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

Anonymous Author(s)

Affiliation Address email

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.

9 1 Instruction

2

6

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

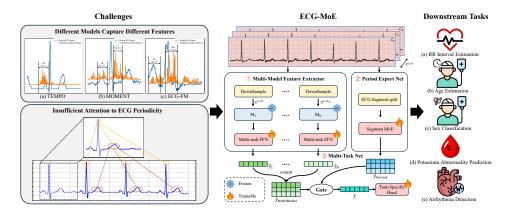


Figure 1: Existing foundation models have limitations because different models capture distinct ECG features, hindering multi-task performance within a single model. Furthermore, despite the strong periodicity inherent in ECG signals, this characteristic is often overlooked by existing ECG foundation 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.

Foundation models have shown promise in automated ECG diagnosis but struggle to comprehensively represent clinically nuanced, segment-specific features [9]. 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 18 signals, this characteristic is often overlooked by existing ECG foundation models. They also incur 19 high computational costs, hindering clinical adoption. Moreover, they often fail to align with ECG's 20 inherent periodicity, where each heartbeat carries distinct diagnostic information [10, 11, 12], making 21 it difficult to leverage rhythmic consistency while accommodating pathological variations [13, 14]. 22 To address these issues, as shown in Figure 1, we introduce ECG-MoE, a hybrid architecture that integrates multi-model temporal features [15] with a dual-path Mixture-of-Experts (MoE) using task-24 conditioned gating [16]. Specifically, we propose an ensemble learning method based on Period MoE. 25 We employ Low-Rank Adaptation (LoRA) for parameter-efficient fusion [17, 18]. 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.

Related Works 2 29

26

27

28

Electrocardiogram Foundation Models. Recent deep learning advances have developed ECG foun-30 dation models using CNNs [19] and Transformers [15] for tasks like arrhythmia detection. While 31 pretraining and self-supervision [20] learn general features, these approaches often treat ECGs as 32 generic time-series, neglecting the inherent periodicity critical to cardiac electrophysiology [10, 13]. 33 This results in suboptimal handling of phase-localized abnormalities and inefficient rhythm modeling, 34 which is a gap our period-conditioned framework directly addresses. 35

ECG Periodicity Modeling. ECG's quasi-periodicity has been exploited via segmentation [14] and 36 rhythm-tracking methods, yet these struggle with pathological variations and holistic modeling. 37 Although RNNs [21] capture temporal dependencies, they are inefficient and sensitive to period 38 variability. No current framework incorporates periodicity as a core inductive bias for foundation 39 models, despite its diagnostic importance [7, 12]. 40

Multi-Task ECG Analysis. Multi-task learning (MTL) addresses simultaneous cardiac condition 41 detection, but shared-backbone approaches [22] suffer from interference, especially in complex 42 diagnostics like ST-segment analysis [6]. Modular or parameter-efficient designs improve scalability 43 but ignore ECG's phase-specific biomarkers. Existing models fail to combine task-specific feature 44 extraction with periodicity constraints. ECG-MoE overcomes this via hierarchical fusion and task-45 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 48 that synergistically integrates heterogeneous feature representations while explicitly modeling cardiac 49 periodicity. The model comprises three modules, each with distinct mathematical formulations:

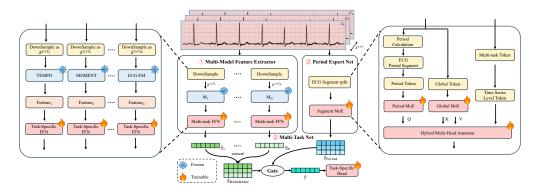


Figure 2: ECG-MoE framework: ① Multi-model feature extraction with adaptive downsampling, ② Period-aware gating weight generation via specialized MoE, 3 Multi-Task fusion using LoRA.

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

53 adaptive downsampling:

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

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,
- 56 enhancing feature diversity.
- 57 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
- 59 CNN experts with different kernel sizes process morphological features within beats, while two
- 60 dilated CNNs capture inter-beat rhythm patterns. A task-conditioned gating mechanism dynamically
- weights expert contributions based on global signal characteristics and task embeddings:

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

- 62 Features are integrated via multi-head attention to produce the final periodic representation Fperiodic.
- 63 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:

$$\mathbf{F} \mathbf{f} \mathbf{u} \mathbf{s} \mathbf{e} \mathbf{d} = \hat{\alpha} m \mathbf{F} m \oplus \hat{\alpha} p \mathbf{F} \mathbf{p} \mathbf{e} \mathbf{r} \mathbf{i} \mathbf{o} \mathbf{d} \mathbf{c}$$

$$\tag{4}$$

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:

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

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

69 4 Experiments

- 70 In this section, we evaluate the effectiveness of our proposed model from two key perspectives:
- 71 performance and efficiency. Specifically, we compare its predictive accuracy across multiple tasks
- 72 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.

