficient deployment. We extend this idea to bilateral sensing, where a teacher model learns from complementary dual-sensor inputs during training, and a single-sensor student inherits these motion-compensation priors at inference. Unlike standard compression, the student must approximate decisions learned from richer multi-site inputs using only one sensor.

3. Method

Motivated by the complementary and time-varying signal quality across wrists under motion, we propose a bilateral PPG framework that jointly models synchronized signals and leverages this supervision to train a robust single-wrist estimator via knowledge distillation. The following sections describe the dataset, preprocessing pipeline, model architecture, and training procedure.

3.1. Dataset

First, we utilize the PPG Sports dataset (Anonymous, 2026), a bilateral PPG dataset with ECG ground truth for motion-robust heart rate estimation under realistic sports conditions. The dataset provides PPG signals from both wrists of 30 healthy participants across six activities: stationary, walking, running, badminton, table tennis, and basketball, each performed for 8 minutes in indoor and outdoor settings, yielding approximately 48 hours of data. It captures key real-world challenges, including non-periodic motion, high-intensity wrist dynamics, and bilateral asymmetry between dominant and non-dominant wrists.

3.2. Signal Preprocessing

All PPG signals are processed using a unified pipeline prior to segmentation to ensure consistency across recordings. Specifically, signals are first resampled to 125 Hz, then clipped at 5σ to suppress transient outliers. A fourth-order Butterworth bandpass filter ($[0.5, 8.0]$ Hz) is subsequently applied to isolate the physiological frequency range, followed by second-order polynomial detrending to remove slow baseline drift. Each segment is finally min–max normalized to $[0, 1]$. Ground-truth heart rate labels are derived from the corresponding chest ECG signals. These are processed using a bandpass filter ($[0.1, 30.0]$ Hz), after which R-peaks are detected using NeuroKit2 (Makowski et al., 2021) to compute heart rate. Here, the signals are segmented into 10-second windows with an 8-second step. Finally, to ensure subject-independent evaluation and prevent data leakage, the dataset is split by subject into training (S01–S21), validation (S22–S25), and test (S26–S30) sets.

3.3. Signal Quality Index Features

Despite the availability of bilateral signals, the quality of each wrist remains highly variable under intense motion. To account for this time-varying reliability, we also prepare Signal Quality Index (SQI) features that guide the model to emphasize cleaner regions while suppressing degraded ones (Schmith et al., 2023). For each PPG segment, SQI features are computed over 1.5-second sub-windows with a 0.5-second stride. We extract two complementary measures: (i) **Skewness**, which captures waveform asymmetry and distortion, and (ii) **PSD ratio**, defined as the fraction of spectral power within the cardiac band $[0.66, 3.4]$ Hz relative to total power estimated via Welch's method. These form a temporal quality matrix $\mathbf{S} \in \mathbb{R}^{T_{\text{win}} \times 2}$.

Among common SQI candidates, skewness and PSD ratio have shown strong performance for assessing PPG quality (Elgendi, 2016; Beh et al., 2023). Based on prior literature and SHAP analysis, these two features were selected as the most informative inputs to the gating module (see Appendix 5) (Lundberg & Lee, 2017).

3.4. Bilateral Teacher Architecture

With the dataset prepared and preprocessed, we propose a bilateral architecture that jointly processes synchronized PPG signals from the dominant and non-dominant wrists (Figure 1). The design is motivated by the observation that motion artifacts affect the two wrists asymmetrically, creating complementary signal quality over time. To leverage this, the model first extracts wrist-specific representations, then suppresses low-quality segments, and finally fuses the two streams to produce a robust heart rate estimate.

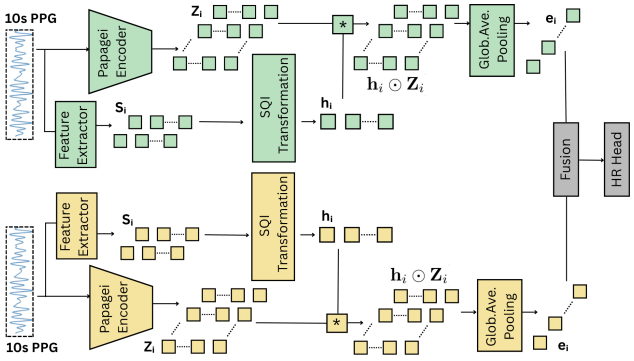


Figure 1. Overview of the proposed bilateral PPG-based heart rate estimation framework.

