MSS-Net: Learning Hierarchical Multi-Scale and Multimodal Representations for Complex Cardiac Disease Classification

Yuling Chen[1], Jiahao Xia[2], Xingwang Yao[1], Zhuangye Luo[1], Tianli Zhao[3], and Feng Zeng[1⋆]

[1] School of Computer Science and Engineering, Central South University, Changsha, China
[2] School of Computer Science and Engineering, Southeast University, Nanjing, China
[3] The Second Xiangya Hospital of Central South University, Changsha, China
fengzeng@csu.edu.cn

Abstract. Recent breakthroughs in deep learning have catalyzed the evolution of multi-label Electrocardiogram (ECG) classification for cardiovascular diseases. However, prevailing paradigms often suffer from feature fragmentation, which limits their efficacy in identifying complex comorbidities. Specifically, it manifests as a discontinuity in representation, driven by the scale discrepancy between local morphological details and global heartbeat rhythms. Addressing the difficulty of aggregating these fragmented representations for diagnosing complex comorbidities, we propose the Multi-scale Spectral-Semantic Network (MSS-Net), a framework comprising three key components: Adaptive Gated Spectral Injection (AGSI) module, Dual Cross-scale Interaction (DCI) module, and Prior-Guided Feature Aggregation (PGFA) module. Specifically, the AGSI module modulates rebalanced global features in the frequency domain to capture long-range rhythmic dependencies, while DCI module establishes dual interaction channels to synergize local morphological details with global rhythmic contexts, effectively mitigating feature fragmentation to capture features across scales. Subsequently, the PGFA module incorporates a co-learning mechanism that uses individualized clinical contexts as auxiliary supervision signals during training to calibrate complex pathological features. Extensive experiments on three datasets show the effectiveness of our method. MSS-Net outperforms the state-of-the-art by +2.03% in AUC, indicating robust discriminative capability in complex clinical scenarios.

Keywords: ECG Data · Complex Cardiac Disease Classification · Multi-scale Fusion · Feature Fragmentation · Multimodal Learning.

1 Introduction

Cardiovascular diseases remain a leading cause of global mortality, necessitating precise diagnostic systems for timely intervention [15, 24]. To alleviate the

⋆ Corresponding Author.

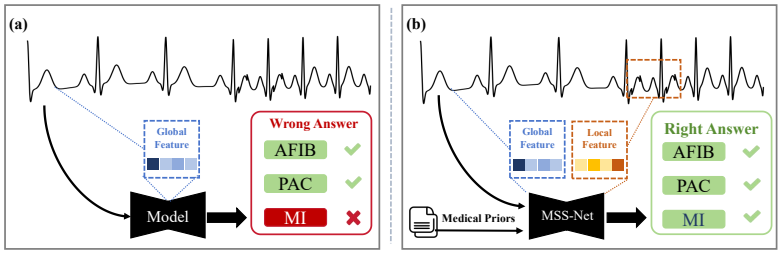

Fig. 1. Motivation and Challenges in Complex Cardiac Disease Classification. (a) Single-scale features often bias the model towards common diseases(e.g., AFIB and PAC), while neglecting the complex pathologies(e.g., MI). (b) The proposed MSS-Net enables precise recognition of complex diseases by synergizing multi-scale global-local features and calibrating representations with multimodal clinical text guidance.

unsustainable burden of manual interpretation, deep learning models like Convolutional Neural Networks (CNNs) have been widely adopted for clinical decision support [2,4]. However, clinical ECG diagnosis inherently requires observing both local morphological details and global rhythms to identify complex pathologies. Constrained by their localized receptive fields, standard CNNs struggle to capture long-term rhythmic trends. To overcome this limitation, hybrid architectures combining CNNs with other novel designs have emerged to simultaneously capture these dual-domain dependencies [8].

