Wavelet-Based Masked Multiscale Reconstruction for PPG Foundation Models

Anonymous Author(s)
Affiliation
Address
email

Abstract

We introduce Masked Multiscale Reconstruction (MMR), a self-supervised pretraining framework for photoplethysmography (PPG) signals that leverages the discrete wavelet transform. MMR is pretrained on ∼18M unlabeled 10-second PPG segments collected from over ∼41K smartwatch users largely in naturalistic field settings. The pretraining task is defined to randomly mask out subsets of wavelet coefficients derived from multi-resolution decomposition of raw PPG signals and train the encoder to reconstruct them. This enables the model to capture patterns across scales from fine-grained waveform morphology to long-term temporal dynamics crucial for diverse downstream tasks. On 10 of 13 health-related tasks, MMR trained on large-scale wearable PPG data outperforms or matches state-of-the-art open-source PPG foundation models and other self-supervised baselines. An ablation study of wavelet design further underscores the value of wavelet-based representations, paving the way toward robust and generalizable PPG foundation models.

1 Introduction

Photoplethysmography (PPG) has emerged as a key sensing modality in wearables, powering applications from cuffless blood pressure estimation [Song et al., 2019], arrhythmia detection [Bashar et al., 2019] to stress monitoring [Namvari et al., 2022]. Its ubiquity in consumer devices creates an opportunity for large-scale, continuous monitoring of cardiovascular health and the development of digital biomarkers [Charlton et al., 2022a, Lee and Akamatsu, 2025]. Traditional PPG models relied on handcrafted features or small datasets [Shao et al., 2021, Han et al., 2020]; however, recent foundation models such as [Abbaspourazad et al., 2024a, Pillai et al., 2024, Saha et al., 2025] have leveraged vast amounts of unlabeled datasets, establishing a new paradigm for large-scale representation learning.

While seminal works such as [Abbaspourazad et al., 2024a] explore patient-wise contrastive learning and [Pillai et al., 2024] introduce morphology-awareness through proxies like sVRI binning and SQI regression, these approaches remain primarily rooted in the time domain. Time-only representations overlook the spectral structure of PPG, where physiological rhythms unfold across multiple frequency bands. Standard Fourier methods attempt to capture these patterns but impose a stationarity assumption, leading to poor localization and resolution in non-stationary signals [Mallat, 2002]. Explicitly modeling the spectral domain is therefore beneficial for capturing the hierarchical, multi-resolution structure of PPG—rich information that purely temporal features or proxy objectives may fail to represent [Chen et al., 2025].

Wavelet decomposition [Daubechies, 1992] provides a natural way to analyze non-stationary signals in the time–frequency domain. By adaptively trading time and frequency resolution, wavelets capture short-lived fine details at higher frequencies while preserving coarse, long-term dynamics at lower frequencies. This multi-resolution view is critical, as physiological signals carry information across

Submitted to 39th Conference on Neural Information Processing Systems (NeurIPS 2025). Do not distribute.

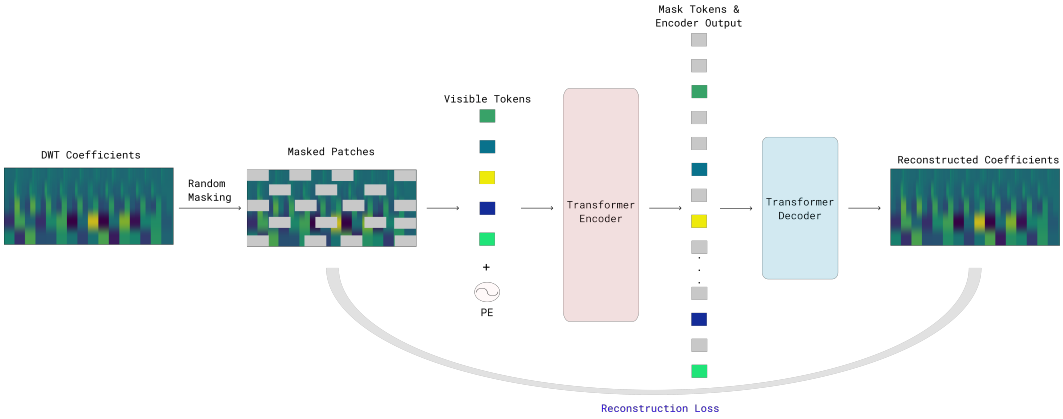

Figure 1: Masked Multiscale Reconstruction for Photoplethysmography (PPG) signals.