3.4.1. Dual Encoder Backbone

Each wrist signal $\mathbf{x} \in \mathbb{R}^{1\times T}$ is independently encoded using a PapaGei backbone (Pillai et al., 2025), where T denotes the number of data points in a 10-second segment. The two encoders share the same architecture but maintain separate weights, enabling wrist-specific specialization. We adopt the PapaGei-S variant due to its strong empirical performance and its morphology-aware pretraining, which captures physiologically relevant waveform characteristics.

3.4.2. SQI-Conditioned Temporal Gating

Given the encoder outputs, we next incorporate signal quality information to suppress unreliable regions. For each wrist, the SQI feature matrix $\mathbf{S} \in \mathbb{R}^{T_{\text{win}}\times 2}$, where T_{win} is the number of subwindows, which is processed by a lightweight transformation network to produce a temporal gate $\mathbf{h} \in \mathbb{R}^{T_{\text{res}}\times 1}$, where T_{res} denotes the temporal resolution of the final backbone layer. The network maps SQI features to scalar quality scores, which are interpolated to match the encoder resolution, smoothed via a 1D convolution, and normalized with a softmax over time.

The gated embedding is computed as:

$$\mathbf{e}_i = \big(\mathbf{h}_i \odot \mathbf{Z}_i\big).\text{mean}(-1), \quad i \in \{\text{dom}, \text{non-dom}\}, \quad (1)$$

where $\mathbf{Z}_i$ denotes the encoder features. This step enables quality-aware temporal aggregation by down-weighting artifact-corrupted segments.

3.4.3. Cross-Wrist Attention Fusion

Finally, the two gated embeddings are fused to exploit their complementarity. We form a two-token sequence from $\mathbf{e}_{\text{dom}}$ and $\mathbf{e}_{\text{non-dom}}$ and process it using a Transformer encoder with multi-head self-attention. This allows the model to capture interactions between the two streams and dynamically emphasize the more reliable signal. The fused representation is obtained via average pooling and passed through a linear layer to predict the heart rate.

The model is trained end-to-end using the Huber loss to improve robustness to outliers caused by motion artifacts. We use the Adam optimizer with a learning rate of 10^{-3} and weight decay of 10^{-4}, and apply cosine annealing over 200 epochs with a batch size of 32. All activities are jointly used during training, and the best model is selected based on validation performance to ensure optimal generalization and avoid overfitting.

3.5. Knowledge Distillation to Single-Wrist Student

While the bilateral teacher achieves strong accuracy, deploying two synchronized wrist sensors is impractical for general consumer use. We address this through a knowledge distillation framework that transfers the bilateral model's learned motion-compensation priors into a lightweight single-wrist student, enabling accurate HR estimation with standard single-sensor hardware.

3.5.1. Student Architecture

The student architecture is an EfficientNet-1D (Tan & Le, 2019) model comprising a 7-stage MBConv backbone with depthwise-separable convolutions, Squeeze-and-Excitation (SE) channel attention, and a linear HR regression head. This architecture is selected for its compound scaling of depth, width, and resolution, which together deliver strong representational capacity at low computational cost. At $\sim$4.8M parameters, which is 64% smaller than the bilateral teacher ($\sim$13.7M), the EfficientNet-1D student achieves faster inference and lower memory footprint, with further quantization or pruning needed for full on-device wearable deployment. The MBConv blocks use stochastic depth regularization to mitigate small-dataset overfitting. The backbone outputs a feature map $(B, 512)$. A hint projector $(512 \to 512)$ is appended during training to align student embeddings with the teacher's representation space; it is discarded at inference.

