ShapeEFM: A Shape-based Pretrained Foundation Model on ECG Morphology Comprehension

Huan Liu[1], Jiaqi Fan[2], Siyan Guo[1], Junyi Wang[1], Xiaoyu Bai[1], and Le Lu[1]

[1] Ant Group, Hangzhou, China
[2] The Second Affiliated Hospital of Zhejiang University School of Medicine, Hangzhou, China
{hliu8335}@gmail.com

Abstract. Electrocardiogram (ECG) mass screening plays a crucial role in the early detection of asymptomatic cardiovascular abnormalities, enabling timely intervention and prevention of sudden cardiac events in seemingly healthy individuals. Nowadays, ECG self-supervised learning (eSSL) is the cutting-edge frontier technique for representation learning on unannotated ECG data. Despite extensive efforts in modeling heartbeat rhythms from ECG signals, the continuous temporal morphology-based dynamics of ECG patterns remain incompletely represented in current modeling framework. In this work, we introduce a novel shape-based eSSL approach to capture the intricate wave deflection characteristics of ECG signals, in which temporal trend shifts are discretized and treated as word tokens in a symbolic sequence. We propose ShapeEFM: a novel transformer-based foundation model pretrained using generative eSSL approach. We evaluated ShapeEFM across EchoNext and CSN datasets, where it demonstrated robust competitiveness and enhanced interpretability against the existing eSSL methods. The code is available at: `https://github.com/AQ-MedAI/ShapeEFM`

Keywords: Foundation Model · Shape · ECG Morphology · eSSL.

1 Introduction

The implementation of electrocardiogram (ECG) mass screening is essential for shifting the paradigm of cardiovascular care from reactive treatment to proactive prevention. By identifying asymptomatic abnormalities in apparently healthy populations, clinicians can initiate timely interventions that significantly reduce the incidence of sudden cardiac death [20]. Recent advances in artificial intelligence have further expanded the feasibility of such initiatives. AI-driven models are now capable of interpreting ECG data with expert-level precision, identifying subtle signatures of arrhythmias [19], and structural heart disease [17]. This technological evolution broadens the scope of detectable conditions and enhances the scalability of screening programs, offering a cost-effective pathway to improved population health outcomes.

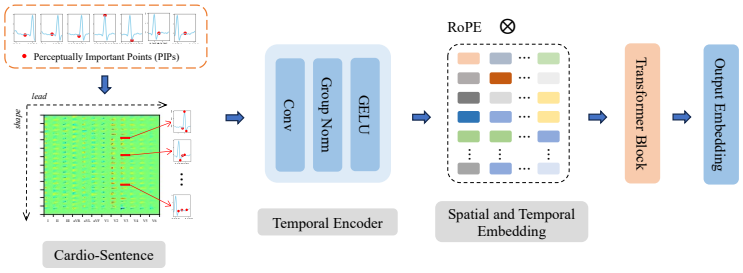

Fig. 1. The Cardio-Morpheme Transformer structures input ECG signals into Cardio-Sentences using PIPs, encoding each Cardio-Morpheme via a temporal encoder enriched with spatial and temporal embeddings. By employing RoPE to model cardiac cycle information, the resulting embeddings are processed by a patch-wise attention Transformer to generate the final diagnostic output.

The development of ECG Self-Supervised Learning (eSSL) has undergone a paradigm shift, evolving from task-specific architectures into large-scale foundation models that harness massive unlabeled datasets to learn robust, generalizable representations [1,6,15,18,24]. Current research indicates that hybrid SSL frameworks, which integrate generative reconstruction with contrastive discrimination, offer the most effective approach by simultaneously capturing local structural patterns and high-level functional semantics [13]. Recent methodologies like ST-MEM [14] have further propelled the field by explicitly modeling spatio-temporal relationships via lead-specific embeddings, while novel approaches such as HeartLang [10] treat heartbeats as "words" in an ECG vocabulary to decipher latent semantic relationships. Ultimately, these advancements underscore the transformative power of eSSL: by establishing a versatile, universal feature extractor, these models bridge the gap between data acquisition and clinical interpretation, ensuring reliable diagnostic performance across heterogeneous environments and diverse patient populations.