Despite the progress achieved by hybrid modeling, mainstream parallel extraction paradigms suffer from feature fragmentation, which refers to the semantic inconsistency and insufficient interaction between local morphological features and global rhythmic features and fails to dynamically align them [3,14,28]. This lack of complementary calibration disperses unified pathological patterns into independent, uncoordinated segments. Consequently, while parallel architectures may identify isolated conditions, they inherently struggle to capture the intricate dependencies required for classifying complex comorbidities. As illustrated in Fig. 1, omitting local details causes conventional models to miss subtle pathologies. While global features correctly identify rhythm disorders, they fail to detect complex pathologies like MI. Conversely, incorporating multi-scale ECG features and multi-modal text feature fusion explicitly captures these morphological variations, successfully recovering the missed comorbidities.

To summarize, current complex cardiac disease classfication faces the following challenges: (1) ECG signals are lengthy and contain intricate information, where global features often overshadow subtle yet crucial local fluctuations, making complex pathologies difficult to capture; (2) the representation capability of relying solely on ECG signals is limited, while clinical text descriptions lack precise localization. To address these challenges, we propose a novel Multi-scale Spectral-Semantic Network (MSS-Net) for precise complex cardiac disease classification. Specifically, for the input ECG signals, a morphological feature encoder is first employed to map them into semantic space, forming initial global fea-

tures. After rebalancing to address the class imbalance in global features, the Adaptive Gated Spectral Injection (AGSI) module is utilized to modulate and analyze features in frequency domain. Subsequently, the Dual Cross-scale Interaction (DCI) module leverages parallel branches to bridge and fuse the rebalanced global features with the modulated local features, enabling the selective aggregation of shallow morphological details. Finally, the Prior-Guided Feature Aggregation (PGFA) module extracts additional prior information from medical priors. Through alignment within the latent space, the module achieves multimodal supervision and fusion, ensuring robust prediction of complex diseases.

Our main contributions are summarized as follows:

- We propose MSS-Net, the first to combine cross-modal semantic alignment with spectral-temporal synergistic learning for complex cardiac disease classification, indicating promising potential for clinical diagnostic support.
- We design three novel modules: AGSI, DCI, and PGFA. These modules improve conventional ECG analysis through frequency-domain modulation, multi-scale feature fusion, and multi-modal information calibration, respectively. This integrated approach reduces bias in global signal representation, enhances the discriminative power of subtle local features, and incorporates medical priors to ensure the robustness of predictions.
- Extensive comparative experiments and ablation studies are conducted to validate our framework. MSS-Net achieves significant performance gains, including a 2.03% of AUC improvement on the PTB-XL [21] and 1.17% of AUC improvement on the CPSC2018 [11], showing the superior performance of MSS-Net and the individual effectiveness of its architectural components.

2 Related Work

2.1 Current Paradigms in ECG Classification

Early ECG classification primarily utilized CNNs for localized feature extraction [7, 13]. To overcome the receptive field limitations of CNNs in capturing global rhythms, hybrid architectures integrating Transformers have emerged [1, 27]. For instance, ProtoECGNet [20] decouples global rhythms and local waveforms through parallel branches. ECG-Mamba [18] leverages selective scan mechanisms to model extensive temporal dependencies with linear complexity. However, most prevailing paradigms adhere to a parallel design, relying on straightforward concatenation. Such a decoupled structure often leads to feature fragmentation, where morphological and rhythmic features remain semantically isolated, failing to capture the deep synergies essential for diagnosing complex comorbidities.

2.2 Hybrid Architectures with Multi-scale Modeling

To capture both subtle details and global trends, researchers have explored diverse multi-scale architectural designs, including multi-branch convolutions with

varying kernel sizes [5, 25], dual-resolution attention networks [9], and hierarchical feature extraction frameworks [6, 12, 17, 19, 23]. However, these paradigms using simplistic aggregation may exacerbate feature fragmentation in complex comorbidities, scattered feature fragments are difficult to reconstruct into a comprehensive pathological representation.

In summary, although existing hybrid architectures have improved the richness of features, they have not addressed the underlying issue of feature fragmentation. There is a need for a unified architecture capable of facilitating the interaction between morphology and rhythm alongside cross-modal semantic guidance. To this end, we propose MSS-Net, a method using multi-scale and multi-modal features for precise classification of complex cardiac diseases.

