ECG-MoE: Mixture-of-Expert Electrocardiogram Foundation Model

Yuhao Xu
Department of Computer Science
Emory University
Atlanta, GA 30307
yxu81@emory.edu

Xiaoda Wang
Department of Computer Science
Emory University
Atlanta, GA 30307
xiaoda.wang@emory.edu

Yi Wu
School of Computer Science
University of Oklahoma
Norman, OK 73019
yi.wu-1@ou.edu

Wei Jin
Department of Computer Science
Emory University
Atlanta, GA 30307
wei.jin@emory.edu

Xiao Hu
Center of Data Science
Emory University
Atlanta, GA 30307
xiao.hu@emory.edu

Carl Yang
Department of Computer Science
Emory University
Atlanta, GA 30307
j.carlyang@emory.edu

Abstract

Electrocardiography (ECG) analysis is crucial for cardiac diagnosis, yet existing foundation models often fail to capture the periodicity and diverse features required for varied clinical tasks. We propose ECG-MoE, a hybrid architecture that integrates multi-model temporal features with a cardiac period-aware expert module. Our approach uses a dual-path Mixture-of-Experts to separately model beat-level morphology and rhythm, combined with a hierarchical fusion network using LoRA for efficient inference. Evaluated on five public clinical tasks, ECG-MoE achieves state-of-the-art performance with 40% faster inference than multi-task baselines.

1 Instruction

Cardiovascular diseases remain the leading cause of global mortality [1]. Electrocardiography (ECG) is a primary non-invasive tool for cardiac assessment, with increasing relevance due to advances in wearable monitoring [2, 3]. ECG interpretation relies on waveform-feature-disease correlations, where specific morphologies like QT prolongation or ST elevation indicate particular pathologies [4, 5, 6, 7]. These features are often phase-localized within the periodic P-QRS-T cycle [8, 9].

Foundation models have shown promise in automated ECG diagnosis but struggle to comprehensively represent clinically nuanced, segment-specific features [10]. In particular, existing foundation models have limitations because different models capture distinct ECG features, hindering multi-task performance within a single model. Furthermore, despite the strong periodicity inherent in ECG signals, this characteristic is often overlooked by existing ECG foundation models. They also incur high computational costs, hindering clinical adoption. Moreover, they often fail to align with ECG's inherent periodicity, where each heartbeat carries distinct diagnostic information [11, 12, 13], making it difficult to leverage rhythmic consistency while accommodating pathological variations [14, 15].
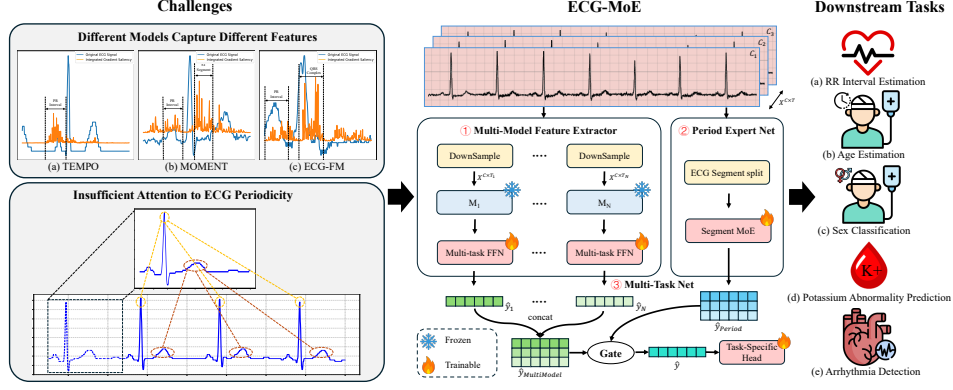
Figure 1: Existing foundation models have limitations because different models capture distinct ECG features, hindering multi-task performance within a single model. Furthermore, despite the strong periodicity inherent in ECG signals, this characteristic is often overlooked by existing ECG foundation models. To address these issues, we propose an ensemble learning method based on Period MoE. The effectiveness of our approach is validated across five common downstream tasks.