Although the prevailing eSSL paradigm excels at capturing high-level semantic relationships, it remains fundamentally constrained in its representation of real-world cardiac events [10,11]. Unlike HeartLang, which compresses each beat into one word and largely relies on the accurate detection of the QRS complex, ShapeEFM treats sub-beat morphology changes such as QRS widening and ST deviation as reconstruction units, making these clinically relevant transitions explicit targets of SSL rather than implicit within heartbeat tokens. And these morphology-level shapelets enhance model's representation capability for ECG morphology.

To fully capture the granular semantic representation of ECG morphology, we propose ShapeEFM. We conceptualize individual waveform components, such as the P, Q, R, S, and T waves, as fundamental linguistic units, termed "Cardio-Morpheme". These morphemes serve as the atomic "words" of the physiological signal and are sequentially concatenated to construct a "Cardio-Sentence", a structured representation that captures the syntactic relationships between

adjacent ECG morphology. The main contributions of this paper are threefold: **First**, we propose a novel Cardio-Morpheme Transformer that decomposes ECG signals into semantic units, preserving the morphological and clinical integrity of ECG waves, as shown in Fig. 1. **Second**, by aggregating these units into Cardio-Sentences, we introduce a syntactic structure that captures the temporal dependencies between ECG waves. **Third**, the proposed method enhances model interpretability, allowing cardiac abnormalities to be visualized and understood as semantic anomalies within a physiological narrative.

2 Methods

2.1 Model Architecture

Shape Detection. This method learns shapes in an encoder-freeze-like manner, enabling the segmentation of ECGs into a series of signal corpora that describe the signal's trajectory. The process is achieved by identifying PIPs through the following steps [12] , assuming a time series $\mathcal{T} \in \mathbb{R}^{1\times T}$. Initially, the first and last time points are added to the PIP set $\mathbb{S}$. Then, for remaining time point i, the point with the highest reconstruction distance RD is iteratively added to $\mathbb{S}$ until the size of $\mathbb{S}$ reaches N_{pip}, which is set to 250 for 10s ECG.

$$RD(i,\mathbb{S}) = \frac{\frac{\mathcal{T}_e-\mathcal{T}_s}{P_e-P_s} * (P_i - P_e) - (\mathcal{T}_i - \mathcal{T}_e)}{\sqrt{(\frac{\mathcal{T}_e-\mathcal{T}_s}{P_e-P_s})^2 + 1}}, \quad i \notin \mathbb{S} \tag{1}$$

$\forall j \in \mathbb{S}$, P_j is the z-normalized value of the list of positions $[1,\ldots,j]$ and s, e denote the two points in $\mathbb{S}$ nearest to i.

By treating each major trend inflection, which is defined as the ECG signal segment between three adjacent PIPs, as a Cardio-Morpheme, the raw ECG signal is reshaped into a Cardio-Sentence $\mathcal{X} \subset \mathbb{R}^{N_{pip}\times L}$. Here, $L = k \times n$, where k denotes the maximum length of a Cardio-Morpheme (set to 40, if the length of a Cardio-Morpheme is smaller than k, it is padded with zeros to match the required size) and n represents the number of leads. The PIP discovery is performed exclusively on Lead I, and the identified points are subsequently applied to all 12 leads.

Temporal Encoder. ECG signals exhibit high temporal resolution, and each token within a Cardio-Sentence captures temporal trend shifts across all leads. To extract and encode distinguishing features, it is necessary to embed the data into a higher-dimensional token feature space. Therefore, we employ a temporal encoder comprising several temporal convolution blocks to encode each Cardio-Morpheme into a patch embedding. Each temporal convolution block consists of a 1-D convolution layer, a group normalization layer [23], and a GELU activation function [9]. Finally, the Cardio-Sentence is embedded into $\mathcal{E} \subset \mathbb{R}^{L_{sen}\times L_e}$, where $L_{sen} = N_{pip} \times 12$ represents the sequence length and $L_e = 256$ denotes the dimension of the token feature space.