3.5.2. Four-Component KD Loss

The student is trained to simultaneously match the bilateral teacher's behaviour at four levels:

$$\begin{aligned}\mathcal{L} = \lambda_{\text{hard}}\,\mathcal{L}_{\text{Huber}}(\hat{y}_s, y) + \lambda_{\text{resp}}\,\mathcal{L}_{\text{resp}}(\hat{y}_s, \hat{y}_t) \\ + \lambda_{\text{hint}}\,\mathcal{L}_{\text{hint}}(\mathbf{e}_s, \mathbf{e}_t) + \lambda_{\text{KL}}\,\mathcal{L}_{\text{KL}}(p_s \| p_t)\end{aligned} \quad (2)$$

where $\hat{y}_s$ and $\hat{y}_t$ are student and teacher HR predictions, y is the ECG ground truth, $\mathbf{e}_s$ and $\mathbf{e}_t$ are the projected student and teacher embeddings, and p_s, p_t are soft Gaussian HR distributions centred on $\hat{y}_s$ and $\hat{y}_t$ respectively (temperature $\tau = 4.0$). The loss weights are empirically set to $\lambda_{\text{hard}} = \lambda_{\text{resp}} = 0.35$, $\lambda_{\text{hint}} = 0.20$, $\lambda_{\text{KL}} = 0.10$. Importantly, the hard loss grounds predictions on ECG labels and the response loss aligns the student's output with the teacher's bilateral-informed estimates. The hint loss aligns internal representations, transferring the teacher's motion-compensation features, and the KL loss enforces distributional consistency beyond point-estimate matching. Here, the teacher is frozen, and the student is optimized using AdamW (LR 3×10^{-4}, weight decay 10^{-4}).

3.6. Baselines

We compare the proposed bilateral model against twelve unilateral baselines built on two backbone architectures, PapaGei (Pillai et al., 2025), and PulsePPG (Saha et al., 2025), under two training regimes. In the *frozen* setting, pretrained backbone weights are fixed and only a linear regression

Table 1. Performance comparison across models and activities (MAE↓). Abbreviations: Frz = Frozen, FT = Fine-Tuned (Non-Frozen), ND = Non-dominant, Dom = Dominant.

Model	Stat.	Walk	Run	Badmin.	T.Tennis	Basket.	Comb.
Papagei-s (Frz, ND)	9.27	19.93	40.62	16.35	18.66	7.11	18.66
Papagei-s (Frz, Dom)	10.69	13.88	38.26	20.75	20.41	7.52	18.59
Papagei-s (FT, ND)	2.04	2.89	1.67	7.93	2.97	2.38	3.32
Papagei-s (FT, Dom)	2.62	3.98	5.05	10.65	5.63	6.81	5.86
Papagei-p (Frz, ND)	20.83	15.36	33.81	18.33	14.50	12.13	19.16
Papagei-p (Frz, Dom)	20.22	13.63	35.80	15.96	13.05	12.72	18.56
Papagei-p (FT, ND)	2.05	3.91	2.41	8.46	3.90	2.70	3.91
Papagei-p (FT, Dom)	1.89	4.50	4.80	10.33	5.67	6.32	5.60
PulsePPG (Frz, ND)	4.48	8.55	4.09	14.42	13.69	7.78	8.84
PulsePPG (Frz, Dom)	5.12	10.50	11.51	12.57	9.20	7.13	9.34
PulsePPG (FT, ND)	1.85	4.87	2.12	7.98	2.97	2.12	3.65
PulsePPG (FT, Dom)	1.88	2.78	6.43	9.20	5.93	5.73	5.32
ResNet18 (ND)	1.61	9.76	2.29	8.90	6.35	4.41	5.57
Efficientnet (ND)	1.52	4.08	1.77	8.19	1.90	2.85	3.39
Ours	**1.26**	**2.76**	**1.23**	7.99	**1.86**	**2.02**	**2.85**

head is trained. In the *non-frozen* setting, the full network is fine-tuned end-to-end. All unilateral baselines use either the dominant or non-dominant wrist alone. We additionally include ResNet18 (He et al., 2016) and EfficientNet (Tan & Le, 2019) baselines evaluated on the non-dominant wrist. All models use identical data splits, loss functions, and optimization settings.