multiple scales, from local waveform morphology, which has been linked to vascular health [Charlton et al., 2022b], to longer-term rhythm dynamics, crucial for tasks such as heart rate variability [Namvari et al., 2022]. Motivated by these insights, we propose a masked multi-scale reconstruction framework for PPG, in which raw signals are decomposed into multiple resolution bands and a foundation model is trained to reconstruct masked coefficients across scales. To summarize, our contributions are: *(i)* We pretrain a large-scale *wavelet-based PPG foundation model* on ∼50K hours of PPG data with a masked multiscale reconstruction objective, enabling the model to capture rich time–frequency information across multiple scales. *(ii)* We demonstrate strong generalization across 13 diverse downstream tasks and provide detailed ablations that examine the impact of design choices such as wavelet family, decomposition scales, and patch size.

2 Background

Foundation models (FMs) for biosignals have recently shown strong promise, with large-scale self-supervised pretraining on ECG and PPG data from wearables demonstrating transferable representations across diverse downstream tasks [Abbaspourazad et al., 2024b, Pillai et al., 2024, Saha et al., 2025, Yang et al., 2023]. Yet most of these approaches treat signals purely in the time domain, overlooking spectral information that carries important physiological cues. A growing body of frequency-aware FMs [Zhang et al., 2022, Liu et al., 2023, Kara et al., 2024, Cheng et al., 2025, Fu and Hu, 2025, Duan et al., 2024] shows that explicitly modeling spectral content improves robustness and transferability. Wavelet analysis offers a natural way to capture both temporal localization and multi-resolution frequency structure, and has long been applied to PPG for denoising, feature extraction, and disease detection [Alafeef and Fraiwan, 2020, Singh et al., 2023, Shao et al., 2021]. Recent deep learning approaches incorporate wavelets in end-to-end pipelines, such as wavelet-informed tokenization [Masserano et al., 2024], and the combination of learnable wavelet decompositions with frequency-guided masking for biosignal foundation models [Chen et al., 2025]. Our work extends this line and introduces a multi-resolution masked pretraining framework for large-scale PPG data collected from smartwatches in real-world settings. By leveraging the fact that health tasks rely on information at multiple signal granularities, our approach provides more physiologically grounded and transferable representations (Refer Appendix A).

3 Method

PPG wavelet coefficients are patched, encoded, and reconstructed in a multi-scale framework for robust representation learning as shown in Fig. 1.

Discrete wavelet transform (DWT). The discrete wavelet transform (DWT) decomposes a signal into an approximation A_J and detail bands $\{D_j\}_{j=1}^{J}$ using paired low- and high-pass filters, with each level downsampled by half to provide joint time–frequency localization. At sampling rate f_s, the j-th level spans the frequency range $[f_s/2^{j+1}, f_s/2^j]$. We apply a level-4 Haar DWT (selected as

Table 1: Linear probing results across downstream tasks. Best scores are **bold**, second best underlined, with 95% CIs in gray brackets.

