TIME-SERIES FOUNDATION MODEL EMBEDDINGS AS MEANS FOR PHYSIOLOGICAL FEATURE EXTRACTION

Yujin Lee
Department of Industrial Engineering
Sungkyunkwan University
bani99@skku.edu

Fanurs Chi-En Teh*
Auton Lab, Robotics Institute
Carnegie Mellon University
fanurs@cmu.edu

Joo Heung Yoon
Department of Medicine
University of Pittsburgh
yoonjh@upmc.edu

Artur Dubrawski
Auton Lab, Robotics Institute
Carnegie Mellon University
awd@cs.cmu.edu

ABSTRACT

Physiological time-series, such as electrocardiograms, photoplethysmograms, or respiratory waveforms, are widely used for clinical assessment, prognosis, and monitoring. Recently, time-series foundation models (TSFMs) have demonstrated strong performance across various downstream tasks in the medical domain. However, it remains unclear whether TSFM embeddings include physiologically meaningful information. In this study, we systematically examine physiological features to evaluate whether TSFM embeddings provide a significant representation premium over raw signals by mapping temporal data into a more learnable representation space. We demonstrate this by training regressors to predict clinically informative diagnostic features. Addressing the limitation that TSFM embeddings lack amplitude information due to normalization, we propose reinjecting global scale statistics to the embeddings. Experimental results show that embedding-based feature extraction achieves on average 17.90% improvement in relative RMSE compared to raw signal-based extraction, with particularly strong gains in real-world noisy datasets. Specifically, periodicity-based features exhibited a substantial average performance improvement of 26.02%, while amplitude-based features showed an increase of 18.28%. In comparison, morphology-based features showed comparatively limited improvements, achieving gains of 11.82% from raw signals and 2.60% from second-derivative signals.

Track: Research

1 INTRODUCTION

Physiological time-series, such as electrocardiograms (ECG), photoplethysmograms (PPG), and respiratory waveforms (Resp), have been widely used in clinical settings for patient monitoring and clinical decision-making (Serhani et al., 2020; Alian & Shelley, 2022; Iqbal et al., 2022). With recent advances, time-series foundation models (TSFMs) have been increasingly applied to various downstream tasks in the medical domain (Zhang et al., 2025; Chen et al., 2025). However, existing studies have largely focused on downstream task performance, leaving the utility of internal representations of these models underexplored. Without confirming whether TSFMs capture stable physiological biomarkers, their applicability as general-purpose medical encoders would remain speculative.

Prior studies indicate that general-purpose TSFMs can be effective at medical tasks, sometimes outperforming specialized TSFMs. Specifically, Kataria et al. (2024) had shown that MOMENT (Goswami et al., 2024), a general-purpose TSFM, demonstrates superior classification performance compared to the specialized model GPT-PPG (Chen et al., 2025). Pucher et al. (2025) also reported

*Corresponding author.

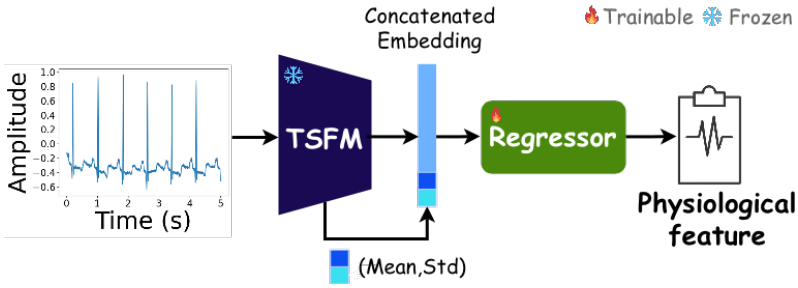


Figure 1: Overview of the embedding-based feature extraction framework. Raw signals are encoded by MOMENT, a general-purpose TSFM, to generate embeddings which are concatenated with the mean and standard deviation of raw signal and used to regress for physiological features.

that general-purpose TSFMs exhibit more robust forecasting performance on external electronic health record (EHR) datasets compared to specialized models.

In this study, we systematically analyze embeddings obtained from a general-purpose TSFM to characterize the clinical information they encode. To this end, we formulate a physiological feature extraction framework from ECG, PPG, and Resp signals as a regression task. The objective of this study is not to replace established feature extraction pipelines, such as peak detection or fourier based analysis, but to assess whether TSFM embeddings capture physiological information and provide a more learnable representation than raw signals. Furthermore, we identify inherent limitations of embeddings in capturing certain types of information and explore strategies to mitigate it.

Experimental results demonstrate that physiological features derived from embeddings yield 16.34% performance improvement compared to those obtained directly from raw signals. These findings suggest that general-purpose TSFM embeddings provide a more effective representation space than raw signals, supporting their use in downstream clinical tasks. We further find that the embeddings effectively capture periodicity-based features. By contrast, their ability to represent amplitude-based features is comparatively limited, mostly due to the typical scaling of raw signals fed into TSFMs. To mitigate this effect, we incorporate global scale statistics, specifically the mean and standard deviation of the raw signal, alongside embeddings, improving the extraction of amplitude-based features by 14.28%.

2 METHODS

Let the dataset be defined as $\mathcal{D} = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{N}$, where $i \in \{1, \ldots, N\}$ indexes individual samples and N denotes the number of samples. Each sample $\mathbf{x}_i \in \mathbb{R}^T$ represents a univariate physiological time-series of fixed length T. For each time-series $\mathbf{x}_i$, a set of physiological features is defined as $\mathbf{y}_i = (y_i^{(1)}, \ldots, y_i^{(K)}) \in \mathbb{R}^K$, where each $y_i^{(k)} \in \mathbb{R}$ represents an individual feature value. Here, $k \in \{1, \ldots, K\}$ indexes the individual features and K denotes the number of features.

Figure 1 presents an overview of the proposed framework, where MOMENT is employed as a TSFM. For a given input time-series $\mathbf{x}_i$, we extract a fixed-dimensional embedding $\mathbf{z}_i = f_{\boldsymbol{\theta}}(\mathbf{x}_i)$ using a pretrained MOMENT $f_{\boldsymbol{\theta}}$ with frozen parameters $\boldsymbol{\theta}$. MOMENT adopts Reversible Instance Normalization (RevIN) (Kim et al., 2021) for sample-wise scale normalization of the input time-series. Thus, the amplitude information of input $\mathbf{x}_i$ is not retained in the embedding $\mathbf{z}_i$. To preserve scaling information in the embeddings, we concatenate the mean and standard deviation of each raw sample, which capture the absolute magnitudes necessary for accurate clinical feature regression. The concatenated embedding is defined as $\tilde{\mathbf{z}}_i = [\mathbf{z}_i \,;\, \mu_i \,;\, \sigma_i]$, where μ_i and σ_i denote the mean and standard deviation used by RevIN to normalize the input time-series $\mathbf{x}_i$.

Physiological features are estimated from $\tilde{\mathbf{z}}_i$ using a dedicated regressor $g^{(k)}(\cdot)$ for each individual feature $y^{(k)}$, with a prediction given by $\hat{y}_i^{(k)} = g^{(k)}(\tilde{\mathbf{z}}_i)$.

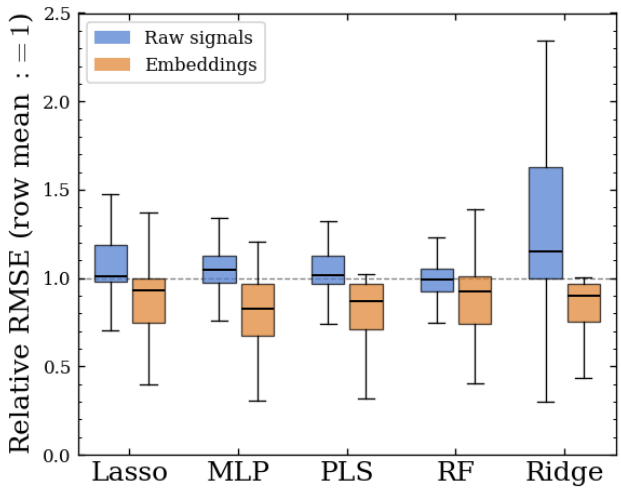


Figure 2: Distribution of relative RMSE across regressors (the lower the better). Blue boxes correspond to raw signal-based feature extraction, and orange boxes correspond to MOMENT embedding-based feature extraction. The dashed horizontal line at 1 indicates that the embedding-based RMSE equals the average RMSE of raw signal-based feature extraction across regressors.