To address these issues, as shown in Figure 1, we introduce ECG-MoE, a hybrid architecture that integrates multi-model temporal features [16] with a dual-path Mixture-of-Experts (MoE) using task-conditioned gating [17]. Specifically, we propose an ensemble learning method based on Period MoE. We employ Low-Rank Adaptation (LoRA) for parameter-efficient fusion [18, 19]. The effectiveness of our approach is validated across five common downstream tasks. Validated on five clinical tasks, ECG-MoE achieves state-of-the-art performance with 40% faster inference and improved robustness.

2 Related Works

Electrocardiogram Foundation Models. Recent deep learning advances have developed ECG foundation models using CNNs [20] and Transformers [16] for tasks like arrhythmia detection. While pretraining and self-supervision [21] learn general features, these approaches often treat ECGs as generic time-series [22, 23, 24], neglecting the inherent periodicity critical to cardiac electrophysiology [11, 14]. This results in suboptimal handling of phase-localized abnormalities and inefficient rhythm modeling, which is a gap our period-conditioned framework directly addresses.

ECG Periodicity Modeling. ECG's quasi-periodicity has been exploited via segmentation [15] and rhythm-tracking methods, yet these struggle with pathological variations and holistic modeling. Although RNNs [25] capture temporal dependencies, they are inefficient and sensitive to period variability. No current framework incorporates periodicity as a core inductive bias for foundation models, despite its diagnostic importance [8, 13].

Multi-Task ECG Analysis. Multi-task learning (MTL) addresses simultaneous cardiac condition detection, but shared-backbone approaches [26] suffer from interference, especially in complex diagnostics like ST-segment analysis [7]. Modular or parameter-efficient designs improve scalability but ignore ECG's phase-specific biomarkers. Existing models fail to combine task-specific feature extraction with periodicity constraints. ECG-MoE overcomes this via hierarchical fusion and task-conditioned gating, enabling efficient and accurate multi-task inference.

3 Method

As shown in Figure 2, our proposed ECG analysis framework employs a multi-branch architecture that synergistically integrates heterogeneous feature representations while explicitly modeling cardiac periodicity. The model comprises three modules, each with distinct mathematical formulations:

Multi-Model Feature Extraction. Our framework integrates five time-series foundation models to capture diverse ECG characteristics. For input $\mathbf{X} \in \mathbb{R}^{C \times T}$, each model extracts features after
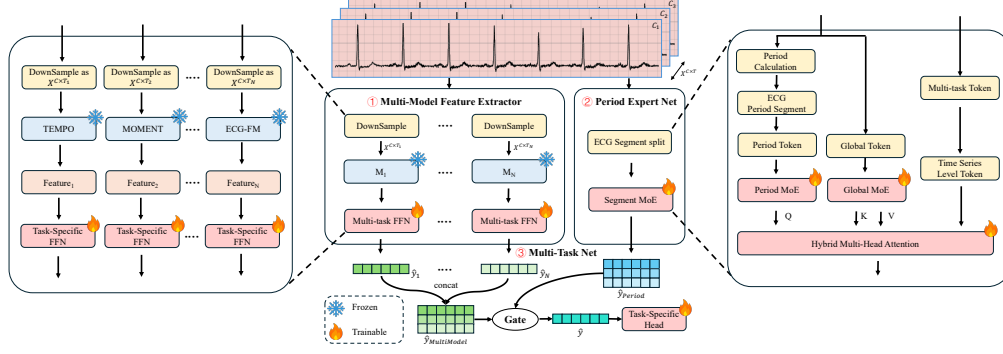
Figure 2: ECG-MoE framework: ① Multi-model feature extraction with adaptive downsampling, ② Period-aware gating weight generation via specialized MoE, ③ Multi-Task fusion using LoRA.

adaptive downsampling:

$$\mathbf{F}k = f\theta_k(\mathrm{DS}_k(\mathbf{X})) \tag{1}$$

Features are projected and concatenated into a unified representation:

$$\mathbf{F}m = \bigoplus k \in \mathcal{M}(\mathbf{W}_k\mathbf{F}_k + \mathbf{b}_k) \tag{2}$$

This ensemble approach captures complementary temporal patterns from different model architectures, enhancing feature diversity.