Classification - AUROC (↑)	**SimCLR**	**PaPaGei-P**	**PaPaGei-S**	**LSM**	**MMR**
Hypertension - Lab	57.93 [54.2 – 61.2]	**67.78 [64.7 – 71.2]**	56.69 [52.4 – 59.9]	54.74 [51.0 – 58.4]	**67.47 [63.9 – 70.9]**
Hypertension	**64.12 [57.4 – 71.1]**	62.10 [54.2 – 68.6]	61.46 [54.3 – 67.6]	54.28 [51.0 – 58.4]	60.69 [46.8 – 60.9]
PVC Detection	71.78 [71.1 – 72.5]	80.38 [79.7 – 80.9]	74.61 [73.9 – 75.3]	72.29 [71.5 – 72.9]	**82.47** [81.8 – 83.1]
HDL	41.12 [38.4 – 45.9]	49.71 [46.2 – 54.0]	33.43 [29.8 – 36.8]	56.53 [52.9 – 59.9]	**62.41** [58.1 – 66.7]
LDL	49.41 [46.5 – 52.5]	**64.30** [61.1 – 67.5]	50.94 [47.7 – 54.0]	56.51 [53.1 – 59.5]	61.50 [58.7 – 64.6]
Platelets	61.49 [59.0 – 63.9]	**74.31** [72.0 – 76.8]	62.14 [59.4 – 64.7]	56.30 [53.7 – 58.8]	66.18 [63.9 – 68.5]
Potassium	64.55 [62.2 – 66.3]	81.20 [79.6 – 82.8]	71.53 [69.4 – 73.6]	67.34 [65.3 – 69.5]	**82.85** [81.1 – 84.3]
Sodium	50.88 [46.7 – 54.8]	60.71 [57.4 – 64.5]	50.65 [47.0 – 54.7]	49.04 [45.2 – 53.0]	**71.90** [69.0 – 74.8]
Triglyceride	51.51 [49.4– 53.6]	44.36 [42.4 – 46.3]	50.63 [48.3 – 53.0]	**55.23** [53.2 – 57.2]	47.50 [45.4 – 49.6]
Average	56.97 ± 8.97	64.98 ± 11.93	56.89 ± 11.74	58.02 ± 6.76	**66.99** ± 10.47
Regression - MAE (↓)					
Sys. BP (Lab)	12.08 [11.6 – 12.6]	**11.99** [11.5 – 12.5]	12.15 [11.6 – 12.7]	12.13 [11.6 – 12.6]	13.12 [12.6 – 13.6]
Dias. BP (Lab)	10.80 [10.4 – 11.4]	10.38 [10.0 – 10.7]	10.78 [10.4 – 11.1]	10.72 [10.4 – 11.0]	**9.66** [9.2 – 9.9]
Sys. BP	13.12 [11.9 – 14.4]	12.93 [11.7 – 14.3]	13.04 [11.8 – 14.4]	13.12 [11.9 – 14.4]	**12.80** [11.6 – 14.1]
Dias. BP	**10.19** [9.3 – 11.0]	10.28 [9.3 – 11.1]	10.30 [9.4 – 11.1]	10.37[9.5 – 11.2]	10.28 [9.4 – 11.1]
Average	11.55 ± 1.14	**11.40** ± 1.12	11.57 ± 1.09	11.59 ± 1.10	11.47 ± 1.52

optimal setting based on ablations in Section 4.2) using PyWavelets [Lee et al., 2019], yielding one approximation and four detail subbands. The high-frequency subbands (2) dominated by noise and with close to zero coefficients are discarded. The remaining subbands are interpolated and arranged in order of increasing frequency to form a 2-D coefficient map of shape $[n_{\text{bands}}, \text{time}]$.

Masked Multiscale Reconstruction – MMR. We adopt a Vision Transformer (ViT) encoder–decoder within the masked autoencoder framework [He et al., 2022]. The 2-D wavelet coefficient map is divided into non-overlapping patches of size $(1, 25)$ along the temporal axis, producing a sequence of tokens for each subband. Fixed 2-D sine–cosine positional embeddings are added to encode temporal and spectral structure. During pretraining, 75% of patches are randomly masked, and the decoder reconstructs the missing coefficients from the visible context. This Masked Multiscale Reconstruction (MMR) objective encourages the encoder to model dependencies across wavelet scales, enabling coarse bands to support fine-scale recovery and fine bands to refine coarse trends. Training minimizes mean-squared error (MSE) between reconstructed and original coefficients over the masked patches.

Experimental Setting We pretrain our encoder with unlabeled 10-second PPG segments collected from different [REDACTED] smartwatches, where the majority of the segments (∼95%) are sampled at a low rate of 25 Hz (due to battery constraints in the wild). Each segment is upsampled, band-pass filtered, and z-score normalized before wavelet decomposition is applied. The dataset is split 80:20 into training and validation sets for pretraining. To evaluate generalization, the pretrained encoder is linearly probed for a set of 13 downstream clinically motivated tasks. Classification tasks include the detection of hypertension, premature ventricular contractions (PVCs), and abnormal laboratory measures (e.g., high/low lipids, electrolytes, platelets), whereas for regression tasks, we perform the prediction of systolic and diastolic blood pressure in both field and laboratory data collection settings.