3 Experiments and Results

We compared two feature extraction paradigms under identical regression settings: (1) embedding-based feature extraction, where regressors are trained on embeddings obtained from TSFM, and (2) raw signal-based feature extraction, where regressors are trained directly on raw time-series. We adopted MOMENT-small (Goswami et al., 2024) as the TSFM backbone, where RevIN was used to standardize input signals. To extract physiological features from the concatenated embeddings, we employed multiple regressors, including Lasso, Multilayer Perceptron (MLP), Partial Least Squares (PLS), Random Forest (RF), and Ridge.

We analyzed ECG, PPG, and Resp signals using synthetic and real-world benchmark datasets. Synthetic ECG, PPG, and Resp datasets are generated using `NeuroKit2` (Makowski et al., 2021). For the real-world benchmark datasets, we used the MIT-BIH arrhythmia dataset (Moody & Mark, 2001; Goldberger et al., 2000) for ECG signals and the BID-MC dataset (Pimentel et al., 2016; Goldberger et al., 2000) for PPG and Resp signals.

We considered 15 features for ECG, 23 features for PPG, and 4 features for Resp, all selected by an experienced clinician based on their practical utility in medical diagnosis and prognosis. Three additional features were included for PPG and Resp signals in the BID-MC dataset, as clinically meaningful annotations were available for these features. We group the physiological features into four categories: periodicity-based, amplitude-based, morphology-based (raw), morphology-based (2_{nd} deriv.). Periodicity-based features capture temporal periodic patterns of signals, whereas amplitude-based features reflect waveform manitudes. Morphology-based features characterize the shape of physiological signal waveforms, where morphology-based (raw) features describe non-differentiated signals and morphology-based (2^{nd} deriv.) features describe its second derivative. The justification of feature selection and detailed descriptions of all used features are provided in Appendix A.4.2, A.5, and summarized in Table 5.

Performance was evaluated using the relative root mean squared error (RMSE), computed as the ratio of the RMSEs between two methods, with the specific methods compared varying across tables and figures. The relative RMSE reported in Table 1 and Table 3 is defined as the ratio of the RMSE obtained from embedding-based feature extraction to that from raw signal-based feature extraction using the same regressor. For Figure 2, the relative RMSE refers to the RMSE from embedding-based feature extraction divided by the average RMSE of raw signal-based feature extraction across all regressors. Additionally, for Table 2 and Table 4, we use relative RMSE, which measures the

ratio of the RMSE with and without concatenating global statistics to the embeddings. Detailed experimental settings are provided in the Appendix A.1.

The overall comparison is summarized in Table 1 and Figure 2, while the complete results are presented in Table 3. As shown in Table 1, embedding-based feature extraction outperforms raw signal-based extraction, with an average improvement of 17.90% across all experimental settings. For individual regressors, the average relative RMSE improvement was 14.35% for Lasso, 22.82% for MLP, 14.03% for PLS, 7.99% for RF, and 30.32% for Ridge. This trend is consistently observed in Figure 2, which illustrates the performance improvements across all regressors. These findings suggest that TSFM embeddings provide a more effective representation space for physiological feature extraction than raw signals.

Table 1: Comparison of the average relative RMSE across signal types, datasets, and feature categories. Values less than 1 indicate that embedding-based feature extraction outperforms raw signal-based feature extraction and are highlighted in bold.

Signal	Dataset	Feature type	# Feats.	Lasso	MLP	PLS	RF	Ridge
ECG	Synthetic	overall	15	**0.809**	**0.716**	**0.828**	**0.887**	**0.791**
		periodicity	10	**0.848**	**0.654**	**0.820**	**0.888**	**0.761**
		morphology (raw)	4	**0.713**	**0.766**	**0.847**	**0.907**	**0.884**
		amplitude	1	**0.958**	1.068	1.073	1.088	1.082
	MIT-BIH	overall	15	**0.776**	**0.740**	**0.728**	**0.800**	**0.722**
		periodicity	10	**0.791**	**0.731**	**0.758**	**0.825**	**0.752**
		morphology (raw)	4	**0.848**	**0.791**	**0.758**	**0.827**	**0.748**
		amplitude	1	**0.344**	**0.630**	**0.316**	**0.437**	**0.311**
PPG	Synthetic	overall	23	1.013	**0.924**	**0.922**	1.024	**0.922**
		periodicity	5	1.071	**0.930**	**0.931**	1.047	**0.928**
		morphology (raw)	8	1.047	**0.978**	**0.967**	1.025	**0.969**
		morphology (2^{nd} deriv.)	7	1.055	**0.941**	**0.972**	1.012	**0.972**
		amplitude-based	3	**0.996**	**0.935**	**0.912**	**0.972**	**0.906**
	BID-MC	overall	26	**0.806**	**0.675**	**0.912**	**0.927**	**0.451**
		periodicity	7	**0.556**	**0.449**	**0.535**	**0.683**	**0.385**
		morphology (raw)	8	**0.840**	**0.822**	**0.836**	**0.947**	**0.619**
		morphology (2^{nd} deriv.)	7	1.007	**0.797**	1.360	1.314	**0.309**
		amplitude-based	3	**0.890**	**0.662**	1.043	**0.621**	**0.523**
Resp	Synthetic	overall	4	**0.831**	**0.867**	**0.835**	**0.862**	**0.832**
		periodicity	3	**0.776**	**0.849**	**0.783**	**0.818**	**0.779**
		morphology (raw)	1	**0.996**	**0.920**	**0.990**	**0.993**	**0.992**
	BID-MC	overall	7	**0.816**	**0.763**	**0.824**	**0.913**	**0.536**
		periodicity	5	**0.789**	**0.721**	**0.797**	**0.893**	**0.466**
		morphology (raw)	1	**0.980**	1.017	**0.986**	1.031	**0.955**

Notably, embedding-based feature extraction yields greater performance gains on real-world benchmark datasets than on synthetic datasets. The average relative RMSE across regressors improves by 10.52% on synthetic datasets and 24.36% on benchmark datasets. In idealized synthetic settings, high signal quality enables simple regressors to perform well on raw waveforms, limiting the additional benefits of embeddings. However, in real-world scenarios, MOMENT embeddings demonstrate a significant representation premium. This suggests that pre-training provides a structured denoising effect, preserving biomarkers in degraded waveforms where direct regression on raw signals may fail.

Moreover, performance gains vary substantially across feature types. The average relative RMSE improvement was 26.02% for periodicity-based features, 18.28% for amplitude-based features, and 11.82% and 2.60% for morphology-based features derived from raw signals and their second derivatives, respectively. These findings indicate that embeddings are especially effective in capturing periodicity-based features. Despite this overall improvement, morphology-based features obtained from the second derivative of the signal remain challenging, suggesting that embeddings may not fully capture subtle higher-order morphological variations.

Table 2: Comparison of the average relative RMSE reflecting the difference between RMSE values with and without global scale statistics (mean and standard deviation). Values less than 1 indicate that embedding-based feature extraction outperforms raw signal-based feature extraction and are highlighted in bold.

Signal	Dataset	Feature type	Lasso	MLP	PLS	RF	Ridge
ECG	Synthetic	Non-amplitude	1.000	**0.817**	1.001	**0.999**	1.000
		Amplitude	1.000	**0.982**	**0.747**	**0.796**	**0.828**
	MIT-BIH	Non-amplitude	**0.992**	**0.870**	**0.992**	**0.970**	**0.992**
		Amplitude	**0.635**	**0.775**	**0.632**	**0.595**	**0.633**
PPG	Synthetic	Non-amplitude	1.000	1.000	1.000	1.000	1.000
		Amplitude	1.000	**0.998**	**0.980**	**0.986**	**0.985**
	BID-MC	Non-amplitude	**0.996**	1.003	**0.997**	**0.985**	1.000
		Amplitude	1.000	**0.998**	**0.980**	**0.986**	**0.985**
Resp	Synthetic	Non-amplitude	**0.969**	**0.997**	**0.969**	**0.927**	**0.975**
	BID-MC	Non-amplitude	**0.995**	1.028	**0.994**	**0.917**	**0.994**

As shown in Table 2 and Table 4, concatenating embeddings with mean and standard deviation consistently improves performance, particularly for amplitude-based features. Larger gains are observed for amplitude-based features, with an average improvement of 14.03%. This improvement is expected, as MOMENT employs RevIN, which removes scale information during embedding extraction. By re-injecting the global scale statistics, we restore amplitude cues that are otherwise suppressed. In contrast, performance improvements for non-amplitude-based features are relatively modest, with an average gain of 1.31%.