4. Results

4.1. Comparison with Baselines

We begin by comparing the proposed method with a range of unilateral baselines. Table 1 reports MAE for all models across six activities and the combined test set. The bilateral model achieves the lowest combined MAE of 2.85 bpm with Pearson $r = 0.96$, outperforming all unilateral baselines across both backbone families and both training regimes. These findings suggest that bilateral modeling effectively captures complementary information and mitigates time-varying signal corruption that cannot be addressed by selecting a single wrist alone.

4.2. Ablation Study

To better understand the contribution of each component, we perform an ablation study on the combined test set and report the results in Table 2. For fair comparison, single-wrist variants use the same self-attention module as the bilateral model, isolating the effect of input modality. Removing SQI gating increases MAE from 2.85 to 3.20 bpm, confirming the benefit of signal quality conditioning. Using only the dominant or non-dominant wrist yields 3.35 bpm and 6.25 bpm MAE, respectively. Removing self-attention increases MAE to 3.52 bpm, worse than both single-wrist variants, showing that dual inputs alone are insufficient without structured cross-stream fusion. Overall, bilateral fusion, self-attention, and SQI gating provide complementary gains in robustness and accuracy in our framework.

Table 2. Ablation study on the combined test set (MAE↓).

Model Variant	MAE (bpm)
Without SQI	3.20
Non-dominant Only	3.35
Dominant Only	6.25
Without Attention	3.52
Ours	**2.85**

The single-wrist ablation variants share the bilateral architecture and training objective, and are therefore not directly comparable to the independently optimized unilateral baselines in Table 1. They are included only to quantify the marginal benefit of the second wrist input. Accordingly, the strongest unilateral baselines serve as references for the knowledge distillation results in Section 4.3.

4.3. Knowledge Distillation Results

Table 3 reports the per-activity and combined MAE for the distilled single-wrist students evaluated on the PPG sports dataset. The non-dominant student achieves a combined MAE of 2.79 bpm, outperforming its direct unilateral baseline of 3.35 bpm (16% improvement) and approaching the bilateral teacher's 2.85 bpm, despite operating with only one wrist at inference. The dominant student achieves 3.72 bpm, again substantially better than its unilateral baseline of 5.86 bpm (35.9% improvement). The dramatic improvement on the dominant wrist, where unilateral performance was severely degraded, directly demonstrates that bilateral training successfully transfers motion-compensation knowledge from the cleaner non-dominant stream into the dominant-wrist student.

Table 3. Knowledge distillation results on the PPG sports test set (MAE↓).

Model	Stat.	Walk	Run	Badmin.	T.Tennis	Basket.	Comb.
Unilateral ND (baseline)	2.04	2.89	1.67	7.93	2.97	2.38	3.32
Unilateral Dom (baseline)	2.62	3.98	5.05	10.65	5.63	6.81	5.86
KD Student (ND)	1.32	2.75	1.67	8.14	1.39	1.46	**2.79**
KD Student (Dom)	1.50	1.83	4.03	8.77	2.38	3.83	**3.72**

5. Conclusion

In this paper, we present a bilateral PPG framework for motion-robust heart rate estimation. By jointly modeling signals from both wrists with SQI-conditioned gating and cross-wrist attention, the model outperforms all unilateral baselines. We further show that bilateral knowledge can be distilled into a lightweight single-wrist model, achieving strong performance while requiring only one sensor at inference. These results highlight bilateral sensing as an effective training paradigm for robust single-wrist physiological monitoring.

References

Anonymous. Anonymous authors, "ppg sports dataset", 2026. Accepted, to appear. Details omitted for anonymous review.

Beh, W.-K., Yang, Y.-C., Lo, Y.-C., Lee, Y.-C., and Wu, A.-Y. Machine-aided ppg signal quality assessment (sqa) for multi-mode physiological signal monitoring. *ACM Trans. Comput. Healthcare*, 4(2), April 2023. doi: 10.1145/3587256. URL https://doi.org/10.1145/3587256.