4 Results

We compare against various baselines such as SimCLR [Chen et al., 2020], masked auto encoding (time-domain) as done in LSM [Narayanswamy et al., 2024], and the open-source PaPaGei family (-S, -P variants)[Pillai et al., 2024]. We further analyze design choices through ablation studies.

4.1 Main Results

MMR performs competitively across the majority of downstream tasks, matching or surpassing strong baselines such as the PaPaGei family, LSM, and SimCLR. Across 13 classification and regression tasks, MMR achieved the top score in 7 (PVC detection, hypertension-lab, diastolic BP, potassium,

HDL, etc.) and came close in nearly all others (61.50 vs 64.30 for LDL, 10.28 vs 10.19 for Dias. BP. MAE), resulting in top-2 performance in 10/13 tasks overall. In classification, MMR achieved strong AUROC scores, including state-of-the-art performance for PVC detection (82.5) and hypertension-lab (67.5). It also had the highest classification scores for abnormal conditions, such as high HDL(62.41), sodium (71.90), and potassium (82.85). Notably, MMR outperforms SimCLR and LSM (trained on the same wearable data as MMR) by up to +18 AUROC points (e.g., high potassium: 82.8 vs. 64.0) and by +10% for PVC detection, respectively. It also outperforms PaPaGei-S by margins of 10–15 AUROC points on several tasks. For regression, MMR achieved the lowest error in diastolic BP-lab (9.66), in the field setting, ranked second for diastolic BP (10.28 vs. 10.19), and led systolic BP regression. Across most tasks, PaPaGei-P remains a strong baseline, while the other variants (PaPaGei-s) lag significantly behind MMR. Importantly, these results were achieved using real-world wearable PPG data sampled at a low rate, compared to PaPaGei models trained on clean, high-frequency (125–500 Hz) clinical signals. Despite the lower sampling rate and higher noise inherent in field data, MMR not only competes but often surpasses clinical-data-trained models.

4.2 Ablation Studies

We ablated a 1M-sample subset, varying wavelet family, decomposition level, and patch size, and evaluated on two representative tasks (Table 2).

Table 2: AUROC scores for two tasks. Ablation of MMR across wavelet family, decomposition level, and patch size. Pretrained on 5% of the full dataset.

Configuration	Hypertension (avg. lab, field)	PVC
db4 – Level 4 – Patch 50	61.54	**73.65**
db4 – Level 6 – Patch 50	63.77	70.83
db4 – Level 7 – Patch 50	64.82	70.05
db4 – Level 6 – Patch 25	**64.87**	69.94
db4 – Level 6 – Patch 100	62.90	70.92
haar – Level 6 – Patch 50	64.11	**74.84**
bior3.5 – Level 6 – Patch 50	61.28	68.52

Wavelet Families: cn exhibit distinct trade-offs in DWT. The Daubechies-4 (db4) provides stable performance across both tasks (Hypertension 63.8%, PVC 70.8%). Haar wavelet family achieves the highest PVC score (74.8%) while maintaining competitive Hypertension performance (64.1%). This can be attributed to Haar's compact nature and sharp discontinuities, which can emphasize abrupt waveform changes that are critical for detecting ectopic beats [Yang et al., 2019]. By contrast, db4 and biorthogonal3.5 offer smoother base functions, which may miss key transients necessary for PVC detection.

Decomposition Level: governs the number of multi-resolution sub-bands/hierarchy available for analysis. The db4 wavelet attains 61.5% Hypertension accuracy at Level 4, increasing to 64.9% at Levels 7, suggesting that deeper decompositions with additional sub-bands yield more informative representations for classification [Singh et al., 2023, Attivissimo et al., 2023]. In contrast, PVC scores within the db4 family peak at Level 4 and decrease with deeper decompositions (70.0% at Level 7).

Patch size: regulates the granularity of the temporal context provided to the encoder. Smaller patches (25) achieve the best Hypertension score (64.9%), whereas a patch size of 50 offers a balanced trade-off, with reasonable Hypertension (63.4%) and strong PVC (71.1%). Large patches (100) reduce Hypertension to (62.9%), emphasizing that a longer window may average out subtle morphological patterns.

In summary, our experiments show that *different tasks benefit from distinct temporal and frequency scales*. This supports the hypothesis that multi-scale PPG information provides complementary cues, highlighting wavelet-based representations as an effective pretraining strategy for adaptive and robust physiological monitoring with PPG.