4 Conclusion

This study provides quantitative evidence that TSFM embeddings serve as robust proxies to estimate clinically meaningful physiological features. We identify a significant representation premium: while raw signals are sufficient in idealized synthetic settings, TSFM embeddings are more learnable and yield more resilient results in noisy, real-world scenarios. Our analysis also addresses "amplitude erasure" in scale-invariant foundation model architectures. We show that re-injecting global scale metrics helps to restore the information needed for accurate estimation of amplitude-based physiological features.

While using TSFMs in physiological feature extraction pipelines incurs higher computational costs than traditional algorithms, they offer a unified representation that reduces the need for task-specific signal engineering. They also provide greater robustness to signal imperfections, enabling more stable basis for clinical assessment under real-world conditions.

5 Acknowledgement

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korean government (MSIT) (RS-2022-00143911, AI Excellence Global Innovative Leader Education Program). This work has been partially supported by the National Science Foundation (awards 2406231 and 2427948) and by the National Institutes of Health (awards R01NR013912 and R35GM159939).

References

Paul S Addison, Andre Antunes, Dean Montgomery, and Ulf R Borg. Non-Contact Monitoring of Inhalation-Exhalation (I: E) Ratio in Non-Ventilated Subjects. *IEEE Journal of Translational Engineering in Health and Medicine*, 2024.

Aymen Awad Alian and Kirk H Shelley. PPG in clinical monitoring. In *Photoplethysmography*, pp. 341–359. Elsevier, 2022.

Benedikt Alkin, Lukas Miklautz, Sepp Hochreiter, and Johannes Brandstetter. Mim-refiner: A contrastive learning boost from intermediate pre-trained representations. *arXiv preprint arXiv:2402.10093*, 2024.

Abdul Fatir Ansari, Lorenzo Stella, Caner Turkmen, Xiyuan Zhang, Pedro Mercado, Huibin Shen, Oleksandr Shchur, Syama Sundar Rangapuram, Sebastian Pineda Arango, Shubham Kapoor, et al. Chronos: Learning the language of time series. *arXiv preprint arXiv:2403.07815*, 2024.

RM Boyle, HL Adlakha, and DA Mary. Diagnostic value of the maximal ST segment/heart rate slope in asymptomatic factory populations. *Journal of Electrocardiology*, 20:128–134, 1987.

Zhaoliang Chen, Cheng Ding, Saurabh Kataria, Runze Yan, Minxiao Wang, Randall Lee, and Xiao Hu. GPT-PPG: a GPT-based foundation model for photoplethysmography signals. *Physiological Measurement*, 46(5):055004, 2025.

Seokmin Choi, Sajad Mousavi, Phillip Si, Haben G Yhdego, Fatemeh Khadem, and Fatemeh Afghah. Ecgbert: Understanding hidden language of ecgs with self-supervised representation learning. *arXiv preprint arXiv:2306.06340*, 2023.

Abhimanyu Das, Weihao Kong, Rajat Sen, and Yichen Zhou. A decoder-only foundation model for time-series forecasting. In *Forty-first International Conference on Machine Learning*, 2024.

Jason F Deen, Dorothy A Rhoades, Carolyn Noonan, Lyle G Best, Peter M Okin, Richard B Devereux, and Jason G Umans. Comparison of QRS duration and associated cardiovascular events in American Indian men versus women (The Strong Heart Study). *The American journal of cardiology*, 119(11):1757–1762, 2017.

Shanghua Gao, Teddy Koker, Owen Queen, Thomas Hartvigsen, Theodoros Tsiligkaridis, and Marinka Zitnik. Units: Building a unified time series model, 2024.

Azul Garza, Cristian Challu, and Max Mergenthaler-Canseco. TimeGPT-1. *arXiv preprint arXiv:2310.03589*, 2023.

Márton Á Goda, Peter H Charlton, and Joachim A Behar. pyPPG: a Python toolbox for comprehensive photoplethysmography signal analysis. *Physiological Measurement*, 45(4):045001, 2024.

Ary L Goldberger, Luis AN Amaral, Leon Glass, Jeffrey M Hausdorff, Plamen Ch Ivanov, Roger G Mark, Joseph E Mietus, George B Moody, Chung-Kang Peng, and H Eugene Stanley. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *circulation*, 101(23):e215–e220, 2000.

Mononito Goswami, Konrad Szafer, Arjun Choudhry, Yifu Cai, Shuo Li, and Artur Dubrawski. MOMENT: A family of open time-series foundation models. In *International Conference on Machine Learning*, Vienna, Austria, July 2024.

Chan Sik Han and Keon Myung Lee. Leveraging Intermediate Representations of Time Series Foundation Models for Anomaly Detection. *arXiv preprint arXiv:2509.12650*, 2025.

Shi Bin Hoo, Samuel Müller, David Salinas, and Frank Hutter. The tabular foundation model tabpfn outperforms specialized time series forecasting models based on simple features. In *NeurIPS Workshop on Time Series in the Age of Large Models*, 2024.

Takuya Horie, Naoto Burioka, Takashi Amisaki, and Eiji Shimizu. Sample entropy in electrocardiogram during atrial fibrillation. *Yonago Acta Medica*, 61(1):049–057, 2018.

Talha Iqbal, Adnan Elahi, Sandra Ganly, William Wijns, and Atif Shahzad. Photoplethysmography-based respiratory rate estimation algorithm for health monitoring applications. *Journal of medical and biological engineering*, 42(2):242–252, 2022.

Suprabhath Kalahasti, Benjamin Faucher, Boxuan Wang, Claudio Ascione, Ricardo Carbajal, Maxime Enault, Christophe Vincent Cassis, Titouan Launay, Caroline Guerrisi, Pierre-Yves Boëlle, et al. Foundation time series models for forecasting and policy evaluation in infectious disease epidemics. *medRxiv*, pp. 2025–02, 2025.

Saurabh Kataria, Yi Wu, and Xiao Hu. Foundation Models for Physiological Data, 2024. URL https://www.siam.org/publications/siam-news/articles/foundation-models-for-physiological-data/.

Taesung Kim, Jinhee Kim, Yunwon Tae, Cheonbok Park, Jang-Ho Choi, and Jaegul Choo. Reversible instance normalization for accurate time-series forecasting against distribution shift. In *Proceedings of International conference on learning representations*, 2021.

Siva Rama Krishna Kottapalli, Karthik Hubli, Sandeep Chandrashekhara, Garima Jain, Sunayana Hubli, Gayathri Botla, and Ramesh Doddaiah. Foundation Models for Time Series: A Survey. *arXiv preprint arXiv:2504.04011*, 2025.

Hao Li, Bowen Deng, Chang Xu, Zhiyuan Feng, Viktor Schlegel, Yu-Hao Huang, Yizheng Sun, Jingyuan Sun, Kailai Yang, Yiyao Yu, et al. MIRA: Medical Time Series Foundation Model for Real-World Health Data. *arXiv preprint arXiv:2506.07584*, 2025.

Yuxuan Liang, Haomin Wen, Yuqi Nie, Yushan Jiang, Ming Jin, Dongjin Song, Shirui Pan, and Qingsong Wen. Foundation models for time series analysis: A tutorial and survey. In *Proceedings of the 30th ACM SIGKDD conference on knowledge discovery and data mining*, pp. 6555–6565, 2024.

Mingzhu Liu, Angela H Chen, and George H Chen. Generalized Prompt Tuning: Adapting Frozen Univariate Time Series Foundation Models for Multivariate Healthcare Time Series. *arXiv preprint arXiv:2411.12824*, 2024.