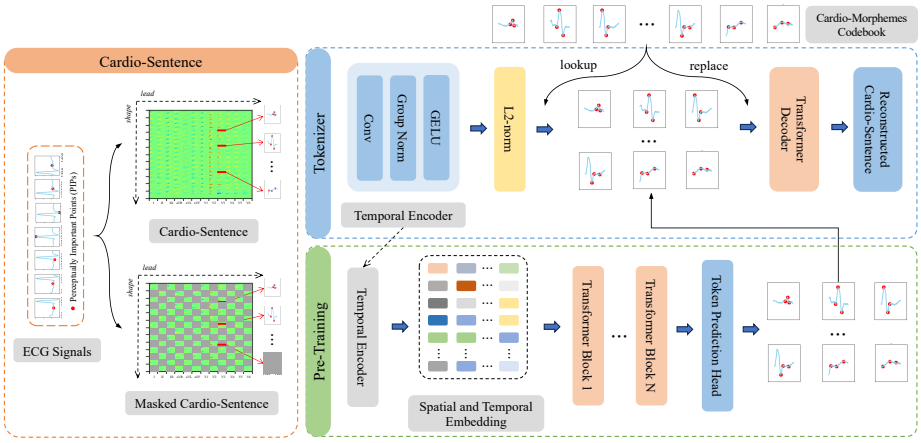

Fig. 2. Overview of the Cardio-Morpheme tokenizer training and ShapeEFM pre-training. **Left:** The construction process of the Cardio-Sentence. **Right Up:** Training of the Cardio-Morpheme tokenizer to discretize ECG signals into discrete shape tokens via shape reconstruction. **Right Down:** During pre-training, a subset of Cardio-Morphemes is masked, with the objective to predict the masked tokens from the visible Cardio-Morphemes.

Spatial & Temporal Embedding. To enable the model to perceive the temporal and spatial information of the patch embeddings, we initialize a temporal embedding list $TE = \{te_1, te_2, \ldots, te_{N_{pip}}\}$ using Rotary Positional Embeddings (RoPE) [21] and a spatial embedding list $SE = \{se_1, se_2, \ldots, se_{12}\}$ using absolute position encoding. Thus, given an arbitrary output embedding $\mathrm{E}_t \in \mathcal{E}$ from the temporal encoder, we add the corresponding temporal and spatial embeddings to obtain the final embedding $\mathcal{F} \subset \mathbb{R}^{L_{sen} \times L_e}$.

$$\mathrm{E}' = \mathrm{E}_t + TE_u + SE_v, \quad u \in \{1, \ldots, N_{pip}\},\ v \in \{1, \ldots, 12\} \tag{2}$$

Transformer Encoder. Finally, the sequence of embeddings is fed directly into the Transformer encoder [22]. To ensure stable and efficient training, we apply QK-Normalization prior to the dot-product attention mechanism, thereby preventing excessively large values in the attention logits [5]:

$$\mathrm{Attention}(Q, K, V) = \mathrm{softmax}\left(\frac{\mathrm{LN}(Q)\mathrm{LN}(K)^T}{\sqrt{d_{\mathrm{head}}}}\right) V, \tag{3}$$

where d_{head} denotes the dimension of a single head in the multi-head attention mechanism, and LN refers to Layer Normalization [2].

2.2 Cardio-Morpheme Tokenizer Training

Individual Cardio-Morphemes lack generalization due to their independence. To enable ShapeEFM to learn cross-subject representations during pre-training,

we introduce a Cardio-Morpheme vocabulary, a codebook of generalized Morphemes. This vocabulary is jointly optimized via quantization and reconstruction, as depicted in the upper right corner of Fig. 2.

Vector Quantization. A quantizer is employed to map inputs to a collective Cardio-Morpheme vocabulary $\mathcal{V} = \{v_i \mid i = 1, \ldots, K\} \subset \mathbb{R}^{K \times L_e}$, where K denotes the number of entries in the vocabulary. The quantizer identifies the nearest neighbor for each interval representation f_i in $\mathcal{F}$ using cosine similarity. This procedure is formulated as: $z_i = \underset{f_i \in \mathcal{F}, v_i \in \mathcal{V}}{\arg\min} \|\ell_2(f_i) - \ell_2(v_i)\|_2$, where ℓ_2 represents l_2-normalization.