75 4.1 Dataset and Baselines

- 76 We utilize the MIMIC-IV-ECG [23] and PTB-XL [24] datasets. MIMIC-IV-ECG is currently the
- 77 largest publicly released ECG repository, which contains 800,035 diagnostic electrocardiograms
- ₇₈ acquired from 161,352 unique patients. The PTB-XL dataset comprises 21,837 clinical 12-lead
- 79 ECGs, each lasting 10 seconds, from 18,885 patients. We select TimesNet [25], DLinear [26],
- MOMENT [27], TEMPO [28] and ECG-FM [29] as baselines and construct an ensemble of them.

4.2 Research Questions and Results

- 82 <u>RQ1: Model Selection Rationale.</u> Why were specific foundation models selected, and what benefits
- 83 does their ensemble provide?

81

- 84 We evaluate SOTA time-series models based on architectural diversity and ECG applicability through
- 25 zero-shot evaluation on 10,000 patient samples.
- 86 Table 1 shows zero-shot performance of baseline models. We can get the key findings of ECG-FM
- 87 excels in clinical diagnostics with domain-specific pretraining. TimesNet captures morphologi-
- 88 cal variations effectively. DLinear shows competence in temporal feature modeling. MOMENT
- 89 underperforms without domain adaptation.
- These results validate our ensemble strategy, combining strengths of different architectures.

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 Age	817.0 ± 2.5 62.28 ± 0.36	816.4 ± 2.9 62.63 ± 0.50	816.6 ± 2.1 62.61 ± 0.37	816.3 ± 1.9 62.33 ± 0.38	816.3 ± 1.9 62.27 ± 0.38
Binary Class (F1†)	Sex Ka	0.60 ± 0.00 0.06 ± 0.00	0.51 ± 0.08 0.05 ± 0.00	0.34 ± 0.00 0.02 ± 0.00	0.42 ± 0.00 0.18 ± 0.00	0.33 ± 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

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

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?

97

98

100

101

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.

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

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 Age	$103.47 \pm 2.4 16.49 \pm 0.87$	78.64 ± 1.4 13.04 ± 0.14	$\begin{array}{ c c c c }\hline 117.46 \pm 3.2 \\ 15.47 \pm 0.85 \\ \hline \end{array}$	86.71 ± 5.8 14.71 ± 0.51	$\begin{array}{ c c c }\hline {\bf 76.37 \pm 4.7} \\ {\bf 12.83 \pm 0.42} \\ \end{array}$
Binary Class (F1\(\epsilon\))	Sex Ka	0.53 ± 0.13 0.39 ± 0.17	0.61 ± 0.10 0.50 ± 0.10	$\begin{array}{ c c c c c }\hline 0.54 \pm 0.01 \\ 0.48 \pm 0.00 \end{array}$	0.61 ± 0.02 0.48 ± 0.00	$\begin{array}{ c c c c c c c c c c c c c c c c c c c$
15 Class (ACC↑)	AD	0.62 ± 0.01	0.74 ± 0.00	0.67 ± 0.00	0.69 ± 0.01	0.73 ± 0.00

As shown in Table 3, we evaluate MoE architectures, and attention mechanisms. Multi-task learning significantly enhances performance across various clinical tasks. ECG-MoE's cardiac rhythm-aware gating mechanism notably outperforms generic mixture-of-experts approaches, while its hybrid attention module enables precision diagnostics by achieving a 53.5% improvement in RR-interval

accuracy. Furthermore, routing heatmaps reveal physiologically plausible patterns of expert activation,
 reinforcing the model's clinical interpretability.