Dominique Makowski, Tam Pham, Zen J Lau, Jan C Brammer, François Lespinasse, Hung Pham, Christopher Schölzel, and SH Annabel Chen. NeuroKit2: A Python toolbox for neurophysiological signal processing. *Behavior research methods*, 53(4):1689–1696, 2021.

George B Moody and Roger G Mark. The impact of the MIT-BIH arrhythmia database. *IEEE engineering in medicine and biology magazine*, 20(3):45–50, 2001.

Junyung Park, Hyeon Seok Seok, Sang-Su Kim, and Hangsik Shin. Photoplethysmogram analysis and applications: an integrative review. *Frontiers in physiology*, 12:808451, 2022.

Juyoung Park, Seunghan Lee, and Kyungtae Kang. Arrhythmia detection using amplitude difference features based on random forest. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 5191–5194. IEEE, 2015.

Marco AF Pimentel, Alistair EW Johnson, Peter H Charlton, Drew Birrenkott, Peter J Watkinson, Lionel Tarassenko, and David A Clifton. Toward a robust estimation of respiratory rate from pulse oximeters. *IEEE Transactions on Biomedical Engineering*, 64(8):1914–1923, 2016.

Gernot Pucher, Amin Dada, Felix Nensa, Martin Schuler, Christian Reinhardt, Jens Kleesiek, and Christopher M Sauer. Evaluating zero-shot foundation models for time series forecasting in clinical settings: A simulation study with electronic health records. *Studies in health technology and informatics*, 329:820–824, 2025.

Peter Vibe Rasmussen, Jonas Bille Nielsen, Morten Wagner Skov, Adrian Pietersen, Claus Graff, Bent Lind, Johannes Jan Struijk, Morten Salling Olesen, Stig Haunsø, Lars Køber, et al. Electrocardiographic PR interval duration and cardiovascular risk: results from the Copenhagen ECG study. *Canadian Journal of Cardiology*, 33(5):674–681, 2017.

Kashif Rasul, Arjun Ashok, Andrew Robert Williams, Arian Khorasani, George Adamopoulos, Rishika Bhagwatkar, Marin Biloš, Hena Ghonia, Nadhir Hassen, Anderson Schneider, et al. Lag-llama: Towards foundation models for time series forecasting. In *R0-FoMo: Robustness of Few-shot and Zero-shot Learning in Large Foundation Models*, 2023.

Mohamed Adel Serhani, Hadeel T. El Kassabi, Heba Ismail, and Alramzana Nujum Navaz. ECG monitoring systems: Review, architecture, processes, and key challenges. *Sensors*, 20(6):1796, 2020.

Je-Min Song, Gye-Hwan Jin, Sung-Bo Seo, Jeong-Seok Park, Sang-Bock Lee, and Keun-Ho Ryu. Design and implementation of a prediction system for cardiovascular diseases using PPG. *Journal of the Korean Society of Radiology*, 5(1):19–25, 2011.

Ethan Steinberg, Jason Fries, Yizhe Xu, and Nigam Shah. MOTOR: A time-to-event foundation model for structured medical records. *arXiv preprint arXiv:2301.03150*, 2023.

Toshiyo Tamura. Current progress of photoplethysmography and SPO2 for health monitoring. *Biomedical engineering letters*, 9(1):21–36, 2019.

Yanqin Wang, Guoyu Huang, Ceran Chen, Qiu Li, and Ximing Xu. A comparative study of time series foundation models for hand, foot, and mouth disease forecasting: TimesFM, Moirai, and traditional approaches. *Frontiers in Public Health*, 13:1634138, 2025.

Katherine D Wick, Michael A Matthay, and Lorraine B Ware. Pulse oximetry for the diagnosis and management of acute respiratory distress syndrome. *The Lancet Respiratory Medicine*, 10(11): 1086–1098, 2022.

Gerald Woo, Chenghao Liu, Akshat Kumar, Caiming Xiong, Silvio Savarese, and Doyen Sahoo. Unified training of universal time series forecasting transformers. PMLR, 2024.

Joo Heung Yoon, Jueun Kim, Theodore Lagattuta, Michael R Pinsky, Marilyn Hravnak, and Gilles Clermont. Early Physiologic Numerical and Waveform Characteristics of Simulated Hemorrhagic Events With Healthy Volunteers Donating Blood. *Critical Care Explorations*, 6(4):e1073, 2024.

Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. How transferable are features in deep neural networks? *Proceddings of Advances in neural information processing systems*, 27, 2014.

Xiao Zhang, Srinath Sridharan, Nur Hakim Bin Zahrin, Narayan Venkataraman, and Goh Siang Hiong. Benchmarking a Time-Series Foundation Model (TimeGPT) for Real-World Forecasting Applications. *Machine Learning with Applications*, pp. 100801, 2025.

Yiyi Zhang, Wendy S Post, Elena Blasco-Colmenares, Darshan Dalal, Gordon F Tomaselli, and Eliseo Guallar. Electrocardiographic QT interval and mortality: a meta-analysis. *Epidemiology*, 22(5):660–670, 2011.

A Appendix

A.1 Experiment details

A.1.1 Datasets

We evaluate the proposed method using both synthetic and real-world benchmark datasets to examine its effectiveness under noise-free and noisy conditions. Synthetic ECG, PPG, and Resp datasets are generated using `NeuroKit2` (Makowski et al., 2021). For each modality, heart rate values are uniformly sampled from the range of 30 to 150 beats per minute, and corresponding physiological signal waveforms are generated conditioned on the selected heart rate. For real-world benchmark datasets, we use the MIT-BIH arrhythmia dataset for ECG signals (Moody & Mark, 2001; Goldberger et al., 2000) and the BID-MC dataset for PPG and Resp signals (Pimentel et al., 2016; Goldberger et al., 2000).

For both synthetic and benchmark datasets, the signal lengths were determined based on the minimum requirements of the feature extraction packages. Specifically, fixed-length windows of 5 seconds for ECG and 15 seconds for PPG and respiration signals were used. Signals were directly generated at these fixed lengths for synthetic datasets, whereas real-world records were segmented into fixed-length windows for benchmark datasets. MIT-BIH dataset was segmented into 5 second windows, resulting in 17,328 data points, and the sampling rate is 360 Hz. The BID-MC dataset

was segmented into 15 second windows, yielding 1,696 segments, with a sampling rate of 125 Hz. For the synthetic datasets, the sampling rate was set to match that of the corresponding benchmark dataset for each signal type, and each dataset consisted of 5,000 data points.

The ground-truth feature values for the ECG and Resp signals are computed using `the NeuroKit2` toolkit (Makowski et al., 2021) by detecting biomarkers and then computing derived physiological features based on these detected markers. For PPG signals, physiological features are extracted using the `pyPPG` toolkit (Goda et al., 2024). Features not directly supported by `pyPPG` were either computed through additional calculations or approximated using closely related features. Specifically, *Crest time* is replaced by the *u-point time* provided by `pyPPG`, which is defined as the time interval between the pulse onset and u-point. The u-point corresponds to the point of maximum amplitude between the pulse onset and the systolic peak in the first derivative of the PPG signal. In addition, the feature $(b-e)/a$ is computed as $b/a - e/a$, and $(b-c-d)/a$ is computed as $b/a - c/a - d/a$. Both the MIT-BIH and BID-MC dataset provide annotated heart rate values. Additionally, the BID-MC dataset includes annotations for *Oxygen saturation*, *Respiratory rate*, and *Pulse rate*, which were also utilized in this study.

A.1.2 Implementation details

We use MOMENT-small (Goswami et al., 2024) as a TSFM and extract embeddings from an intermediate layer, specifically layer 4. Recent studies report that deeper layers in modern TSFMs tend to specialize in the original pre-training objective and may lose generalizability (Alkin et al., 2024; Yosinski et al., 2014). In contrast, intermediate layers of TSFMs better preserve morphological and semantic features (Han & Lee, 2025).

We employ a set of regressors, including Multilayer Perceptrons (MLP), Partial Least Squares (PLS), Ridge, Lasso, and Random Forest regressor to extract physiological features from both TSFM embeddings and raw signals. Hyperparameters for all regressors are optimized via grid search. We utilized the sample-wise mean and standard deviation stored during the RevIN normalization process when extracting embeddings with MOMENT-small.