Periodic Expert Network. We design a dual-path MoE architecture to model ECG periodicity. R-peak detection segments the signal into individual heartbeats, which are normalized to fixed length. Three CNN experts with different kernel sizes process morphological features within beats, while two dilated CNNs capture inter-beat rhythm patterns. A task-conditioned gating mechanism dynamically weights expert contributions based on global signal characteristics and task embeddings:

$$\mathbf{g}_p = \mathrm{softmax}(\mathbf{U}_p[\bar{\mathbf{X}} \oplus \mathbf{e}t]) \tag{3}$$

Features are integrated via multi-head attention to produce the final periodic representation $\mathbf{F}$periodic.

Multi-Task Fusion. A hierarchical gating network combines multi-model and periodic features based on task requirements. The gating weights $\hat{\alpha}_m$ and $\hat{\alpha}p$ are learned through task-conditioned networks:

$$\mathbf{F}\text{fused} = \hat{\alpha}m\mathbf{F}m \oplus \hat{\alpha}p\mathbf{F}\text{periodic} \tag{4}$$

Task-specific heads then generate predictions from the fused features. The model is optimized with a composite loss that combines task-specific objectives with contrastive regularization:

$$\mathcal{L} = \sum_{k=1}^{K} \mathcal{L}_{\text{task}_k} + \lambda\mathcal{L}_{\text{cont}} \tag{5}$$

This design enables efficient parameter sharing while maintaining specialized processing for diverse clinical tasks through end-to-end training.

4 Experiments

In this section, we evaluate the effectiveness of our proposed model from two key perspectives: performance and efficiency. Specifically, we compare its predictive accuracy across multiple tasks and analyze its computational efficiency in terms of training and inference costs. By examining these aspects, we aim to demonstrate the advantages of our model in achieving a balanced trade-off between accuracy and computational feasibility.

4.1 Dataset and Baselines

We utilize the MIMIC-IV-ECG [27] datasets. MIMIC-IV-ECG is currently the largest publicly released ECG repository, which contains 800,035 diagnostic electrocardiograms acquired from 161,352 unique patients. We select TimesNet [28], DLinear [29], MOMENT [30], TEMPO [31] and ECG-FM [32] as baselines and construct an ensemble of them.

3

4.2 Research Questions and Results

RQ1: Model Selection Rationale. Why were specific foundation models selected, and what benefits does their ensemble provide?

Table 1: Zero-shot performance of baseline models on ECG data. Highlighted are the top first and second results. RR Interval Estimation (RR), Age Estimation (Age), Sex Classification (Sex), Potassium Abnormality Prediction (Ka), Arrhythmia Detection (AD).

		TimesNet	DLinear	MOMENT	TEMPO	ECG-FM
Regression (MAE↓)	RR	817.0 ± 2.5	816.4 ± 2.9	816.6 ± 2.1	816.3 ± 1.9	816.3 ± 1.9
	Age	62.28 ± 0.36	62.63 ± 0.50	62.61 ± 0.37	62.33 ± 0.38	62.27 ± 0.38
Binary Class (F1↑)	Sex	0.60 ± 0.00	0.51 ± 0.08	0.34 ± 0.00	0.42 ± 0.00	0.33 ± 0.00
	Ka	0.06 ± 0.00	0.05 ± 0.00	0.02 ± 0.00	0.18 ± 0.00	0.35 ± 0.00
15 Class (ACC↑)	AD	0.06 ± 0.01	0.03 ± 0.02	0.03 ± 0.02	0.02 ± 0.00	0.07 ± 0.00

We evaluate SOTA time-series models based on architectural diversity and ECG applicability through zero-shot evaluation on 10,000 patient samples.