3 Method

3.1 Overview

We formulate complex cardiac disease classification as a mapping problem from the signal space $\mathcal{X}$ to the pathological label space $\mathcal{Y}$. Given an input signal $X \in \mathbb{R}^{12\times L}$, our goal is to learn a parameterized function $F_\theta : \mathcal{X} \to \mathcal{Y}$ to output a predicted probability vector that approximates the ground truth labels $Y \in \{0, 1\}^K$, where L denotes length of single signal and K indicates the number of classifications. As illustrated in Fig. 2, MSS-Net adopts a hierarchical collaborative architecture comprising three core components: AGSI, DCI and PGFA. For the input ECG signal X, a morphological feature encoder E_m is first employed to obtain initial global feature F_1, which captures the macroscopic spatial-temporal characteristics of the signal. To extract subtle local fluctuations that are easily overlooked in global representations and to compensate for the limited representation capability of unimodal ECG signals, three core components are designed to refine traditional ECG feature extraction from three perspectives: frequency-domain signal modulation, multi-scale feature fusion and multi-modal information calibration. This integrated approach collectively enhances the model's analytical and classification performance for complex diseases.

3.2 Adaptive Gated Spectral Injection

To overcome the limited receptive fields inherent in CNNs, we leverage large-kernel grouped convolutions and channel competition mechanisms to extract high-fidelity morphological features. Given the spatial independence of ECGs, we employ grouped convolutions to model each lead separately. The encoder comprises three layers utilizing large 31×1 kernels. Since ECG features exhibit inherent class imbalance, models often overfocus on common pathological patterns during feature extraction, thereby neglecting the distribution fitting of rare diseases. To mitigate this, we incorporate the Global Response Normalization (GRN) [22] unit to rebalance weights across the feature channel dimension. This mechanism effectively highlights rare disease features and enhances the attention

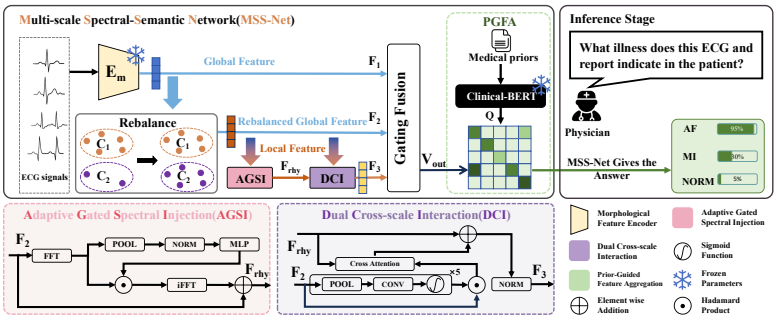

Fig. 2. Overview of MSS-Net. ECG signals are processed by the Morphological Encoder E_m to extract hierarchical features (F_1, F_2). F_2 is modulated by **AGSI** module to capture rhythmic context F_{rhy}, which is subsequently synergized with F_2 via the **DCI** module to yield F_3 to align local morphology with global rhythms. Finally, the gated fusion of $\{F_1, F_2, F_3\}$ yields panoramic features V_{out}, which are calibrated by the **PGFA** module using frozen medical model priors for prediction.

of subsequent modules toward these subtle pathological signatures. Formally, given the input $X \in \mathbb{R}^{12\times L}$, the rebalanced global feature $\mathbf{F}_2$ is expressed as:

$$\mathbf{F}_2 = \mathbf{W}_2 \cdot \mathrm{GRN}(\mathrm{GELU}(\mathbf{W}_1 \cdot \mathrm{LN}(\mathrm{DConv}_{31\times 1}(X)))), \tag{1}$$

where $\mathbf{W}_1, \mathbf{W}_2$ are point-wise convolutions for channel expansion and projection.