The dataset was partitioned into training, validation, and test sets with a ratio of 8:1:1 randomly. All experiments are conducted three times with different random seeds, and the reported results are averaged across runs.

A.2 Results

This section reports the results in terms of relative RMSE, defined as the ratio between the RMSE obtained when features are extracted from MOMENT embeddings and the RMSE obtained when same features are extracted directly from the raw signals using the same regressor in Table 3. The definitions of the features in our study are provided in Table 5, and the experimental details are described in Ablation A.1.

Table 3: Comparison of relative RMSE across features and regressors. Values less than 1 indicate improved performance relative to the raw-signal feature extraction and are highlighted in bold. Red indicates periodicity-based features. Green denotes morphology-based features derived from raw signals, whereas teal represents morphology-based features extracted from the second-derivative of the signals. Blue indicates amplitude-based features.

Signal	Datasets	Features	Lasso	MLP	PLS	RF	Ridge
ECG	Synthetic	Heart rate	**0.526**	**0.344**	**0.380**	**0.612**	**0.245**
		HRV-sdnn	**0.815**	1.128	**0.811**	**0.955**	**0.999**
		HRV-rmssd	**0.789**	1.004	**0.788**	**0.890**	**0.961**
		HRV-pnn20	**0.971**	**0.969**	**0.963**	**0.987**	1.019
		HRV-pnn50	**0.938**	**0.908**	**0.935**	**0.973**	1.051
		HRV-LF power	**1.000**	**0.000**	1.031	**1.000**	**0.753**
		HRV-HF power	**1.000**	**0.049**	1.030	**0.969**	**0.893**
		HRV-LF/HF ratio	**0.912**	**0.977**	**0.914**	**0.978**	**0.412**
		RR interval	**0.683**	**0.509**	**0.530**	**0.627**	**0.519**

Continued on next page

Signals	Datasets	Features	Lasso	MLP	PLS	RF	Ridge
		Sample entropy	**0.697**	**0.718**	**0.577**	**0.598**	**0.400**
		PR interval	**0.736**	**0.737**	**0.760**	**0.913**	**0.802**
		QT intervals	**0.706**	**0.801**	**0.795**	**0.846**	**0.860**
		QRS width	**0.740**	**0.838**	**0.874**	**0.894**	**0.855**
		ST segment slope	**0.671**	**0.689**	**0.959**	**0.975**	1.018
		RS amp diff	**0.958**	1.068	1.073	1.088	1.082
	MIT-BIH	Heart rate	**0.482**	**0.507**	**0.480**	**0.675**	**0.475**
		HRV-sdnn	**0.778**	**0.756**	**0.777**	**0.786**	**0.768**
		HRV-rmssd	**0.773**	**0.766**	**0.772**	**0.797**	**0.762**
		HRV-pnn20	**0.818**	**0.824**	**0.820**	**0.862**	**0.809**
		HRV-pnn50	**0.706**	**0.696**	**0.706**	**0.780**	**0.698**
		HRV-LF power	**1.000**	1.043	**0.911**	**0.922**	**0.910**
		HRV-HF power	**1.000**	**0.217**	**0.905**	**0.913**	**0.902**
		HRV-LF/HF ratio	**1.000**	1.001	**1.000**	1.007	**0.999**
		RR interval	**0.689**	**0.707**	**0.623**	**0.752**	**0.618**
		Sample entropy	**0.662**	**0.788**	**0.583**	**0.761**	**0.583**
		PR interval	**0.852**	**0.791**	**0.815**	**0.837**	**0.806**
		QT interval	**0.823**	**0.779**	**0.795**	**0.826**	**0.786**
		QRS width	**0.718**	**0.687**	**0.605**	**0.754**	**0.599**
		ST segment slope	**1.000**	**0.909**	**0.815**	**0.892**	**0.800**
		RS amp diff	**0.344**	**0.630**	**0.316**	**0.437**	**0.311**
PPG	Synthetic	Heart rate	**0.269**	**0.321**	**0.213**	1.163	**0.210**
		Pulse interval	1.091	**0.912**	**0.918**	1.068	**0.917**
		Time-to-peak	1.081	**0.897**	**0.907**	1.069	**0.905**
		Crest time	1.081	**0.942**	**0.934**	1.048	**0.931**
		ΔT	1.030	**0.970**	**0.963**	1.004	**0.956**
		Systolic width-50	1.058	**0.995**	**0.981**	1.030	**0.979**
		Diastolic width-50	1.043	**0.949**	**0.951**	**0.997**	**0.959**
		Pulse width-50	1.063	**0.940**	**0.958**	1.003	**0.965**
		Total pulse area	**0.999**	**1.000**	**0.925**	**0.786**	**0.913**
		Systolic area	1.007	**1.000**	1.014	1.196	1.030
		Diastolic area	**1.000**	**1.000**	**0.983**	1.020	**0.985**
		Slope	1.208	**0.942**	**0.951**	1.077	**0.947**
		PAI	**1.000**	**0.999**	**0.977**	1.094	**0.973**
		b/a	1.334	**0.730**	**0.956**	1.041	**0.912**
		c/a	1.017	**0.986**	**0.962**	**0.993**	**0.978**
		d/a	1.005	**0.991**	**0.998**	**0.999**	**0.987**
		e/a	1.019	**0.982**	**0.967**	1.008	**0.980**
		$(b-c-d-e)/a$	**0.995**	**0.975**	**0.965**	1.013	**0.986**
		$(b-e)/a$	1.010	**0.912**	**0.962**	1.036	**0.977**
		$(b-c-d)/a$	1.004	1.014	**0.991**	**0.995**	**0.987**
		PAV	**0.950**	**0.994**	**0.940**	**0.896**	**0.914**
		Systolic amp	1.021	**0.843**	**0.876**	**0.999**	**0.882**
		Diastolic amp	1.018	**0.967**	**0.921**	1.021	**0.923**
	BID-MC	SpO_2	**0.632**	**0.264**	**0.636**	**0.681**	**0.350**
		Heart rate	**0.432**	**0.121**	**0.425**	**0.594**	**0.329**
		Resp rate	**0.577**	**0.549**	**0.579**	**0.538**	**0.369**
		Pulse ratae	**0.458**	**0.214**	**0.452**	**0.608**	**0.367**
		Pulse interval	**0.550**	**0.551**	**0.529**	**0.761**	**0.339**
		Time-to-peak	**0.541**	**0.431**	**0.517**	**0.767**	**0.324**
		Crest time	**0.703**	**0.688**	**0.638**	**0.832**	**0.473**
		ΔT	**0.633**	**0.586**	**0.603**	**0.682**	**0.497**
		Systolic width-50	**0.605**	**0.627**	**0.545**	**0.761**	**0.393**
		Diastolic width-50	**0.761**	**0.701**	**0.760**	**0.849**	**0.468**
		Pulse width-50	**0.690**	**0.706**	**0.689**	**0.822**	**0.421**
		Total pulse area	1.001	**0.997**	1.003	1.076	**0.829**