5 Discussion and Future Work

Pretraining on diverse smartwatch data with wavelet-based multi-scale reconstruction of PPG signals provides a strong foundation for robust physiological feature learning and downstream cardiovascular tasks. Future directions include dynamically learning or adapting to different levels of decompistion, incorporating frequency information through positional embeddings or cross-scale reconstruction where subbands are masked and reconstructed, and further analyzing the frequency–time components the model attends to for deeper interpretability. Advancing along these directions could yield richer and more generalizable PPG foundation models for real-world health applications.

References

Salar Abbaspourazad, Anshuman Mishra, Joseph Futoma, Andrew C Miller, and Ian Shapiro. Wearable accelerometer foundation models for health via knowledge distillation. *arXiv preprint arXiv:2412.11276*, 2024a.

Salar Abbaspourazad et al. Large-scale training of foundation models for wearable biosignals. In *ICLR 2024 Workshop or Poster*, 2024b. URL `https://openreview.net/forum?id=pC3WJHf51j`. Self-supervised learning on Apple Heart Movement Study (PPG/ECG).

Maha Alafeef and Mohammad Fraiwan. Smartphone-based respiratory rate estimation using photoplethysmographic imaging and discrete wavelet transform. *Journal of Ambient Intelligence and Humanized Computing*, 11(2):693–703, 2020.

Filippo Attivissimo, Luisa De Palma, Attilio Di Nisio, Marco Scarpetta, and Anna Maria Lucia Lanzolla. Photoplethysmography signal wavelet enhancement and novel features selection for non-invasive cuff-less blood pressure monitoring. *Sensors*, 23(4):2321, 2023.

Syed Khairul Bashar, Dong Han, Shirin Hajeb-Mohammadalipour, Eric Ding, Cody Whitcomb, David D McManus, and Ki H Chon. Atrial fibrillation detection from wrist photoplethysmography signals using smartwatches. *Scientific reports*, 9(1):15054, 2019.

Yong-Mei Cha, Glenn K Lee, Kyle W Klarich, and Martha Grogan. Premature ventricular contraction-induced cardiomyopathy: a treatable condition. *Circulation: Arrhythmia and Electrophysiology*, 5 (1):229–236, 2012.

Peter H Charlton, Panicos A Kyriacou, Jonathan Mant, Vaidotas Marozas, Phil Chowienczyk, and Jordi Alastruey. Wearable photoplethysmography for cardiovascular monitoring. *Proceedings of the IEEE*, 110(3):355–381, 2022a.

Peter H Charlton, Birutė Paliakaitė, Kristjan Pilt, Martin Bachler, Serena Zanelli, Daniel Kulin, John Allen, Magid Hallab, Elisabetta Bianchini, Christopher C Mayer, et al. Assessing hemodynamics from the photoplethysmogram to gain insights into vascular age: a review from vascagenet. *American Journal of Physiology-Heart and Circulatory Physiology*, 322(4):H493–H522, 2022b.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PmLR, 2020.

Yanlong Chen, Mattia Orlandi, Pierangelo Maria Rapa, Simone Benatti, Luca Benini, and Yawei Li. Physiowave: A multi-scale wavelet-transformer for physiological signal representation. *arXiv preprint arXiv:2506.10351*, 2025.

Rui Cheng, Xiangfei Jia, Qing Li, Rong Xing, Jiwen Huang, Yu Zheng, and Zhilong Xie. Fat: Frequency-aware pretraining for enhanced time-series representation learning. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2*, pages 310–321, 2025.

Ingrid Daubechies. *Ten lectures on wavelets*. SIAM, 1992.

Jufang Duan, Wei Zheng, Yangzhou Du, Wenfa Wu, Haipeng Jiang, and Hongsheng Qi. Multi-frequency contrastive learning representation for time series. In *ICML*, 2024.

En Fu and Yanyan Hu. Frequency-masked embedding inference: A non-contrastive approach for time series representation learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 16639–16647, 2025.

Dong Han, Syed Khairul Bashar, Fahimeh Mohagheghian, Eric Ding, Cody Whitcomb, David D McManus, and Ki H Chon. Premature atrial and ventricular contraction detection using photoplethysmographic data from a smartwatch. *Sensors*, 20(19):5683, 2020.

Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16000–16009, 2022.

Denizhan Kara, Tomoyoshi Kimura, Shengzhong Liu, Jinyang Li, Dongxin Liu, Tianshi Wang, Ruijie Wang, Yizhuo Chen, Yigong Hu, and Tarek Abdelzaher. Freqmae: Frequency-aware masked autoencoder for multi-modal iot sensing. In *Proceedings of the ACM Web Conference 2024*, pages 2795–2806, 2024.

Gregory Lee, Ralf Gommers, Filip Waselewski, Kai Wohlfahrt, and Aaron O'Leary. Pywavelets: A python package for wavelet analysis. *Journal of Open Source Software*, 4(36):1237, 2019.

Simon A Lee and Kai Akamatsu. Foundation models for physiological signals: Opportunities and challenges. 2025.

Ran Liu, Ellen L Zippi, Hadi Pouransari, Chris Sandino, Jingping Nie, Hanlin Goh, Erdrin Azemi, and Ali Moin. Frequency-aware masked autoencoders for multimodal pretraining on biosignals. *arXiv preprint arXiv:2309.05927*, 2023.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.

Stephane G Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE transactions on pattern analysis and machine intelligence*, 11(7):674–693, 2002.

Luca Masserano, Abdul Fatir Ansari, Boran Han, Xiyuan Zhang, Christos Faloutsos, Michael W Mahoney, Andrew Gordon Wilson, Youngsuk Park, Syama Rangapuram, Danielle C Maddix, et al. Enhancing foundation models for time series forecasting via wavelet-based tokenization. *arXiv preprint arXiv:2412.05244*, 2024.

Mina Namvari, Jessica Lipoth, Sheida Knight, Ali Akbar Jamali, Mojtaba Hedayati, Raymond J Spiteri, and Shabbir Syed-Abdul. Photoplethysmography enabled wearable devices and stress detection: a scoping review. *Journal of Personalized Medicine*, 12(11):1792, 2022.

Girish Narayanswamy, Xin Liu, Kumar Ayush, Yuzhe Yang, Xuhai Xu, Shun Liao, Jake Garrison, Shyam Tailor, Jake Sunshine, Yun Liu, et al. Scaling wearable foundation models. *arXiv preprint arXiv:2410.13638*, 2024.

National Library of Medicine (US). Medlineplus. `https://medlineplus.gov/`, 2020. [Internet]. Bethesda (MD): National Library of Medicine (US); [updated 2020-06-24; cited 2025-08-31].

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems*, 32, 2019.

Arvind Pillai et al. Papagei: Open foundation models for optical physiological signals. *arXiv preprint arXiv:2410.20542*, 2024. URL `https://arxiv.org/abs/2410.20542`.

Mithun Saha et al. Pulse-ppg: An open-source field-trained ppg foundation model for wearable applications. *arXiv preprint arXiv:2502.01108*, 2025. URL `https://arxiv.org/abs/2502.01108`.

Shiliang Shao, Ting Wang, Lebing Wang, Sinan Li, and Chen Yao. A photoplethysmograph signal preprocess method based on wavelet transform. In *2021 36th Youth Academic Annual Conference of Chinese Association of Automation (YAC)*, pages 771–775. IEEE, 2021.

Bikesh Kumar Singh, Neelamshobha Nirala, et al. Expert diagnostic system for detection of hypertension and diabetes mellitus using discrete wavelet decomposition of photoplethysmogram signal and machine learning technique. *Medicine in Novel Technology and Devices*, 19:100251, 2023.

Kwangsub Song, Ku-young Chung, and Joon-Hyuk Chang. Cuffless deep learning-based blood pressure estimation for smart wristwatches. *IEEE Transactions on Instrumentation and Measurement*, 69(7):4292–4302, 2019.

Maxwell A Xu, Girish Narayanswamy, Kumar Ayush, Dimitris Spathis, Shun Liao, Shyam A Tailor, Ahmed Metwally, A Ali Heydari, Yuwei Zhang, Jake Garrison, et al. Lsm-2: Learning from incomplete wearable sensor data. *arXiv preprint arXiv:2506.05321*, 2025.