Table 1 shows zero-shot performance of baseline models. We can get the key findings of ECG-FM excels in clinical diagnostics with domain-specific pretraining. TimesNet captures morphological variations effectively. DLinear shows competence in temporal feature modeling. MOMENT underperforms without domain adaptation.

These results validate our ensemble strategy, combining strengths of different architectures.

RQ2: Performance Benchmarking. Does ECG-MoE outperform existing models on ECG analysis?

We conduct comparative evaluation across five clinical tasks using 10,000 patients. As shown in Table 2, ECG-MoE achieves SOTA performance.

Table 2: Performance of fine-tuned baseline models on five ECG downstream tasks. Bold values represent the best performance.

		TimesNet	DLinear	MOMENT	TEMPO	ECG-FM	ECG-MoE
Regression (MAE↓)	RR	304.3 ± 4.3	786.0 ± 5.4	146.9 ± 1.3	141.5 ± 2.1	147.3 ± 1.3	$\mathbf{76.37 \pm 4.7}$
	Age	24.89 ± 0.07	28.46 ± 0.74	13.41 ± 0.45	13.52 ± 0.31	13.49 ± 0.17	$\mathbf{12.83 \pm 0.42}$
Binary Class (F1↑)	Sex	0.51 ± 0.05	0.57 ± 0.01	$\mathbf{0.69 \pm 0.02}$	0.54 ± 0.01	0.52 ± 0.05	$\mathbf{0.69 \pm 0.01}$
	Ka	0.41 ± 0.01	0.43 ± 0.01	0.49 ± 0.00	0.50 ± 0.00	0.49 ± 0.00	$\mathbf{0.57 \pm 0.00}$
15 Class (ACC↑)	AD	0.17 ± 0.00	0.48 ± 0.02	0.66 ± 0.03	0.54 ± 0.14	0.49 ± 0.03	$\mathbf{0.73 \pm 0.01}$

ECG-MoE reduces MAE in RR-interval estimation by 46.0% and improves arrhythmia detection accuracy by 10.6%, demonstrating superior diagnostic capability.

RQ3: Computational Efficiency. Can ECG-MoE maintain viability under resource constraints?

We measure GPU memory and throughput during inference. ECG-MoE achieves a breakthrough in efficiency by operating within a constrained 8.2 GB of GPU memory, which represents a significant 35% reduction in resource consumption. It processes data at a rate of 14.7 samples per second, yielding a processing speed that is three times faster than real-time. Most importantly, the model maintains an optimal balance between this exceptional efficiency and high predictive accuracy. This supports operation in resource-limited environments without compromising performance.

RQ4: Ablation Study. How do architectural components impact ECG-MoE's performance?

Our ablation studies highlight key insights into ECG-MoE's architectural design. As shown in left column of Table 3 further shows that the rhythm-conditioned MoE achieves clear clinical advantages over conventional variants. By aligning expert routing with cardiac phase, ECG-MoE outperforms TimeMoE in RR-interval and potassium estimation while improving sex classification, confirming that phase-aware gating benefits morphology-sensitive tasks.

Table 3: Comprehensive performance comparison of various MoE and attention mechanisms against the proposed ECG-MoE on ECG tasks.

		MoE	Time MoE	Self-Attn	Cross-Attn	ECG-MoE
Regression (MAE↓)	RR	103.47 ± 2.4	78.64 ± 1.4	117.46 ± 3.2	86.71 ± 5.8	$\mathbf{76.37 \pm 4.7}$
	Age	16.49 ± 0.87	13.04 ± 0.14	15.47 ± 0.85	14.71 ± 0.51	$\mathbf{12.83 \pm 0.42}$
Binary Class (F1↑)	Sex	0.53 ± 0.13	0.61 ± 0.10	0.54 ± 0.01	0.61 ± 0.02	$\mathbf{0.69 \pm 0.01}$
	Ka	0.39 ± 0.17	0.50 ± 0.10	0.48 ± 0.00	0.48 ± 0.00	$\mathbf{0.57 \pm 0.00}$
15 Class (ACC↑)	AD	0.62 ± 0.01	$\mathbf{0.74 \pm 0.00}$	0.67 ± 0.00	0.69 ± 0.01	0.73 ± 0.00