Shape Reconstruction. After quantization, the normalized discrete embeddings $\{\ell_2(z_i) \mid i = 1, \ldots, l\}$ are fed into the transformer decoder. This process can be represented as $\hat{x} = \bigcup_{i=1}^{l} f_d(\ell_2(v_{z_i}))$, where $\hat{x}$ is the reconstructed Cardio-Sentence and f_d is the decoder. To ensure stable updates for the Cardio-Morpheme vocabulary, we adopt an exponential moving average (EMA) strategy. The mean squared error (MSE) loss is utilized to jointly optimize the quantization and reconstruction processes.

2.3 ShapeEFM Pre-Training

Spatio-Temporal Masking. To enable ShapeEFM to learn Cardio-Morpheme features, we perform random masking on individual Cardio-Morphemes, compelling the model to infer the complete Cardio-Sentence from the visible context. Given a Cardio-Sentence $\mathcal{X} \subset \mathbb{R}^{L_{sen} \times L_e}$ and Cardio-Morpheme $\mathcal{P} \subset \mathbb{R}^{1 \times L_e}$. We randomly generate a mask matrix $\mathcal{M} \subset \mathbb{R}^{L_{sen} \times L_e}$, where each element corresponds to a specific Cardio-Morpheme within a lead.

Collective Cardio-Morpheme Prediction. This stage aims to predict the indices of masked Cardio-Morphemes based on unmasked ones by minimizing the discrepancy between predicted and true indices. We extract the hidden vectors h corresponding to the masked segments and use a linear classifier to predict the collective Cardio-Morpheme indices [16]: $p(v_i \mid P_M) = \text{softmax}(\text{Linear}(h))$, where $P_M \in \mathcal{P}$ is the masked Cardio-Morpheme. The target loss function for this stage is:

$$\mathcal{L} = -\sum_{x \in \mathcal{X}} \log p(v_i \mid P_M) \tag{4}$$

3 Experiments

3.1 Configuration

Pre-training Implementation. To prepare the data for pre-training, we implemented a rigorous preprocessing pipeline for the MIMIC-IV-ECG dataset,

as detailed in Table 1. This process involved: (1) imputation of missing values (*i.e.*, 'NaN' and 'Inf') using the average of the six neighboring data points; and (2) upsampling of all ECG data to 500 Hz using polyphase filtering; and (3) application of a band-pass filter (0.05–40 Hz) to suppress baseline drift and high-frequency noise. For the pre-training phase, ShapeEFM consists of 12 transformer encoder blocks, 8 attention heads, and a dropout rate of 0.1, and we employed the AdamW optimizer with a learning rate of 5×10^{-4} and a weight decay of 1×10^{-4}. The model was trained for 50 epochs using a random masking rate of 50%, while a cosine annealing scheduler was utilized for learning rate adjustment. We used a batch size of 64 per GPU, conducting all experiments on eight NVIDIA A100 GPUs.

Downstream Tasks Implementation. We evaluated transfer performance using linear probing, where the pre-trained Cardio-Morpheme Transformer backbone remained frozen, and only the parameters of a newly initialized linear classifier were updated. To assess model robustness under low-resource conditions, we conducted linear probing for each downstream task using 1%, 10%, and 100% of the training data (see Table 1). For all tasks, we employed the Area Under the Receiver Operating Characteristic Curve (AUROC) and the Area Under the Precision-Recall Curve (AUPRC) as primary evaluation metrics.

Table 1. Overview of the public datasets used for pre-training and downstream tasks.

Dataset	Label	Task	Sampling Frequency	Train	Valid	Test
MIMIC-IV-ECG [7]	-	Pretrain	500Hz	773,120	15,902	-
EchoNext [17]	5	Downstream	250Hz	72,475	4,626	5,442
CSN [25,26]	38	Downstream	500Hz	16,546	1,860	4,620

3.2 Evaluation on Linear Probing

The linear probing evaluation results, summarized in Table 2, demonstrate the superior representation learning capability of ShapeEFM across varying label fractions. On the CSN dataset, ShapeEFM consistently outperforms competing methods in data-efficient scenarios, securing the best AUROC and AUPRC at the 10% label fraction and full 100% supervision level. In the more challenging EchoNext dataset, ShapeEFM exhibits robust generalizability; it attains the highest AUROC of 80.15% and AUPRC of 5.04% at 100% labels, while also maintaining a competitive edge in low-data regimes by securing the second-best AUROC at 10% labels and the best AUPRC at 10% labels. These findings validate that ShapeEFM effectively learns high-quality transferable features that scale effectively with data availability.