Despite CNNs' proficiency in morphological recognition, they often overlook the rhythmic consistency inherent in periodic ECG signals. To address this, we first map the feature into the frequency domain via Fourier Transform: $\mathbf{f} = \mathcal{F}(\mathbf{F}_2)$. To facilitate frequency feature extraction, we construct the spectral frequency statistics $\mathbf{s} = \mathcal{P}(\mathbf{f})$, where $\mathcal{P}$ denotes a frequency statistical operator that performs average pooling and normalization on the spectrum $\mathbf{f}$.

For the statistical information $\mathbf{s}$, we define a learnable spectral response function $\gamma = g(\mathbf{s}) = \mathbf{W}_3\sigma(\mathbf{W}_4\mathbf{s})$, which serves as a modulation weight for each channel. This enables frequency-domain information to be fused with the original ECG features through an inverse transform and a residual connection, yielding the rhythmic context $\mathbf{F}_{rhy} = \mathcal{F}^{-1}(\gamma \cdot \mathbf{f}) + \mathbf{F}_2$.

3.3 Dual Cross-scale Interaction

To bridge the semantic gap between $\mathbf{F}_{rhy}$ and $\mathbf{F}_2$, the DCI module utilizes deep features as contextual guidance to aggregate relevant morphological details $\mathbf{H}_{DCI}$. By aggregating channel-wise statistics, we generate a spatial mask to produce the denoised representation $\tilde{\mathbf{F}}_2$. We then employ cross-scale attention:

$$\mathbf{H}_{DCI} = \mathrm{Softmax}\left(\frac{(\mathbf{F}_{rhy}\mathbf{W}_Q)(\tilde{\mathbf{F}}_2\mathbf{W}_K)^T}{\sqrt{d_k}}\right)(\tilde{\mathbf{F}}_2\mathbf{W}_V). \tag{2}$$

To resolve semantic misalignment, we introduce an adaptive residual injection mechanism with a learnable scale α:

$$\begin{aligned} \alpha &= \sigma(\text{MLP}(\text{Concat}(\mathbf{F}_{rhy}, \mathbf{H}_{DCI}))) \\ \mathbf{F}_3 &= \mathbf{F}_{rhy} + \alpha \cdot \mathbf{H}_{DCI} \end{aligned}. \tag{3}$$

To aggregate the multi-scale feature set $\{\mathbf{F}_1, \mathbf{F}_2, \mathbf{F}_3\}$ generated in the preceding stages, we introduce a Gating Fusion mechanism, which adaptively calibrates the contribution of each scale by calculating dynamic gating weights. Specifically, the gating weights $\mathbf{g} = [g_1, g_2, g_3]$ are captured via a compact bottleneck structure conditioned on the global average pooling of concatenated features:

$$\mathbf{g} = \text{Softmax}(\mathbf{W}_u \sigma(\mathbf{W}_d \cdot \text{AvgPool}([\mathbf{F}_1, \mathbf{F}_2, \mathbf{F}_3]))), \tag{4}$$

where $\mathbf{W}_d$ and $\mathbf{W}_u$ denote the down-projection and up-projection weights.

Finally, the input ECG signal X is transformed into the latent representation $\mathbf{V}_{out} = \sum_{i=1}^{3} g_i \cdot \mathbf{F}_i$ in the feature space, serving as the panoramic representation.

3.4 Prior-Guided Feature Aggregation

To address feature fragmentation, PGFA employs a frozen ClinicalBERT to encode class-level diagnostic descriptions into medical prior anchors $\mathbf{Q}$. Notably, these descriptions only provide disease-related physiological and diagnostic feature rather than sample-level information. Using $\mathbf{Q}$ as the query and the panoramic feature $\mathbf{V}_{out}$ as the key/value, we perform multi-head cross-attention (as in Eq. 2) to aggregate prior-guided pathological features $\mathbf{H}_{PGFA}$. The final prediction is generated via $\hat{y} = \sigma(\text{LN}(\mathbf{H}_{PGFA} + \mathbf{Q})\mathbf{W}_{cls} + \mathbf{b}_{cls})$, where $\mathbf{W}_{cls}$ and $\mathbf{b}_{cls}$ are the weights and bias of the linear classifier.