Chaoqi Yang, M. Brandon Westover, and Jimeng Sun. Biot: Biosignal transformer for cross-data learning in the wild. In *NeurIPS 2023*, 2023. URL https://openreview.net/forum?id=c2LZyTyddi.

Chengming Yang, Cesar Veiga, Juan J Rodriguez-Andina, Jose Farina, Andres Iniguez, and Shen Yin. Using ppg signals and wearable devices for atrial fibrillation screening. *IEEE Transactions on Industrial Electronics*, 66(11):8832–8842, 2019.

Xiang Zhang, Ziyuan Zhao, Theodoros Tsiligkaridis, and Marinka Zitnik. Self-supervised contrastive pre-training for time series via time-frequency consistency. In *NeurIPS*, 2022.

A Appendix

A.1 Extended Related Work

Self-supervised pretraining has emerged as the dominant paradigm for large-scale biosignal modeling. For example, [Abbaspourazad et al., 2024b] trained foundation models on PPG and ECG from ∼141K Apple Watch users, demonstrating the value of contrastive learning at scale. In parallel, [Pillai et al., 2024] introduced PaPaGei , an open-source PPG foundation model trained on 20M unlabeled fingertip PPG segments that explicitly leverages waveform morphology, while [Saha et al., 2025] developed Pulse-PPG using 100 days of field data from 120 participants, showing improved efficiency and generalizability. Beyond single-modality PPG, multimodal biosignal foundations transfer representations across ECG, PPG, and other signals either via knowledge distillation [Abbaspourazad et al., 2024a] or unified embeddings [Yang et al., 2023]. Related work has also applied masked reconstruction on multivariate health time series, yielding strong generative and discriminative performance on tasks such as activity classification [Narayanswamy et al., 2024, Xu et al., 2025]. Together, these advances reflect a shift from task-specific models to general-purpose foundation models for biosignals.

While these foundation models highlight the value of large-scale self-supervision, most treat signals purely in the time domain. A growing body of work shows that explicitly incorporating spectral information provides a powerful inductive bias for robust and transferable representations. For instance, Time-Frequency Consistency [Zhang et al., 2022] proposed aligning time- and frequency-domain views via contrastive loss, while bioFAME [Liu et al., 2023] introduced a frequency-aware transformer encoder with multi-head spectral filters. Similarly, FreqMAE [Kara et al., 2024] leveraged temporal-shifting encoders to model spectral content in multimodal IoT data. More recent approaches, such as FAT [Cheng et al., 2025], FEI [Fu and Hu, 2025], and MF-CLR [Duan et al., 2024], further illustrate how spectral modeling can enhance time-series representation learning. These findings suggest that frequency-aware pretraining can serve as a complementary approach to large-scale training for physiological signals such as PPG.

Wavelet analysis provides a natural way to capture information at different temporal scales by decomposing signals into multi-resolution frequency bands. Earlier PPG studies applied discrete wavelet transforms (DWT) for denoising and handcrafted features, for example in respiratory rate estimation [Alafeef and Fraiwan, 2020], hypertension and diabetes detection [Singh et al., 2023], and peak stabilization pipelines [Shao et al., 2021]. More recently, deep learning models have incorporated wavelets end-to-end, such as wavelet-based tokenization for time-series foundation models [Masserano et al., 2024] and PhysioWave [Chen et al., 2025], which couples learned wavelet decompositions, frequency guided masking with Transformers for physiological signals such as ECG and EMG. Our work extends this line and introduces a multi-resolution masked pretraining framework for large-scale PPG data collected from smartwatches in real-world settings. By leveraging the fact that health tasks rely on information at multiple signal granularities, our approach provides more physiologically grounded and transferable representations.