Chan, G., Cooper, R., Hosanee, M., Welykholowa, K., Kyriacou, P. A., Zheng, D., Allen, J., Abbott, D., Lovell, N. H., Fletcher, R., and Elgendi, M. Multi-site photoplethysmography technology for blood pressure assessment: Challenges and recommendations. *Journal of Clinical Medicine*, 8(11), 2019. ISSN 2077-0383. doi: 10.3390/jcm8111827. URL https://www.mdpi.com/2077-0383/8/11/1827.

Charlton, P. H., Kyriacou, P. A., Mant, J., Marozas, V., Chowienczyk, P., and Alastruey, J. Wearable photoplethysmography for cardiovascular monitoring. *Proceedings of the IEEE*, 110(3):355–381, 2022. doi: 10.1109/JPROC.2022.3149785.

Chen, Y.-S., Lu, W.-A., Hsu, L.-Y., and Kuo, C.-D. Determinants of hand pulse wave velocity and hand pulse transit time in healthy adults. *Scientific Reports*, 14(1):10144, 2024. doi: 10.1038/s41598-024-60927-5.

Du, Y.-C. and Stephanus, A. A novel classification technique of arteriovenous fistula stenosis evaluation using bilateral ppg analysis. *Micromachines*, 7(9), 2016. ISSN 2072-666X. doi: 10.3390/mi7090147. URL https://www.mdpi.com/2072-666X/7/9/147.

Elgendi, M. Optimal signal quality index for photoplethysmogram signals. *Bioengineering*, 3(4):21, 2016. doi: 10.3390/bioengineering3040021.

Fallet, S. and Vesin, J.-M. Robust heart rate estimation using wrist-type photoplethysmographic signals during physical exercise: an approach based on adaptive filtering. *Physiological Measurement*, 38(2):155, jan 2017. doi: 10.1088/1361-6579/aa506e. URL https://doi.org/10.1088/1361-6579/aa506e.

He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

Hinton, G., Vinyals, O., and Dean, J. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015.

Lundberg, S. M. and Lee, S.-I. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30, 2017.

Makowski, D., Pham, T., Lau, Z. J., Brammer, J. C., Lespinasse, F., Pham, H., Schölzel, C., and Chen, S. H. A. NeuroKit2: A python toolbox for neurophysiological signal processing. *Behavior Research Methods*, 53(4):1689–1696, feb 2021. doi: 10.3758/s13428-020-01516-y. URL https://doi.org/10.3758%2Fs13428-020-01516-y.

Meier, M. and Holz, C. Robust heart rate detection via multi-site photoplethysmography. In *2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 1–4, 2024. doi: 10.1109/EMBC53108.2024.10782155.

Meier, M., Demirel, B. U., and Holz, C. Wildppg: a real-world ppg dataset of long continuous recordings. *Advances in Neural Information Processing Systems*, 37: 2246–2266, 2024.

Park, S., Zheng, D., and Lee, U. A ppg signal dataset collected in semi-naturalistic settings using galaxy watch. *Scientific Data*, 12(1):892, 2025. doi: 10.1038/s41597-025-05152-z. URL https://doi.org/10.1038/s41597-025-05152-z.

Pillai, A., Spathis, D., Kawsar, F., and Malekzadeh, M. PaPaGei: Open Foundation Models for Optical Physiological Signals. In *The Thirteenth International Conference on Learning Representations, ICLR 2025*, Singapore, April 2025. URL https://arxiv.org/abs/2410.20542. Accepted. arXiv preprint arXiv:2410.20542.

Saha, M., Xu, M. A., Mao, W., Neupane, S., Rehg, J. M., and Kumar, S. Pulse-ppg: An open-source field-trained ppg foundation model for wearable applications across lab and field settings. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 9 (3):1–35, 2025.