We calibrate the decision boundary by injecting class priors into raw logits: $x'_k = x_k - \tau \cdot \log(\pi_k)$, where π_k denotes the empirical frequency of class k calculated from the training set, and τ is a scaling hyperparameter. Let $p_k = \sigma(x'_k)$ denote the calibrated predicted probability. Loss function, integrated with label smoothing $y_{sm}^{(k)}$, is formulated as:

$$\mathcal{L}_{total} = -\sum_k \left[y_{sm}^{(k)} (1 - p_k)^{\gamma_+} \log(p_k) + (1 - y_{sm}^{(k)}) p_k^{\gamma_-} \log(1 - p_k) \right]. \tag{5}$$

4 Experiments and Results

4.1 Experimental Settings

Datasets. We evaluated our method on three public datasets. PTB-XL [21] contains 21837 records from 18885 patients, spanning 5 diagnostic super-classes, 23 sub-classes and 44 diag-classes. CPSC2018 [11] consists of 6877 records, annotated with 9 categories of cardiac abnormalities. Shaoxing [29] dataset comprises 10646 recordings, providing a diverse set of rhythm and morphological labels for robust validation. All datasets are partitioned by patient ID into training, validation, and testing sets in an 8:1:1 ratio to prevent data leakage. All ECG signals are resampled to 500 Hz and standardized to a fixed duration of 10 seconds.

Table 1. Main results across three datasets. We report AUC and F1 (%).**Sub.**, and **Sup.** denote the two hierarchical sub-categories of PTB-XL which are Sub-diag and Sup-diag. The best results are in **bold**, the second results are in underlined.

Method	PTB-XL (All)		Diag		Sub.		Sup.		CPSC2018		Shaoxing	
	AUC	F1	AUC	F1	AUC	F1	AUC	F1	AUC	F1	AUC	F1
InceptionTime [10]	91.07	71.69	92.81	65.09	92.32	67.29	92.52	76.37	94.14	73.93	81.53	62.36
MVMSNet [25]	92.46	72.43	92.92	66.09	92.77	69.42	92.96	77.35	94.34	75.19	82.14	63.50
ECG-Mamba [18]	91.93	73.14	93.21	68.32	92.89	69.96	92.10	75.89	96.31	79.54	84.39	65.83
SelfONN [19]	92.84	73.24	92.30	67.33	92.41	68.14	92.66	75.21	95.60	80.20	83.97	65.23
ProtoECGNet [20]	91.32	72.15	92.85	66.50	92.55	68.20	93.10	76.80	96.55	79.85	84.86	66.27
L5G-Net [26]	**93.41**	73.50	93.95	67.80	93.80	71.20	93.57	77.06	96.82	**80.85**	85.44	66.86
MTA-Net [16]	93.31	72.63	94.62	68.12	93.70	**72.13**	95.10	78.50	96.95	80.05	85.64	67.16
MSS-Net (Ours)	93.36	**74.65**	**96.65**	**70.45**	**94.85**	71.85	**96.88**	**79.55**	**98.12**	80.55	**87.65**	**69.25**

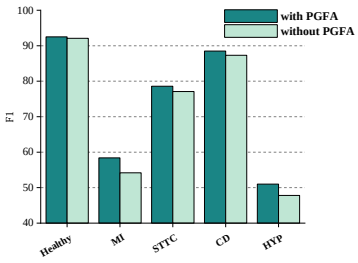

Fig. 3. Impact of PGFA.

Table 2. Ablation study on PTB-XL (Diag). Values (↑) represent the performance gain compared to the baseline.

AGSI	DCI	PGFA	AUC	F1
×	×	×	93.03	63.52
✓	×	×	93.88 ↑0.85	64.91 ↑1.39
✓	✓	×	95.50 ↑2.47	68.33 ↑4.81
✓	×	✓	95.03 ↑2.00	67.46 ↑3.94
×	✓	✓	95.80 ↑2.77	68.72 ↑5.20
✓	✓	✓	**96.65** ↑3.62	**70.45** ↑6.93

Implementation Details. All experiments were performed on an NVIDIA RTX 4090 GPU using PyTorch 2.1. For network optimization, we employed the Adam optimizer with an initial learning rate of 0.001 and a weight decay of 0.1. The batch size was set to 64, and the training process spanned 100 epochs. Model performance is evaluated using macro-averaged AUC and F1-score to assess the performance across the comprehensive spectrum of cardiac abnormalities. The source code is available at https://github.com/origami-one/MSS-Net.