Continued on next page

Signals	Datasets	Features	Lasso	MLP	PLS	RF	Ridge
		Systolic area	**1.000**	**0.995**	1.001	1.059	**0.797**
		Diastolic area	1.001	**0.999**	1.004	1.099	**0.832**
		Slope	**0.674**	**0.556**	**0.683**	**0.945**	**0.383**
		PAI	**0.987**	**1.000**	1.005	**0.968**	**0.834**
		b/a	**0.992**	**0.602**	1.411	1.145	**0.271**
		c/a	**0.921**	**0.791**	**0.911**	1.533	**0.284**
		d/a	1.002	**0.831**	1.210	2.515	**0.425**
		e/a	**0.978**	**0.808**	1.619	1.248	**0.319**
		$(b-c-d-e)/a$	1.035	**0.817**	1.684	**0.835**	**0.291**
		$(b-e)/a$	1.124	**0.797**	1.610	**0.800**	**0.292**
		$(b-c-d)/a$	**0.993**	**0.935**	1.072	1.125	**0.284**
		PAV	**0.995**	**0.948**	1.226	**0.561**	**0.210**
		Systolic amp	**0.780**	**0.390**	1.028	**0.431**	**0.701**
		Diastolic amp	**0.896**	**0.650**	**0.874**	**0.871**	**0.657**
Resp	Synthetic	Heart rate	**0.383**	**0.652**	**0.386**	**0.477**	**0.394**
		Expiration duration	**0.982**	**0.938**	**0.981**	**0.994**	**0.970**
		Inspiration duration	**0.964**	**0.957**	**0.983**	**0.984**	**0.973**
		I/E ratio	**0.996**	**0.920**	**0.990**	**0.993**	**0.992**
	BID-MC	Oxygen saturation (SpO_2)	**0.848**	**0.456**	**0.847**	**0.785**	**0.478**
		Heart rate	**0.791**	**0.726**	**0.798**	**0.837**	**0.325**
		Resp rate	**0.726**	**0.756**	**0.719**	**0.895**	**0.498**
		Pulse rate	**0.730**	**0.645**	**0.733**	**0.824**	**0.385**
		Expiration duration	**0.921**	**0.917**	**0.839**	1.024	**0.667**
		Inspiration duration	**0.719**	**0.825**	**0.844**	**0.994**	**0.445**
		I/E ratio	**0.980**	1.017	**0.986**	1.031	**0.955**

A.3 Impact of Amplitude Information in Embedding Representations

We analyze the effect of global scale statistics, namely mean and standard deviation, on MOMENT embeddings. To this end, we compare model performance with and without incorporating mean and standard deviation information. The comparative results are summarized in Table 4. We employ relative RMSE as an evaluation metric, defined as the ratio of the RMSE obtained with mean and standard deviation information to that obtained without this information.

Table 4: Comparison of relative RMSE with and without mean and standard deviation across features and regressors. Values less than 1 indicate improved performance relative to the raw-signal feature extraction and are highlighted in bold. Red indicates periodicity-based features. Green denotes morphology-based features derived from the raw PPG signal, whereas teal represents morphology-based features extracted from the second-derivative PPG signal. Blue indicates amplitude-based features.

Signal	Datasets	Features	Lasso	MLP	PLS	RF	Ridge
ECG	Synthetic	Heart rate	**1.000**	**0.950**	1.003	1.003	1.001
		HRV-sdnn	1.001	1.039	**1.000**	**1.000**	**0.999**
		HRV-rmssd	1.001	**0.976**	**0.999**	**0.994**	**0.999**
		HRV-pnn20	**1.000**	**1.000**	**1.000**	**0.991**	**0.999**
		HRV-pnn50	**0.999**	**0.957**	**0.998**	**0.996**	**0.998**
		HRV-LF power	**1.000**	**0.068**	**1.000**	**1.000**	**1.000**
		HRV-HF power	**1.000**	**1.000**	**1.000**	1.003	**0.999**
		HRV-LF/HF ratio	**1.000**	1.003	**1.000**	1.035	**1.000**
		RR interval	**1.000**	1.028	1.003	**0.996**	1.001
		Sample entropy	**1.000**	1.055	**1.000**	**0.998**	**0.998**
		PR interval	**1.000**	**0.627**	1.002	**0.996**	**1.000**
		QT interval	**1.000**	**0.956**	1.002	**0.996**	1.001
		QRS width	**1.000**	**0.992**	1.004	**0.998**	**1.000**

Continued on next page

Signals	Datasets	Features	Lasso	MLP	PLS	RF	Ridge
		ST segment slope	**1.000**	**0.544**	**0.998**	**0.984**	**0.999**
		RS amp diff	**1.000**	1.053	**0.747**	**0.796**	**0.828**
	MIT-BIH	Heart rate	**0.963**	**0.953**	**0.964**	**0.930**	**0.965**
		HRV-sdnn	**0.993**	**0.997**	**0.993**	**0.948**	**0.993**
		HRV-rmssd	**0.994**	**0.989**	**0.994**	**0.959**	**0.994**
		HRV-pnn20	**0.993**	1.006	**0.994**	**0.978**	**0.994**
		HRV-pnn50	**0.990**	**0.982**	**0.990**	**0.975**	**0.990**
		HRV-LF power	**1.000**	**0.247**	**0.999**	1.007	**0.999**
		HRV-HF power	**1.000**	**0.821**	**0.998**	1.001	**0.998**
		HRV-LF/HF ratio	**1.000**	**1.000**	**1.000**	**0.993**	**1.000**
		RR interval	**0.978**	**0.988**	**0.985**	**0.934**	**0.984**
		Sample entropy	**0.988**	1.022	**0.987**	**0.939**	**0.986**
		PR interval	**0.997**	**0.994**	**0.998**	**0.989**	**0.998**
		QT interval	**0.995**	**0.996**	**0.996**	**0.967**	**0.995**
		QRS width	**1.000**	1.004	**0.995**	**0.980**	**0.995**
		ST segment slope	**1.000**	**0.744**	**0.994**	**0.983**	**0.995**
		RS amp diff	**0.635**	**0.775**	**0.632**	**0.595**	**0.633**
PPG	Synthetic	Heart rate	**0.999**	1.084	**0.991**	**0.997**	**0.991**
		Pulse interval	**1.000**	**0.978**	1.001	**0.999**	1.001
		Time-to-peak	**1.000**	**0.986**	1.001	1.001	**1.000**
		Crest time	**1.000**	**0.995**	1.001	**0.999**	1.001
		ΔT	**1.000**	**0.986**	**1.000**	**1.000**	**1.000**
		Systolic width-50	**1.000**	**0.992**	1.001	**1.000**	1.001
		Diastolic width-50	**1.000**	**0.995**	1.001	**0.999**	**1.000**
		Pulse width-50	**1.000**	1.017	1.001	**0.999**	1.001
		Total pulse area	**1.000**	**1.000**	**1.000**	1.014	**1.000**
		Systolic area	**1.000**	**1.000**	**1.000**	**0.995**	**1.000**
		Diastolic area	**1.000**	**1.000**	**1.000**	**0.986**	**1.000**
		Slope	**1.000**	**0.990**	**0.997**	**1.000**	**0.998**
		PAI	**1.000**	**1.000**	**1.000**	1.005	**1.000**
		b/a	**1.000**	**0.993**	**1.000**	**1.000**	**1.000**
		c/a	**1.000**	**0.971**	**0.999**	**1.000**	**1.000**
		d/a	**1.000**	**0.973**	1.001	**1.000**	**1.000**
		e/a	**1.000**	1.001	**1.000**	**1.000**	**1.000**
		$(b-c-d-e)/a$	**1.000**	**1.000**	**1.000**	**0.999**	**1.000**
		$(b-e)/a$	**1.000**	1.001	**1.000**	1.001	**1.000**
		$(b-c-d)/a$	**1.000**	1.042	1.001	**1.000**	**1.000**
		PAV	**1.000**	1.002	**1.000**	**0.987**	**1.000**
		Systolic amp	**1.000**	**0.983**	**0.950**	**0.976**	**0.961**
		Diastolic amp	**1.000**	1.062	**0.990**	**0.995**	**0.995**
	BID-MC	SpO_2	**0.998**	1.115	**0.999**	**0.937**	**0.998**
		Heart rate	**0.985**	**0.999**	**0.970**	**0.957**	**0.982**
		Resp rate	1.010	**0.869**	1.015	**0.898**	1.004
		Pulse rate	**0.987**	**0.995**	**0.985**	**0.900**	**0.990**
		Pulse interval	**0.976**	1.031	**0.992**	**0.984**	**0.994**
		Time-to-peak	**0.970**	**0.968**	**0.986**	**0.993**	**0.989**
		Crest time	**0.988**	**0.982**	**0.993**	**0.999**	**0.992**
		ΔT	**0.991**	**0.994**	**0.984**	**0.981**	**0.995**
		Systolic width-50	1.001	1.065	1.002	1.002	1.002
		Diastolic width-50	**1.000**	**0.991**	1.003	**0.988**	1.001
		Pulse width-50	**1.000**	1.010	1.002	**0.995**	**1.000**
		Total pulse area	**1.000**	**1.000**	**0.998**	**0.993**	**1.000**
		Systolic area	**0.999**	1.001	**0.997**	**1.000**	**1.000**
		Diastolic area	**1.000**	1.001	**0.999**	**0.990**	**1.000**
		Slope	**0.999**	**0.951**	**0.988**	**0.962**	**0.997**
		PAI	**1.000**	**1.000**	**1.000**	**0.990**	**1.000**