Schmith, J., Kelsch, C., Cunha, B. C., Prade, L. R., Martins, E. A., Keller, A. L., and de Figueiredo, R. M. Photoplethysmography signal quality assessment using attractor reconstruction analysis. *Biomedical Signal Processing and Control*, 86:105142, 2023. ISSN 1746-8094. doi: https://doi.org/10.1016/j.bspc.2023.105142. URL https://www.sciencedirect.com/science/article/pii/S174680942300575X.

Schäck, T., Muma, M., and Zoubir, A. Computationally efficient heart rate estimation during physical exercise using photoplethysmographic signals. pp. 2478–2481, 09 2017. doi: 10.23919/EUSIPCO.2017.8081656.

Tan, M. and Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pp. 6105–6114. PMLR, 2019.

Wu, J.-X., Lin, C.-H., Wu, M.-J., Li, C.-M., Lim, B.-Y., and Du, Y.-C. Bilateral photoplethysmography analysis for arteriovenous fistula dysfunction screening with fractional-order feature and cooperative game-based embedded detector. *Healthcare Technology Letters*, 2(3): 64–69, 2015. doi: 10.1049/htl.2014.0090.

Appendix

SHAP Analysis of SQI Gate Sub-Network

To further understand the contribution of different SQI features, we analyze the SQI gate using SHAP values. Figure 2 presents the feature importance across both wrists.

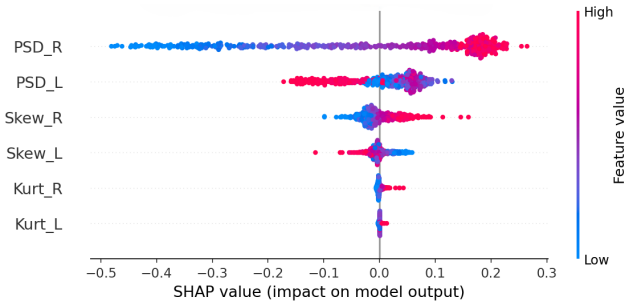
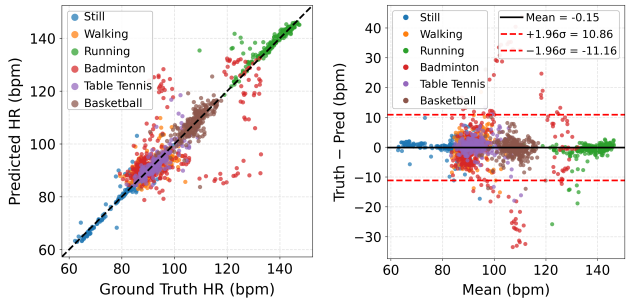


Figure 2. SHAP summary plot for the SQI Gate Sub-Network. Acronyms: PSD (Power Spectral Density), Skew (Skewness), Kurt (Kurtosis). Suffixes _R and _L denote the dominant and non-dominant wrists, respectively.

Correlation and Bland-Altman Analysis

We further assess agreement between predicted and ground-truth heart rate using correlation and Bland–Altman analyses. Figure 3 summarizes the results across all activities.



(a) Correlation analysis *(b)* Bland-Altman analysis

Figure 3. (a) Shows the combined predicted versus ground-truth HR scatter plot for the proposed bilateral Teacher model across all six activities. The predictions exhibit strong linear alignment with the ground truth across the full HR range, achieving a combined MAE of 2.85 bpm, Pearson $r = 0.959$, and $R^2 = 0.916$., (b) Shows the Bland-Altman plot. The mean bias is -0.15 bpm, indicating negligible systematic over- or under-estimation across the HR range.

Extended Evaluation Across Sensing Environments

We report subject-independent evaluations on two additional datasets to assess the generalisability of the proposed framework across different sensing environments and device configurations. In each case, the model is trained and evaluated exclusively within that dataset using held-out test subjects unseen during training.

WildPPG

While our primary focus is real bilateral wrist PPG, we further evaluate whether the proposed framework generalises beyond this setting under distribution shifts in sensor placement and environment. To this end, we use WildPPG (Meier et al., 2024), a large-scale dataset with synchronised PPG from the wrist, ankle, and head collected under unconstrained, free-living conditions, providing a realistic benchmark for robustness.