4.2 Results and Comparison with State-of-the-art

Tables 1 summarizes the classification performance of MSS-Net on three public datasets. Overall, MSS-Net surpasses current SOTA in AUC and F1. The proposed multi-scale fusion architecture facilitates robust feature extraction by mitigating feature fragmentation. Especially in the diag-classes of PTB-XL, it indicates substantial performance improvements. For instance, MSS-Net significantly outperforms the classical InceptionTime by a large margin while surpassing the MTA-Net by +2.03% and +2.33% in AUC and F1, respectively.

For the CPSC2018 dataset, it establishes a new SOTA in AUC (98.12%), showing a +1.17% gain over MTA-Net, which indicates that our architecture can effectively suppress false positives while maintaining high sensitivity in multi-

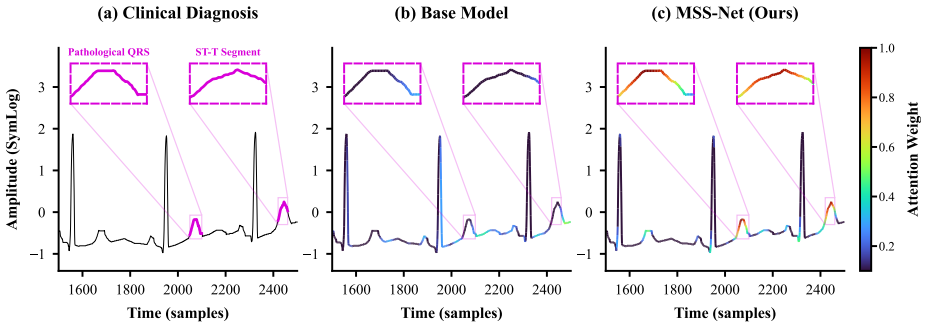

Fig. 4. Feature Heatmap Visualization with MSS-Net.

label arrhythmia classification. For the Shaoxing dataset, MSS-Net achieves a comprehensive breakthrough by yielding both the highest AUC(87.65%) and F1(69.25%), maintaining superior robustness even when faced with the high rhythm diversity and morphological complexity of clinical ECGs.

4.3 Ablation Study

Table 2 evaluates the contribution of each proposed component. Incorporating AGSI leads to an initial improvement of 1.39% in F1, indicating that AGSI effectively captures long-range dependencies in the frequency domain, alleviating the receptive field limitations in time-domain convolutions. DCI triggers the most substantial performance leap, boosting the F1 by 4.81% relative to the baseline. It highlights DCI is the most critical tool for bridging the semantic gap between local morphological waves and global rhythms, which is frequently overlooked by common approaches. PGFA completes the MSS-Net architecture, yielding a cumulative gain of 3.62% in AUC over baseline. The enhancement directly contributed by PGFA is particularly pronounced in categories associated with complex pathologies, such as MI, as visualized in Fig. 3, indicating PGFA leverages medical priors to reconstruct fragmented features for diagnosing complex comorbidities. Therefore, the results confirm that removing any stage disrupts the progressive chain, which proves MSS-Net is not an arbitrary module stack, but a synergistic design targeted at resolving feature fragmentation.

4.4 Feature Heatmap Visualization

As shown in Fig. 4, we evaluate the interpretability of MSS-Net using a myocardial infarction sample with explicitly highlighted abnormal ST-T intervals. (a) is clinical diagnosis of MI. In the magnified dashed boxes of (b), baseline is distracted by irrelevant waveforms, assigning low attention weights (cool colors) to these critical features, potentially leading to a missed diagnosis. In contrast, MSS-Net precisely aggregates high attention weights (warm colors) onto the clinical regions. This indicates our strategy effectively mitigates feature fragmentation by precisely capturing subtle pathological features in complex comorbidities.