Continued on next page

Signals	Datasets	Features	Lasso	MLP	PLS	RF	Ridge
		b/a	**1.000**	1.091	1.001	1.007	1.013
		c/a	**0.995**	**0.990**	**1.000**	**0.996**	**0.998**
		d/a	**1.000**	1.013	**1.000**	1.062	1.003
		e/a	**1.000**	**0.990**	**1.000**	1.030	1.011
		$(b-c-d-e)/a$	**1.000**	1.023	1.010	1.009	1.012
		$(b-e)/a$	**1.000**	1.007	1.001	**0.965**	1.012
		$(b-c-d)/a$	**1.000**	1.042	1.004	1.010	1.001
		PAV	**1.000**	**0.989**	**0.967**	**0.993**	1.001
		Systolic amp	**0.386**	**0.600**	**0.410**	**0.173**	**0.419**
		Diastolic amp	**0.966**	1.034	**0.969**	**0.914**	**0.969**
Res	Synthetic	Heart rate	**0.876**	**0.994**	**0.876**	**0.710**	**0.900**
		Expiration duration	**1.000**	**0.996**	**1.000**	**0.999**	**1.000**
		Inspiration duration	**1.000**	**0.999**	**1.000**	1.001	**1.000**
		I/E ratio	**1.000**	**0.999**	**1.000**	**0.999**	**1.000**
	BID-MC	Oxygen saturation (SpO_2)	**0.987**	1.099	**0.986**	**0.794**	**0.985**
		Heart rate	**0.977**	1.005	**0.979**	**0.888**	**0.979**
		Resp rate	**0.994**	1.040	**0.993**	**0.858**	**0.998**
		Pulse rate	**0.993**	1.004	**0.989**	**0.897**	**0.991**
		Expiration duration	1.009	1.019	1.008	**0.996**	1.011
		Inspiration duration	1.002	**0.989**	1.005	**0.990**	**0.996**
		I/E ratio	**0.999**	1.042	**1.000**	**0.999**	1.001

A.4 Backgrounds

A.4.1 Time-Series Foundation Models

Time-series foundation models (TSFMs) have recently emerged as a powerful paradigm for learning generalizable representations from large-scale, heterogeneous time-series data. Through large-scale pretraining, TSFMs aim to capture universal characteristics of temporal patterns and learn representations that can be transferred across diverse signals, tasks and domains. TSFMs demonstrate strong empirical performance on a wide range of downstream applications, including forecasting, classification, and anomaly detection (Liang et al., 2024; Kottapalli et al., 2025).

Recently, TSFMs have been increasingly adopted in the medical domain (Li et al., 2025; Steinberg et al., 2023). In this context, TSFMs can be broadly categorized into two types: (1) models that leverage general-purpose models for downstream medical tasks, and (2) models specifically trained for particular medical tasks or domain-specific signals.

The first category applies general-purpose TSFMs to various medical tasks. TimesFM (Das et al., 2024) and MOIRAI (Woo et al., 2024) have been utilized to forecast the incidence rate of hand, foot, and mouth disease (Wang et al., 2025). TabPFN-TS (Hoo et al., 2024), Time-GPT (Garza et al., 2023), TimesFM (Das et al., 2024), Lag-Llama (Rasul et al., 2023), and Chronos (Ansari et al., 2024) have been utilized to forecast both short-term and long-term of infectious disease trends, including peak incidence timing (Kalahasti et al., 2025). Liu et al. (2024) adapted MOMENT by using learnable prompt vectors to the input sequence without full fine-tuning. Additionally, UniTS (Gao et al., 2024) has been applied to impute missing values in electronic health record data.

The second category focuses on training signal-specific TSFMs for various medical applications. GPT-PPG (Chen et al., 2025) fine-tunes a model specifically tailored to PPG signals by designing both architecture and loss functions and applied to atrial fibrillation detection, heart rate estimation, respiration rate estimation, and blood pressure estimation. ECGBERT (Choi et al., 2023) models ECG signals as a language by clustering P, QRS, and T waveforms into discrete tokens and applying a BERT-based architecture. It is used for atrial fibrillation detection and heartbeat classification.

A.4.2 Clinically Meaningful Physiological Features

This section focuses on physiological features with well-established clinical relevance in real-world healthcare settings and justifies their selection for this study. The importance of physiological signal

features varies depending on the disease type, diagnostic objective, and clinical context, making it challenging to define a unified feature set applicable across all clinical tasks. Therefore, we selected features whose clinical utility has been validated in prior literature for specific clinical scenarios, based on consultation with experienced clinicians.

Electrocardiography (ECG) is a widely used non-invasive modality to assess cardiovascular status and has been extensively applied for early detection and risk stratification of various diseases. Features extracted from ECG carry distinct physiological meanings depending on the clinical objective, and consequently, the types of features employed differ across applications.

Yoon et al. (2024) reported that autonomic nervous system responses, including increased heart rate and decreased heart rate variability (HRV), serve as key indicators for the early detection of acute hemorrhage in clinical settings. Accordingly, *Heart rate*, *RR interval*, time-domain HRV features including standard deviation of normal-to-normal intervals (*HRV-sdnn*), Root mean square of successive differences (*HRV-rmssd*), percentage of successive normal-to-normal (*HRV-pNN50*, *HRV-pNN20*), as well as frequency-domain HRV features such as *low-frequency (LF) power*, *high-frequency (HF) power* and *LF/HF ratio* have been reported as important features reflecting acute hemorrhagic states. *PR interval* reflects atrioventricular conduction status (Rasmussen et al., 2017), whereas *QRS duration* is a well-established marker of intraventricular conduction delay (Deen et al., 2017). *QT interval* has also been widely employed for cardiovascular risk assessment (Zhang et al., 2011). In addition, *ST segment slope* is used for the diagnosis of myocardial ischemia (Boyle et al., 1987), while *RS amplitude difference* has been utilized for arrhythmia diagnosis in ventricular depolarization (Park et al., 2015). *Sample entropy* has been used for the detection of chronic atrial fibrillation (Horie et al., 2018).

Photoplethysmography (PPG) is a non-invasive physiological signal that reflects changes in blood volume and vascular elasticity in peripheral vessels and has been widely used to assess circulatory status and vascular function. PPG signals are sensitive to hemodynamic changes and response of autonomic nervous system, and a variety of physiological features are utilized depending on the specific clinical objective.

PPG signals have been widely used to monitor fundamental physiological indicators, including *Respiratory rate*, *Pulse rate*, and *Oxygen saturation* (Tamura, 2019). In particular, heart rate derived from PPG signals has been utilized for cardiovascular disease prediction (Song et al., 2011). In addition, features derived from PPG have been widely applied in the diagnosis and assessment of various pathological conditions. Park et al. (2022) showed that various PPG-derived features, including *Systolic width*, *Diastolic width*, *Pulse width*, *Total pulse area*, *Systolic area*, *Diastolic area*, *Crest time*, ΔT, and *PPG augmentation index*, b/a, c/a, d/a, e/a, $(b-c-d-e)/a$, $(b-e)/a$, and $(b-c-d)/a$, have been widely used across a variety of physiological and clinical tasks. Moreover, *Pulse interval*, *Time-to-peak*, *Slope*, and *PPG amplitude variation* exhibit significant changes in the early stage of acute hemorrhage and serve as useful indicators for its detection (Yoon et al., 2024).