(a) Self-Attention (b) Cross Attention (c) Hybrid Multi-Head Attention
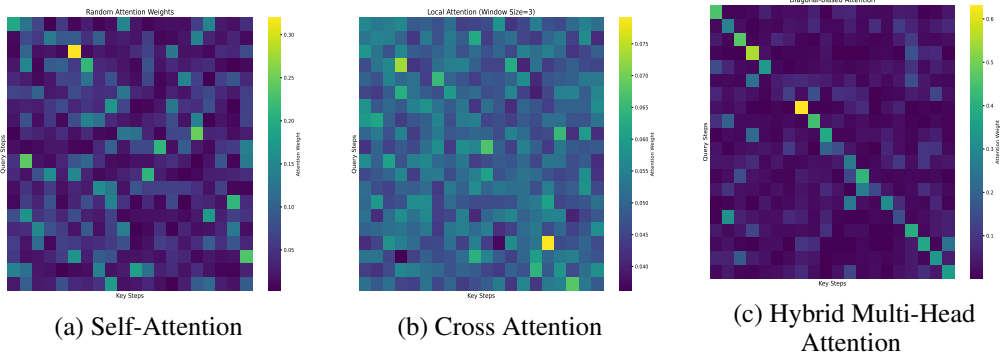
Figure 3: Comparative Attention Heatmaps for Different Attention Mechanisms: An Analysis Based on the RR Interval Task in the MIMIC-IV Dataset.

As shown in middle column of Table 3, hybrid attention is vital for clinical precision. It consistently surpasses self- and cross-attention in RR-interval and age estimation and achieves the highest F1 across classification tasks. Figure 3 illustrates that hybrid attention effectively integrates local morphology and global rhythm context, producing interpretable, physiologically grounded predictions.

These findings show that ECG-MoE's integration of multi-task periodicity learning, phase-aligned gating, and hybrid attention enables superior diagnostic accuracy and interpretability, confirming that ECG-specific inductive biases are essential for clinically reliable AI cardiology.

5 Conclusion

This study introduces ECG-MoE, a hybrid architecture that advances multi-task ECG analysis through specialized foundation integration. The model achieves SOTA performance across diverse tasks, reducing RR-interval MAE by 46.0% compared to TEMPO and improving arrhythmia detection by 10.6% over MOMENT, validating the effectiveness of its period-aware design and hybrid attention mechanisms. In addition, ECG-MoE establishes new efficiency benchmarks for clinical deployment, requiring only 8.2 GB of GPU memory and processing 14.7 samples per second, which enables real-time analysis on consumer-grade hardware and broadens accessibility in resource-limited settings. While the model demonstrates robust performance, challenges remain in potassium abnormality detection. Future efforts will focus on adaptive ensemble refinement and the incorporation of electrophysiological priors to enhance clinical decision support. The open-source release of ECG-MoE highlights the promise of domain-specific architectural innovations in advancing globally accessible, high-precision cardiac care.

6 Acknowledgement

References

[1] Ala Alwan. *Global status report on noncommunicable diseases 2010.* 2011.

[2] Josef Stehlik, Carsten Schmalfuss, Biykem Bozkurt, Jose Nativi-Nicolau, Peter Wohlfahrt, Stephan Wegerich, Kevin Rose, Ranjan Ray, Richard Schofield, Anita Deswal, et al. Continuous wearable monitoring analytics predict heart failure hospitalization: the link-hf multicenter study. *Circulation: Heart Failure*, 13(3):e006513, 2020.

[3] Juntong Ni, Saurabh Kataria, Shengpu Tang, Carl Yang, Xiao Hu, and Wei Jin. Ppg-distill: Efficient photoplethysmography signals analysis via foundation model distillation. In *NeurIPS 2025 Workshop on Learning from Time Series for Health*, 2025.