5 Conclusion

We propose MSS-Net, a novel framework for complex cardiac disease classification that mitigates feature fragmentation by synergizing local morphological details with global rhythms. By integrating multi-scale features and leveraging medical priors, MSS-Net accurately captures complex pathological representations. Extensive experiments on large clinical datasets indicate its superior discriminative capability. Although MSS-Net effectively improves diagnostic accuracy, it cannot be denied that its complex structure reduces efficiency. Therefore, future work will explore lightweight adaptations to reduce computational overhead, enabling deployment on real-time wearable devices.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

Acknowledgment. This work is supported in part by the National Science Foundation of China (Grant No. 62172450).

References

1. Ashhad, M., Rahmani, S., Fayiz, M., Etemad, A., Hashemi, J.: Uncertainty-aware multi-view arrhythmia classification from ecg. In: 2024 International Joint Conference on Neural Networks (IJCNN). pp. 1–6. IEEE (2024)
2. Bai, S.: An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv preprint arXiv:1803.01271 (2018)
3. Chen, J., Dong, X., Wang, W., Zhou, S., Yu, L., Hu, X.: Deri: Cross-modal ecg representation learning with deep ecg-report interaction. In: 34th Internationa Joint Conference on Artificial Intelligence, IJCAI 2025. pp. 4824–4832. International Joint Conferences on Artificial Intelligence (2025)
4. Chmelevsky, M., Budanova, M., Khamzin, S.: Development of a novel machine learning-based methodology for the differential diagnosis of wide qrs complex arrhythmias using automated analysis of 12-lead ecg. In: 2023 Computing in Cardiology (CinC). vol. 50, pp. 1–4. IEEE (2023)
5. Fan, X., Yao, Q., Cai, Y., Miao, F., Sun, F., Li, Y.: Multiscaled fusion of deep convolutional neural networks for screening atrial fibrillation from single lead short ecg recordings. IEEE journal of biomedical and health informatics **22**(6), 1744–1753 (2018)
6. Guo, L., Zhan, Q., Ma, N., An, Y.: Damal: Data augmentation aided multi-channel attention fusion network for multi-label myocardial infarction localization. In: 2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp. 1988–1995. IEEE (2024)
7. Hannun, A.Y., Rajpurkar, P., Haghpanahi, M., Tison, G.H., Bourn, C., Turakhia, M.P., Ng, A.Y.: Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nature medicine **25**(1), 65–69 (2019)