Table 2. AUROC and AUPRC results for ShapeEFM and competing methods. The best and second-best results are in **bold** and shaded in gray, respectively.

Metric	Method	CSN			EchoNext		
		1%	10%	100%	1%	10%	100%
AUROC	SimCLR [3]	65.20	71.62	84.02	64.31	69.95	74.71
	BYOL [8]	63.12	72.19	84.96	61.17	70.03	75.54
	MoCo-v3 [4]	63.54	80.74	82.77	62.12	72.76	79.02
	ST-MEM [14]	**68.89**	76.32	77.36	**65.12**	70.28	73.01
	HeartLang [10]	66.04	78.02	89.76	63.32	**75.29**	76.96
	ShapeEFM	67.09	**81.51**	**89.80**	64.05	74.29	**80.15**
AUPRC	SimCLR [3]	5.82	9.45	15.23	4.12	3.58	3.42
	BYOL [8]	5.45	9.80	15.60	3.95	3.65	3.55
	MoCo-v3 [4]	5.60	11.25	14.88	4.02	3.92	3.78
	ST-MEM [14]	**6.45**	10.15	13.20	4.35	3.75	3.30
	HeartLang [10]	6.26	10.50	16.11	**4.49**	4.08	3.90
	ShapeEFM	6.30	**12.05**	**16.85**	4.20	**4.25**	**5.04**

3.3 Cardio-Morpheme Vocabulary Visualization

The Cardio-Morpheme vocabulary visualization reveals a clear tripartite structure in the latent space, with well-separated clusters corresponding to QRS, RST, and ST segments, each representing a distinct phase of the cardiac cycle, as shown in Fig. 3. The QRS cluster (in teal) exhibits high internal variability, as reflected by the dispersed yet cohesive waveforms, consistent with the rapid, variable depolarization patterns across individuals and heart rates. In contrast, the ST segment cluster (in yellow) shows tighter morphological consistency, with nearly flat, horizontally aligned waveforms, aligning with its physiological role as an isoelectric interval between ventricular depolarization and repolarization. The RST cluster (in purple) occupies an intermediate position, both spatially and morphologically, capturing the transition from QRS termination to ST onset; its waveforms show gradual slope variations, reflecting the dynamic nature of early repolarization. The clean separation in the t-SNE space suggests that our feature representation effectively encodes clinically relevant morphological distinctions, enabling unsupervised discovery of functionally coherent Cardio-Morpheme for ECG analysis. Importantly, this expanded view preserves the clear pentapartite structure identified in our initial analysis, further validating the robustness of the feature representation.

3.4 Ablation Study

We conducted an ablation study to validate ShapeEFM's components (Table 3). Removing the Cardio-Morpheme Vocabulary (replaced by MSE reconstruction) caused significant performance drops, particularly in low-data regimes, high-