A.2 Training Setup

We pretrain MMR using the AdamW [Loshchilov and Hutter, 2017] optimizer with a base learning rate of 1×10^{-4}, cosine decay schedule, and linear warmup over the first 10% of steps. Training is performed for ∼69K steps with a batch size of 512, weight decay of 1e-5, and gradient clipping at 1.0, while adopting the same augmentations (i.e., time-flip, adding Gaussian noise, and stretching along the temporal axis) as LSM [Narayanswamy et al., 2024] to the PPG signal before wavelet decomposition. The backbone follows a ViT-Small configuration (∼7M parameters) with 8 encoder blocks (hidden size 256, 4 heads, feedforward size 1024) and a lightweight decoder of 2 blocks (hidden size 192, 4 heads) used only during pretraining for reconstruction. Hyperparameter tuning was minimal, limited to a small grid search over 2–3 learning rates ∈ {1e-2, 1e-3, 1e-4}and decay values ∈ 1e-3, 1e-4, 1e-5}a, with sweeps and ablations run on a subset of the pretraining data (∼ 1M data points). All experiments are conducted on 4 Tesla T4 GPUs (16GB each) with distributed data parallel (DDP) training in PyTorch [Paszke et al., 2019]. For the baselines, we use the pretrained weights for the PaPaGei family while we pretrain SimCLR and LSM on our pretraining data.

Table 3: Downstream datasets. Counts are (#positive / #negative) segments

Task	Setting	Train (pos/neg)	Test (pos/neg)
Hypertension	Lab (protocol)	746 / 3124	561 / 398
	Naturalistic (field)	542 / 373	140 / 104
PVC Detection	Wearable	14419 / 166270	4491 / 37947
Laboratory Tests			
HDL	Clinical reports	4117 / 3805	283 / 807
LDL	Clinical reports	2518 / 3155	779 / 613
Sodium	Clinical reports	3928 / 2700	1115 / 239
Potassium	Clinical reports	4755 / 5746	1689 / 835
Paleteletes	Clinical reports	3096 / 3590	1291 / 712

A.3 Dataset details

We pretrain on data collected from various types of [REDACTED] smartwatches where PPG is sampled at different sampling frequencies (e.g., 100, 25 Hz). These datasets provide diverse signals collected under [REDACTED] distinct studies and user groups. Such data closely reflect real-world conditions, making them highly representative for PPG-based wearable applications.

We evaluate on several downstream datasets collected in different settings:

- Hypertension-Lab: 63 users (50 train / 13 test) with protocolized data collection; segments: 746 positive / 3124 negative for training, and 561 positive / 398 negative for testing.
- Hypertension-Field: 915 train users / 244 test users; segments: 542 positive / 373 negative for training, and 140 positive / 104 negative for testing.
- PVC detection: 14,419 positive / 166,270 negative segments for training, and 4,491 positive / 37,947 negative segments for testing.
- Laboratory biomarkers (HDL, LDL, Sodium, Potassium): smaller user-sparse datasets ($\approx$15–30 users per task) with class imbalance; segment splits are provided in Table 3.

All analyses/tasks are performed at the segment level. To mitigate imbalance during linear probing, we downsample the majority class in the training (not in the test) split. Random forest classifiers and linear regression models are used for probing with a 4-fold cross-validation. We also compute 95% confidence intervals via bootstrapping with 500 resampling runs similar to [Pillai et al., 2024].

Hypertension Classification We define hypertension as a binary classification task based on clinical guidelines: individuals are labeled as *Hypertensive* (label 1) if their systolic blood pressure is ≥ 130 mmHg or diastolic blood pressure is ≥ 80 mmHg, and *Normal* (label 0) otherwise. We apply buffer thresholds of ± 8 mmHg around the diagnostic cutoffs.

PVC Detection Premature Ventricular Contractions (PVCs) are early heartbeats originating in the ventricles [Cha et al., 2012]. They can indicate underlying cardiac conditions or increased risk of arrhythmias. We label high PVC burden as class 1 and low PVC burden as class 0.

Laboratory Tests For various laboratory tests (explained below as per [National Library of Medicine (US), 2020]), we adopt a binary classification scheme where high values are labeled as class 1 and class 0 otherwise.

- Sodium: Elevated sodium (hypernatremia) is linked to dehydration or adrenal gland/kidney dysfunction.
- Potassium: High potassium (hyperkalemia) may cause cardiac arrhythmias; low potassium (hypokalemia) is associated with muscle weakness, fatigue and rhythm disturbances.
- Platelets: Elevated platelet counts can signal inflammation or clotting risk.
- Low-Density Lipoprotein LDL: High LDL is a risk factor for peripheral artery disease and heart stroke.
- High-Density Lipoprotein HDL: High HDL helps lower heart disease and heart attack.

- Triglycerides: Elevated triglycerides increase cardiovascular risk and are often associated with metabolic syndrome