[4] Craig T January, L Samuel Wann, Joseph S Alpert, Hugh Calkins, Joaquin E Cigarroa, Joseph C Cleveland, Jamie B Conti, Patrick T Ellinor, Michael D Ezekowitz, Michael E Field, et al. 2014 aha/acc/hrs guideline for the management of patients with atrial fibrillation: a report of the american college of cardiology/american heart association task force on practice guidelines and the heart rhythm society. *Journal of the American College of Cardiology*, 64(21):e1–e76, 2014.

[5] John R Montford and Stuart Linas. How dangerous is hyperkalemia? *Journal of the American Society of Nephrology*, 28(11):3155–3165, 2017.

[6] Peter J Schwartz, Arthur J Moss, G Michael Vincent, and Richard S Crampton. Diagnostic criteria for the long qt syndrome. an update. *Circulation*, 88(2):782–784, 1993.

[7] Kristian Thygesen, Joseph S Alpert, Allan S Jaffe, Bernard R Chaitman, Jeroen J Bax, David A Morrow, Harvey D White, and Executive Group on behalf of the Joint European Society of Cardiology (ESC)/American College of Cardiology (ACC)/American Heart Association (AHA)/World Heart Federation (WHF) Task Force for the Universal Definition of Myocardial Infarction. Fourth universal definition of myocardial infarction (2018). *Journal of the American college of cardiology*, 72(18):2231–2264, 2018.

[8] Galen S Wagner, Peter Macfarlane, Hein Wellens, Mark Josephson, Anton Gorgels, David M Mirvis, Olle Pahlm, Borys Surawicz, Paul Kligfield, Rory Childers, et al. Aha/accf/hrs recommendations for the standardization and interpretation of the electrocardiogram: part vi: acute ischemia/infarction a scientific statement from the american heart association electrocardiography and arrhythmias committee, council on clinical cardiology; the american college of cardiology foundation; and the heart rhythm society endorsed by the international society for computerized electrocardiology. *Journal of the American College of Cardiology*, 53(11):1003–1011, 2009.

[9] David S Rosenbaum, Lance E Jackson, Joseph M Smith, Hasan Garan, Jeremy N Ruskin, and Richard J Cohen. Electrical alternans and vulnerability to ventricular arrhythmias. *New England journal of medicine*, 330(4):235–241, 1994.

[10] Zhongwei Wan, Che Liu, Xin Wang, Chaofan Tao, Hui Shen, Zhenwu Peng, Jie Fu, Rossella Arcucci, Huaxiu Yao, and Mi Zhang. Meit: Multi-modal electrocardiogram instruction tuning on large language models for report generation. *arXiv preprint arXiv:2403.04945*, 2024.

[11] Borys Surawicz and Timothy Knilans. *Chou's electrocardiography in clinical practice: adult and pediatric*. Elsevier Health Sciences, 2008.

[12] Antoni Bayes De Luna, Miquel Fiol-Sala, Antoni Bayes-Genis, and Adrian Baranchuk. *Clinical electrocardiography: a textbook*. John Wiley & Sons, 2021.

[13] Gan-Xin Yan and Charles Antzelevitch. Cellular basis for the normal t wave and the electrocardiographic manifestations of the long-qt syndrome. *Circulation*, 98(18):1928–1936, 1998.

[14] Gari D Clifford, Chengyu Liu, Benjamin Moody, Li-wei H Lehman, Ikaro Silva, Qiao Li, Alistair E Johnson, and Roger G Mark. Af classification from a short single lead ecg recording: The physionet/computing in cardiology challenge 2017. In *2017 computing in cardiology (CinC)*, pages 1–4. IEEE, 2017.

[15] Leif Sörnmo and Pablo Laguna. *Bioelectrical signal processing in cardiac and neurological applications*. Academic press, 2005.

[16] Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 11106–11115, 2021.

[17] Noam Shazeer, Azalia Mirhoseini, Krzysztof Maziarz, Andy Davis, Quoc Le, Geoffrey Hinton, and Jeff Dean. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. *arXiv preprint arXiv:1701.06538*, 2017.