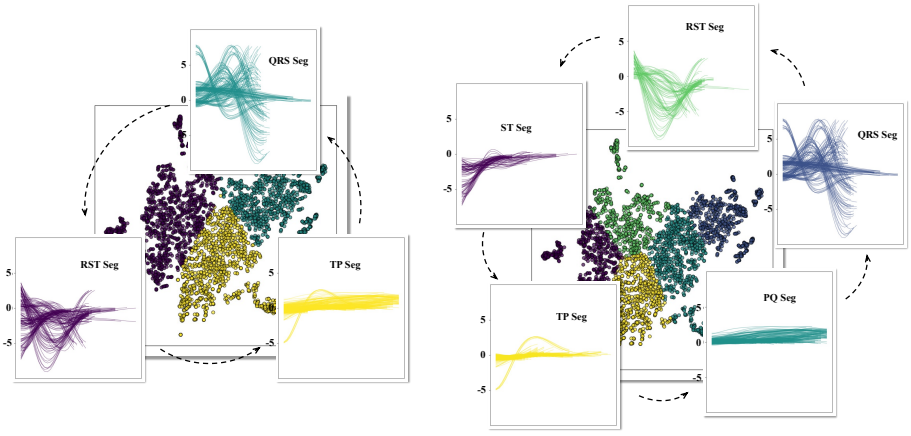

Fig. 3. Cardio-morpheme vocabulary visualization via t-SNE embedding of Cardio-Morpheme, revealing three distinct clusters corresponding to physiologically meaningful segments: QRS (teal), RST (purple), and TP (yellow); and five distinct clusters corresponding to: QRS (blue), RST (green), ST (purple), TP (yellow), and PQ (teal).

lighting the importance of our semantic tokenization. Similarly, eliminating Pre-training resulted in the lowest AUROC scores across all settings, confirming that pre-training is essential for learning robust general representations. Finally, ablating the Spatio-temporal Embedding consistently degraded performance, demonstrating that our specialized architecture effectively captures critical ECG dynamics. In all cases, the full ShapeEFM model achieved superior results, validating the synergistic contribution of each component.

Table 3. Linear probing AUROC results of the ablation study. The best and second-best results are in **bold** and shaded in gray, respectively.

Method	CSN			EchoNext		
	1%	10%	100%	1%	10%	100%
Cardio-Morpheme Vocabulary	64.12	79.80	88.10	61.88	72.50	77.92
Pre-training	63.80	77.90	85.90	60.55	70.15	76.80
Spatio-temporal Embedding	65.20	78.35	86.45	62.90	71.80	78.50
ShapeEFM	**67.09**	**81.51**	**89.80**	**64.05**	**74.29**	**80.15**

4 Conclusion

We presented ShapeEFM, a novel shape-based foundation model that overcomes the limitations of discrete segmentation by conceptualizing ECG waveforms as

semantic "Cardio-Morphemes." Through generative pre-training, our approach achieves competitive performance on the EchoNext and CSN datasets while significantly enhancing interpretability by mapping cardiac abnormalities to semantic anomalies. ShapeEFM effectively bridges the gap between data-driven representations and clinical physiology, offering a robust and explainable framework for cardiovascular diagnosis. In future work, we aim to explore the integration of multi-modal clinical data to further refine the Cardio-Morpheme vocabulary and expand the model's applicability to a broader spectrum of cardiovascular pathologies.

Acknowledgments. This work was supported by Ant Group, Ant Group Research Fund and The Second Affiliated Hospital of Zhejiang University School of Medicine.

Disclosure of Interests. The authors have no competing interests to declare.

References

1. Abbaspourazad, S., Elachqar, O., Miller, A.C., Emrani, S., Nallasamy, U., Shapiro, I.: Large-scale training of foundation models for wearable biosignals. In: The 12th International Conference on Learning Representations (2024)
2. Ba, J.L., Kiros, J.R., Hinton, G.E.: Layer normalization. arXiv preprint arXiv:1607.06450 (2016)
3. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: International conference on machine learning. pp. 1597–1607 (2020)
4. Chen, X., Xie, S., He, K.: An empirical study of training self-supervised vision transformers. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 9640–9649 (2021)
5. Dehghani, M., Djolonga, J., Mustafa, B., Padlewski, P., Heek, J., Gilmer, J., Steiner, A.P., Caron, M., Geirhos, R., Alabdulmohsin, I., et al.: Scaling vision transformers to 22 billion parameters. In: Proceedings of the 40th International Conference on Machine Learning. pp. 7480–7512 (2023)
6. Gedon, D., Ribeiro, A.H., Wahlström, N., Schön, T.B.: First steps towards self-supervised pretraining of the 12-lead ecg. In: 2021 Computing in Cardiology. vol. 48, pp. 1–4 (2021)
7. Gow, B., Pollard, T., Nathanson, L.A., Johnson, A., Moody, B., Fernandes, C., Greenbaum, N., Waks, J.W., Eslami, P., Carbonati, T., et al.: Mimic-iv-ecg: Diagnostic electrocardiogram matched subset. PhysioNet (2023)
8. Grill, J.B., Strub, F., Altché, F., Tallec, C., Richemond, P., Buchatskaya, E., Doersch, C., Avila Pires, B., Guo, Z., Gheshlaghi Azar, M., et al.: Bootstrap your own latent-a new approach to self-supervised learning. Advances in neural information processing systems **33**, 21271–21284 (2020)
9. Hendrycks, D., Gimpel, K.: Gaussian error linear units (gelus). arXiv preprint arXiv:1606.08415 (2016)
10. Jin, J., Wang, H., Li, H., Li, J., Pan, J., Hong, S.: Reading your heart: Learning ecg words and sentences via pre-training ecg language model. arXiv preprint arXiv:2502.10707 (2025)