As WildPPG does not include bilateral wrist measurements, we approximate bilateral input via cross-location pairing, where signals from different anatomical sites are fed into the two encoder branches. Although this deviates from the original bilateral assumption, it allows us to assess whether the model can exploit complementary information across heterogeneous inputs. Table 4 reports the MAE for each configuration.

Single-location results show that wrist PPG is most affected by noise (8.42 bpm), while ankle (3.48 bpm) and head (5.44 bpm) provide stronger signals. Cross-location pairs consistently improve performance (ankle–wrist: 3.32 bpm, head–wrist: 4.43 bpm, head–ankle: 2.63 bpm), suggesting that the fusion module captures complementary signal characteristics beyond a fixed bilateral wrist configuration.

Table 4. Evaluation on WildPPG (MAE ↓)

Configuration	MAE (bpm)
Ankle Only	3.48
Head Only	5.44
Wrist Only	8.42
Ankle–Wrist Pair	**3.32**
Head–Wrist Pair	**4.43**
Head–Ankle Pair	**2.63**

The Head–Ankle pair achieves the lowest MAE of 2.63 bpm among all bilateral configurations, indicating that ankle and head signals provide the most complementary motion artifact profiles across the WildPPG activity spectrum. This pair is therefore selected as the teacher for knowledge distillation, as it offers the strongest bilateral supervision signal for training single-location students.

Table 5 reports the KD results on WildPPG. The ankle student achieves 3.19 bpm, an 8.3% improvement over its unilateral baseline (3.48 bpm), and the head student achieves 4.38 bpm, a 19.5% improvement over its unilateral baseline (5.44 bpm), demonstrating that bilateral teacher supervision effectively transfers complementary priors to single-location students.

Table 5. Knowledge distillation results on the WildPPG dataset (MAE↓).

Model	MAE (bpm)
Unilateral Ankel	3.48
Unilateral Head	5.44
KD: Ankel Student	**3.19**
KD: Head Student	**4.38**

GALAXYPPG

To further validate the proposed framework under real bilateral wrist conditions, we evaluate on a semi-naturalistic settings dataset comprising PPG recordings from both wrists simultaneously (Park et al., 2025). Unlike WildPPG, this dataset provides genuine bilateral wrist PPG: the Right wrist is recorded using a Galaxy Watch and the Left wrist using an Empatica E4 providing a direct evaluation of the bilateral assumption under semi-realistic motion conditions.

Table 6 reports the HR estimation MAE for unilateral and bilateral configurations. The Right wrist (Galaxy) achieves MAE of 5.24 bpm and the Left wrist (Empatica) achieves 6.56 bpm. The bilateral model achieves MAE of 4.71 bpm, outperforming both unilateral baselines, confirming that the SQI-conditioned temporal gating and self-attention fusion effectively exploit complementary signal characteristics from the two wrists even when recorded by heterogeneous devices operating at different sampling rates.

Table 6. Evaluation on Galaxy–Empatica bilateral wrist dataset (MAE↓)

Configuration	MAE (bpm)
Left Wrist (Empatica)	5.24
Right Wrist (Galaxy Watch)	6.56
Bilateral (Ours)	**4.71**

Table 7 reports the KD results on GalaxyPPG. The right wrist student achieves 6.36 bpm and the left wrist student achieves 5.06 bpm, corresponding to improvements of 3.05% and 3.44% respectively over their unilateral baselines, confirming that bilateral teacher supervision transfers effectively to single-wrist students even across heterogeneous devices operating at different sampling rates.

Table 7. Knowledge distillation results on the GalaxyPPG dataset (MAE↓).

Model	MAE (bpm)
Unilateral Left (Empatica)	5.24
Unilateral Right (Galaxy Watch)	6.56
KD: Left Student	**5.06**
KD: Right Student	**6.36**

The modest gains are primarily attributable to device heterogeneity between the Galaxy Watch and Empatica E4 — differences in sampling rate and optical design reduce bilateral complementarity in the teacher, limiting the knowledge available for transfer.