[18] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*, 2021.

[19] Tim Dettmers, Mike Lewis, Younes Belkada, and Luke Zettlemoyer. Gpt3. int8 (): 8-bit matrix multiplication for transformers at scale. *Advances in neural information processing systems*, 35:30318–30332, 2022.

[20] Pranav Rajpurkar, Awni Y Hannun, Masoumeh Haghpanahi, Codie Bourn, and Andrew Y Ng. Cardiologist-level arrhythmia detection with convolutional neural networks. *arXiv preprint arXiv:1707.01836*, 2017.

[21] Zachi I Attia, Peter A Noseworthy, Francisco Lopez-Jimenez, Samuel J Asirvatham, Abhishek J Deshmukh, Bernard J Gersh, Rickey E Carter, Xiaoxi Yao, Alejandro A Rabinstein, Brad J Erickson, et al. An artificial intelligence-enabled ecg algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. *The Lancet*, 394(10201):861–867, 2019.

[22] Xiaoda Wang, Yuji Zhao, Kaiqiao Han, Xiao Luo, Sanne van Rooij, Jennifer Stevens, Lifang He, Liang Zhan, Yizhou Sun, Wei Wang, et al. Conditional neural ode for longitudinal parkinson's disease progression forecasting. In *Proceedings of The 19th IEEE International Conference on Bioinformatics and Biomedicine*, (BIBM'25), 2025.

[23] Zewen Liu, Xiaoda Wang, Bohan Wang, Zijie Huang, Carl Yang, and Wei Jin. Graph odes and beyond: A comprehensive survey on integrating differential equations with graph neural networks. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2*, pages 6118–6128, 2025.

[24] Xiaoda Wang, Yuji Zhao, Kaiqiao Han, Xiao Luo, Sanne van Rooij, Jennifer Stevens, Lifang He, Liang Zhan, Yizhou Sun, Wei Wang, et al. Conditional neural ode for longitudinal parkinson's disease progression forecasting. In *Abstract in the Organization for Human Brain Mapping Annual Meeting*, 2025.

[25] Jonathan Rubin, Saman Parvaneh, Asif Rahman, Bryan Conroy, and Saeed Babaeizadeh. Densely connected convolutional networks and signal quality analysis to detect atrial fibrillation using short single-lead ecg recordings. In *2017 Computing in cardiology (cinc)*, pages 1–4. IEEE, 2017.

[26] Jinlong Ji, Xuhui Chen, Changqing Luo, and Pan Li. A deep multi-task learning approach for ecg data analysis. In *2018 IEEE EMBS International conference on biomedical & health informatics (BHI)*, pages 124–127. IEEE, 2018.

[27] Brian Gow and et al. Mimic-iv-ecg: Diagnostic electrocardiogram matched subset, 2023.

[28] Haixu Wu, Tengge Hu, Yong Liu, Hang Zhou, Jianmin Wang, and Mingsheng Long. Timesnet: Temporal 2d-variation modeling for general time series analysis. *arXiv preprint arXiv:2210.02186*, 2022.

[29] Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. Are transformers effective for time series forecasting? In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pages 11121–11128, 2023.

[30] Mononito Goswami, Konrad Szafer, Arjun Choudhry, Yifu Cai, Shuo Li, and Artur Dubrawski. Moment: A family of open time-series foundation models. *arXiv preprint arXiv:2402.03885*, 2024.

[31] Defu Cao, Furong Jia, Sercan O Arik, Tomas Pfister, Yixiang Zheng, Wen Ye, and Yan Liu. Tempo: Prompt-based generative pre-trained transformer for time series forecasting. *arXiv preprint arXiv:2310.04948*, 2023.

[32] Kaden McKeen, Laura Oliva, Sameer Masood, Augustin Toma, Barry Rubin, and Bo Wang. Ecg-fm: An open electrocardiogram foundation model. *arXiv preprint arXiv:2408.05178*, 2024.