110 References

- 1 [1] Ala Alwan. Global status report on noncommunicable diseases 2010. 2011.
- 112 [2] Josef Stehlik, Carsten Schmalfuss, Biykem Bozkurt, Jose Nativi-Nicolau, Peter Wohlfahrt,
 113 Stephan Wegerich, Kevin Rose, Ranjan Ray, Richard Schofield, Anita Deswal, et al. Continuous
 114 wearable monitoring analytics predict heart failure hospitalization: the link-hf multicenter study.
 115 Circulation: Heart Failure, 13(3):e006513, 2020.
- 116 [3] Craig T January, L Samuel Wann, Joseph S Alpert, Hugh Calkins, Joaquin E Cigarroa, Joseph C
 117 Cleveland, Jamie B Conti, Patrick T Ellinor, Michael D Ezekowitz, Michael E Field, et al. 2014
 118 aha/acc/hrs guideline for the management of patients with atrial fibrillation: a report of the
 119 american college of cardiology/american heart association task force on practice guidelines and
 120 the heart rhythm society. *Journal of the American College of Cardiology*, 64(21):e1–e76, 2014.
- [4] John R Montford and Stuart Linas. How dangerous is hyperkalemia? *Journal of the American Society of Nephrology*, 28(11):3155–3165, 2017.
- [5] 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.
- [6] 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.
- [7] Galen S Wagner, Peter Macfarlane, Hein Wellens, Mark Josephson, Anton Gorgels, David M 131 Mirvis, Olle Pahlm, Borys Surawicz, Paul Kligfield, Rory Childers, et al. Aha/accf/hrs recom-132 mendations for the standardization and interpretation of the electrocardiogram: part vi: acute 133 ischemia/infarction a scientific statement from the american heart association electrocardio-134 135 graphy and arrhythmias committee, council on clinical cardiology; the american college of cardiology foundation; and the heart rhythm society endorsed by the international society for 136 computerized electrocardiology. Journal of the American College of Cardiology, 53(11):1003– 137 1011, 2009. 138
- [8] 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.
- [9] 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.
- [10] Borys Surawicz and Timothy Knilans. Chou's electrocardiography in clinical practice: adult
 and pediatric. Elsevier Health Sciences, 2008.
- 147 [11] Antoni Bayes De Luna, Miquel Fiol-Sala, Antoni Bayes-Genis, and Adrian Baranchuk. *Clinical electrocardiography: a textbook.* John Wiley & Sons, 2021.
- 149 [12] 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.
- [13] 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.

- [14] Leif Sörnmo and Pablo Laguna. Bioelectrical signal processing in cardiac and neurological
 applications. Academic press, 2005.
- [15] 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.
- [16] 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.
- 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.
- [18] 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.
- 171 [19] Pranav Rajpurkar, Awni Y Hannun, Masoumeh Haghpanahi, Codie Bourn, and Andrew Y Ng.
 172 Cardiologist-level arrhythmia detection with convolutional neural networks. *arXiv preprint*173 *arXiv:1707.01836*, 2017.
- [20] 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.
- 179 [21] Jonathan Rubin, Saman Parvaneh, Asif Rahman, Bryan Conroy, and Saeed Babaeizadeh.
 180 Densely connected convolutional networks and signal quality analysis to detect atrial fibrillation
 181 using short single-lead ecg recordings. In 2017 Computing in cardiology (cinc), pages 1–4.
 182 IEEE, 2017.
- [22] 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.
- 186 [23] Brian Gow and et al. Mimic-iv-ecg: Diagnostic electrocardiogram matched subset, 2023.
- [24] Patrick Wagner, Nils Strodthoff, Ralf-Dieter Bousseljot, Dieter Kreiseler, Fatima I. Lunze,
 Wojciech Samek, and Tobias Schaeffter. Ptb-xl, a large publicly available electrocardiography
 dataset. Scientific Data, May 2020.
- 190 [25] Haixu Wu, Tengge Hu, Yong Liu, Hang Zhou, Jianmin Wang, and Mingsheng Long. Times191 net: Temporal 2d-variation modeling for general time series analysis. *arXiv preprint*192 *arXiv:2210.02186*, 2022.
- [26] 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.
- [27] 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.
- 199 [28] Defu Cao, Furong Jia, Sercan O Arik, Tomas Pfister, Yixiang Zheng, Wen Ye, and Yan Liu.
 200 Tempo: Prompt-based generative pre-trained transformer for time series forecasting. *arXiv*201 *preprint arXiv:2310.04948*, 2023.
- [29] 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.