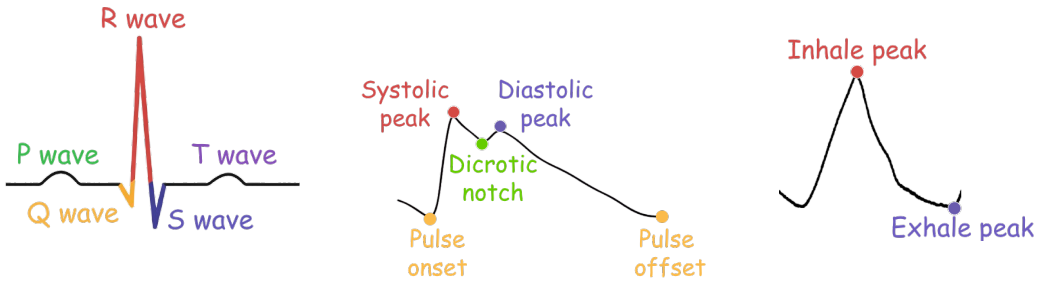
Respiratory (Resp) signals provide direct information about breathing patterns and respiratory function and are widely used to monitor cardiopulmonary status in clinical settings. Features extracted from Resp signals are used to assess autonomic nervous system regulation, stress levels, and respiratory control, and their physiological significance varies depending on the clinical context.

Respiratory rate has been utilized as an indicator for detecting acute hemorrhage (Yoon et al., 2024). *Oxygen saturation* is widely used as an important indicator for assessing respiratory distress (Wick et al., 2022). Furthermore, respiratory pattern features such as *Inspiration duration*, *Expiration duration*, and *I/E ratio* are commonly used in the diagnosis of respiratory diseases (Addison et al., 2024).

A.5 Description of physiological features

This section describes the physiological features used in this study, including their names, acronyms, and definitions. The physiological features are configured differently depending on the signal modality, such as ECG, PPG, and Resp signals. The rationale for selecting the feature sets is discussed in Ablation A.4.2.

Prior to describing the features presented in Table 5, Figure 3 illustrates the biomarker points extracted from the ECG, PPG, and Resp signals, respectively. These biomarker points form the basis



(a) ECG signal (b) PPG signal (c) Respiratory signal

Figure 3: Physiological signals used in this study: (a) ECG, (b) PPG, and (c) Resp. Representative biomarker points extracted are illustrated.

for deriving features from each signal type. Additionally, we define the biomarker points a, b, c, d, and e which represents the second derivative of the PPG signal (SDPTG). These points are defined as follows (Goda et al., 2024):

- a: Peak point between the pulse onset and the systolic peak on the SDPTG.
- b: First local minimum following the a-point on the SDPTG.
- c: First local maximum following the b-point on the SDPTG.
- d: Second local minimum following the c-point on the SDPTG.
- e: Second local maximum following the d-point on the SDPTG.

The extracted physiological features considered in this study are summarized in Table 5, and additional PPG biomarkers are described in detail in (Park et al., 2022). Extracted features are broadly categorized into four types: periodicity-based features, morphology-based features derived from the raw signal, morphology-based features derived from the second derivative signal, and amplitude-based features. A detailed mapping of features to their corresponding categories is provided in Table 6.

Table 5: Physiological feature descriptions from ECG, PPG, and Resp signals. Abbreviations in parentheses indicate feature acronyms. Red indicates periodicity-based features. Green denotes morphology-based features derived from the raw PPG signal, whereas teal represents morphology-based features extracted from the second-derivative PPG signal. Blue indicates amplitude-based features.

Signals	Features	Descriptions
ECG	Heart rate	Number of heart beats per minute.
	HRV-sdnn	Standard deviation of RR intervals.
	HRV-rmssd	Root mean square standard deviation of RR intervals.
	HRV-pNN50	Percentage of RR intervals that the difference exceeds 50 ms.
	HRV-pNN20	Percentage of successive RR intervals that the difference exceeds 20 ms.
	HRV-LF power	Low-frequency power of HRV.
	HRV-HF power	High-frequency power of HRV.
	HRV-LF/HF ratio	Ratio of low-frequency power to high-frequency power.
	RR interval	Average temporal interval between consecutive R-peaks.

Continued on next page

	Sample entropy	Degree of signal regularity based on temporal pattern similarity.
	PR interval	Average duration between P-onset to Q-onset.
	QT interval	Average duration from Q-onset to T-offset.
	QRS width	Average duration from Q-onset to S-offset.
	ST segment slope	Average slope of the ST segment.
	RS amplitude difference (RS amp diff)	Average amplitude difference between R-peaks and S-peaks.
PPG	Oxygen saturation (SpO_2)	Percentage of oxygen-saturated hemoglobin in blood.
	Heart rate	Number of heart beats per minute.
	Respiratory rate (Resp rate)	Number of respiration cycles per minute.
	Pulse rate	Number of pulse beats per minute.
	Pulse interval	Average duration from pulse onset to pulse offset.
	Time-to-peak	Average temporal interval between two consecutive systolic peaks.
	Crest time	Average duration from pulse onset to the first zero-crossing in the first derivative of the PPG waveform.
	ΔT	Average duration from systolic peak to diastolic peak.
	Systolic width-50	Pulse width at 50% of the systolic peak amplitude between pulse onset and systolic peak.
	Diastolic width-50	Pulse width at 50% of systolic peak amplitude between systolic peak and pulse offset.
	Pulse width-50	Sum of the systolic width-50 and the diastolic width-50.
	Total pulse area	Pulse wave area between pulse onset and pulse offset.
	Systolic area	Pulse wave between the pulse onset and the dicrotic notch .
	Diastolic area	Pulse wave between the dicrotic notch and pulse offset .
	Slope	Ratio of time to systolic peak to systolic peak amplitude.
	PPG augmentation index (PAI)	Ratio of the systolic peak amplitude to the diastolic peak amplitude.
	b/a	Ratio of the amplitude of b-point to a-point of SDPTG.
	c/a	Ratio of the amplitude of c-point to a-point of SDPTG.
	d/a	Ratio of the amplitude of d-point to a-point of SDPTG.
	e/a	Ratio of the amplitude of e-point to a-point of SDPTG.
	$(b-c-d-e)/a$	Composite amplitude ratio defined as b/a - c/a - d/a - e/a of SDPTG.
	$(b-e)/a$	Composite amplitude ratio defined as b/a - e/a of SDPTG.
	$(b-c-d)/a$	Composite amplitude ratio defined as b/a - c/a - d/a of SDPTG.
	PPG amplitude variation (PAV)	Degree of fluctuation in pulse amplitude.

Continued on next page

	Systolic amplitude (Systolic amp)	Amplitude difference between the pulse onset and systolic peak.
	Diastolic amplitude (Diastolic amp)	Amplitude difference between the pulse onset and diastolic peak.
Resp	Oxygen saturation (SpO_2)	Percentage of oxygen-saturated hemoglobin in blood.
	Heart rate	Number of heart beats per minute.
	Respiratory rate (Resp rate)	Number of respiration cycles per minute.
	Pulse rate	Number of pulse beats per minute.
	Expiration duration	Average duration from inhale-peak to exhale-peak.
	Inspiration duration	Average duration from exhale-peak to inhale-peak.
	I/E ratio	Ratio of Inspiration duration to Expiration duration.

Table 6: Extracted features classified into four categories: periodicity-based features, morphology-based features in the raw signal, morphology-based features in the second derivative signal, and amplitude-based features.

Signals	Feature types	Features
ECG	Periodicity-based	Heart rate, HRV-sdnn, HRV-rmssd, HRV-pnn20, HRV-pnn50, LF power, HF power, LF/HF ratio, RR interval, Sample entropy.
	Morphology-based (raw)	PR interval, QT interval, QRS width, ST segment slope.
	Amplitude-based	RS amp diff.
PPG	Periodicity-based	Heart rate, Resp rate, Pulse rate, Pulse interval, Time-to-peak, Crest time, ΔT.
	Morphology-based (raw)	Systolic width-50, Diastolic width-50, Pulse width-50, Total pulse area, Systolic area, Diastolic area, Slope, PAI.
	Morphology-based (2^{nd} deriv.)	b/a, c/a, d/a, e/a, $(b-c-d-e)/a$, $(b-e)/a$, $(b-c-d)/a$.
	Amplitude-based	PAV, Systolic area, Diastolic area.
Resp	Periodicity-based	Heart rate, Pulse interval, Time-to-peak, Crest time, ΔT.
	Morphology-based (raw)	Systolic width-50, Diastolic width-50, Pulse width-50, Total pulse area, Systolic area, Diastolic area, Slope, PAI.
	Morphology-based (2^{nd} deriv.)	b/a, c/a, d/a, e/a, $(b-c-d-e)/a$, $(b-e)/a$, $(b-c-d)/a$.