8. Hu, R., Chen, J., Zhou, L.: A transformer-based deep neural network for arrhythmia detection using continuous ecg signals. Computers in Biology and Medicine **144**, 105325 (2022)
9. Huang, W., Wang, N., Feng, P., Wang, H., Wang, Z., Zhou, B.: A multi-resolution mutual learning network for multi-label ecg classification. In: 2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp. 3303–3306. IEEE (2024)
10. Ismail Fawaz, H., Lucas, B., Forestier, G., Pelletier, C., Schmidt, D.F., Weber, J., Webb, G.I., Idoumghar, L., Muller, P.A., Petitjean, F.: Inceptiontime: Finding alexnet for time series classification. Data Mining and Knowledge Discovery **34**(6), 1936–1962 (2020)
11. Liu, F., Liu, C., Zhao, L., Zhang, X., Wu, X., Xu, X., Liu, Y., Ma, C., Wei, S., He, Z., et al.: An open access database for evaluating the algorithms of electrocardiogram rhythm and morphology abnormality detection. Journal of Medical Imaging and Health Informatics **8**(7), 1368–1373 (2018)
12. Liu, S., Yu, H., Liao, C., Li, J., Lin, W., Liu, A.X., Dustdar, S.: Pyraformer: Low-complexity pyramidal attention for long-range time series modeling and forecasting. In: International conference on learning representations (2021)
13. Liu, X., Wang, H., Li, Z., Qin, L.: Deep learning in ecg diagnosis: A review. Knowledge-Based Systems **227**, 107187 (2021)
14. Liu, Y., Hu, T., Zhang, H., Wu, H., Wang, S., Ma, L., Long, M.: itransformer: Inverted transformers are effective for time series forecasting. arXiv preprint arXiv:2310.06625 (2023)
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. PMLR (2022)
16. Pang, S., Kan, Z., Zhao, Z., Wu, W., Ding, H., Qiao, S.: Mta-net: A multi-scale temporal-attentive network with semantic-structural fusion for multi-label ecg classification. In: 2025 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp. 1–6. IEEE (2025)
17. Prabhakararao, E., Dandapat, S.: Multi-scale convolutional neural network ensemble for multi-class arrhythmia classification. IEEE Journal of Biomedical and Health Informatics **26**(8), 3802–3812 (2021)
18. Qiang, Y., Dong, X., Liu, X., Yang, Y., Hu, F., Wang, R.: Ecgmamba: Towards ecg classification with state space models. In: 2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp. 6498–6505. IEEE (2024)
19. Qin, K., Huang, W., Zhang, T., Zhang, H., Cheng, X.: A lightweight selfonn model for general ecg classification with pretraining. Biomedical signal processing and control **89**, 105780 (2024)
20. Sethi, S., Chen, D., Statchen, T., Burkhart, M.C., Bhandari, N., Ramadan, B., Beaulieu-Jones, B.: Protoecgnet: case-based interpretable deep learning for multi-label ecg classification with contrastive learning. Proceedings of machine learning research **298**, https–proceedings (2025)
21. Wagner, P., Strodthoff, N., Bousseljot, R.D., Kreiseler, D., Lunze, F.I., Samek, W., Schaeffter, T.: Ptb-xl, a large publicly available electrocardiography dataset. Scientific data **7**(1), 1–15 (2020)
22. Woo, S., Debnath, S., Hu, R., Chen, X., Liu, Z., Kweon, I.S., Xie, S.: Convnext v2: Co-designing and scaling convnets with masked autoencoders. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 16133–16142 (2023)

23. Wu, H., Hu, T., Liu, Y., Zhou, H., Wang, J., Long, M.: Timesnet: Temporal 2d-variation modeling for general time series analysis. arXiv preprint arXiv:2210.02186 (2022)
24. Xia, J., Zhang, X., He, Y., Qi, Y., Hu, Y., Haigron, P., Tang, C., Zhang, L., Yang, G.: Aegis: Using conditional multi-view diffusion model to achieve angiographic enhancement in non-contrast ct. IEEE Transactions on Circuits and Systems for Video Technology (2025)
25. Yang, S., Lian, C., Zeng, Z., Xu, B., Zang, J., Zhang, Z.: A multi-view multi-scale neural network for multi-label ecg classification. IEEE Transactions on Emerging Topics in Computational Intelligence **7**(3), 648–660 (2023)
26. Yang, T., Xie, C.X., Huang, H.M., Wang, Y., Fan, M.H., Kuo, I.C., Chen, T.Y., Chen, S.L., Chen, C.A., Abu, P.A.R., et al.: Lead analysis for the classification of multi-label cardiovascular diseases and neural network architecture design. Electronics **14**(16), 3211 (2025)
27. Yang, Z., Jin, A., Li, Y., Yu, X., Xu, X., Wang, J., Li, Q., Guo, X., Liu, Y.: A coordinated adaptive multiscale enhanced spatio-temporal fusion network for multi-lead electrocardiogram arrhythmia detection. Scientific Reports **14**(1), 20828 (2024)
28. Zeng, A., Chen, M., Zhang, L., Xu, Q.: Are transformers effective for time series forecasting? In: Proceedings of the AAAI conference on artificial intelligence. vol. 37, pp. 11121–11128 (2023)
29. Zheng, J., Zhang, J., Danioko, S., Yao, H., Guo, H., Rakovski, C.: A 12-lead electrocardiogram database for arrhythmia research covering more than 10,000 patients. Scientific data **7**(1), 48 (2020)