11. Lan, X., Ng, D., Hong, S., Feng, M.: Intra-inter subject self-supervised learning for multivariate cardiac signals. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 36, pp. 4532–4540 (2022)
12. Le, X.M., Luo, L., Aickelin, U., Tran, M.T.: Shapeformer: Shapelet transformer for multivariate time series classification. In: Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. pp. 1484–1494. Barcelona, Spain (2024)
13. McKeen, K., Masood, S., Toma, A., Rubin, B., Wang, B.: Ecg-fm: An open electrocardiogram foundation model. Jamia Open **8**(5), ooaf122 (2025)
14. Na, Y., Park, M., Tae, Y., Joo, S.: Guiding masked representation learning to capture spatio-temporal relationship of electrocardiogram. In: The 12th International Conference on Learning Representations (2024)
15. Oh, J., Chung, H., Kwon, J.m., Hong, D.g., Choi, E.: Lead-agnostic self-supervised learning for local and global representations of electrocardiogram. In: Conference on Health, Inference, and Learning. pp. 338–353 (2022)
16. Peng, Z., Dong, L., Bao, H., Ye, Q., Wei, F.: Beit v2: Masked image modeling with vector-quantized visual tokenizers. arXiv preprint arXiv:2208.06366 (2022)
17. Poterucha, T.J., Jing, L., Ricart, R.P., Adjei-Mosi, M., Finer, J., Hartzel, D., Kelsey, C., Long, A., Rocha, D., Ruhl, J.A., et al.: Detecting structural heart disease from electrocardiograms using ai. Nature **644**(8075), 221–230 (2025)
18. Sarkar, P., Etemad, A.: Self-supervised ecg representation learning for emotion recognition. IEEE Transactions on Affective Computing **13**(3), 1541–1554 (2020)
19. Serhal, H., Abdallah, N., Marion, J.M., Chauvet, P., Oueidat, M., Humeau-Heurtier, A.: Overview on prediction, detection, and classification of atrial fibrillation using wavelets and ai on ecg. Computers in Biology and Medicine **142**, 105168 (2022)
20. Siontis, K.C., Noseworthy, P.A., Attia, Z.I., Friedman, P.A.: Artificial intelligence-enhanced electrocardiography in cardiovascular disease management. Nature Reviews Cardiology **18**(7), 465–478 (2021)
21. Su, J., Ahmed, M., Lu, Y., Pan, S., Bo, W., Liu, Y.: Roformer: Enhanced transformer with rotary position embedding. Neurocomputing **568**, 127063 (2024)
22. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)
23. Wu, Y., He, K.: Group normalization. In: Proceedings of the European conference on computer vision. pp. 3–19 (2018)
24. Yue, Z., Wang, Y., Duan, J., Yang, T., Huang, C., Tong, Y., Xu, B.: Ts2vec: Towards universal representation of time series. In: Proceedings of the AAAI conference on artificial intelligence. vol. 36, pp. 8980–8987 (2022)
25. Zheng, J., Chu, H., Struppa, D., Zhang, J., Yacoub, S.M., El-Askary, H., Chang, A., Ehwerhemuepha, L., Abudayyeh, I., Barrett, A., et al.: Optimal multi-stage arrhythmia classification approach. Scientific reports **10**(1), 2898 (2020)
26. Zheng, J., Guo, H., Chu, H.: A large scale 12-lead electrocardiogram database for arrhythmia study (version 1.0. 0). PhysioNet **23**, 7 (2022)