

Versatile and Risk-Sensitive Cardiac Diagnosis via Graph-Based ECG Signal Representation

Yue Wang, Yuyang Xu, Renjun Hu, Fanqi Shen, Hanyun Jiang, Jun Wang, Jintai Chen, Danny Z. Chen *Fellow, IEEE*, Jian Wu, *Member, IEEE*, and Haochao Ying, *Member, IEEE*

Abstract—Despite the rapid advancements of electrocardiogram (ECG) signal diagnosis and analysis methods through deep learning, two major hurdles still limit their clinical adoption: the lack of versatility in processing ECG signals with diverse configurations, and the inadequate detection of risk signals due to sample imbalances. Addressing these challenges, we introduce **Versatile and Risk-Sensitive cardiac diagnosis (VARS)**, an innovative approach that employs a graph-based representation to uniformly model heterogeneous ECG signals. VARS stands out by transforming ECG signals into versatile graph structures that capture critical diagnostic features, irrespective of signal diversity in the lead count, sampling frequency, and duration. This graph-centric formulation also enhances diagnostic sensitivity, enabling precise localization and identification of abnormal ECG patterns that often elude standard analysis methods. To facilitate representation transformation, our approach integrates denoising reconstruction with contrastive learning to preserve raw ECG information while highlighting pathognomonic patterns. We rigorously evaluate the efficacy of VARS on three distinct ECG datasets, encompassing a range of structural variations. The results demonstrate that VARS not only consistently surpasses existing state-of-the-art models across all these datasets but also exhibits substantial improvement in identifying risk signals. Additionally, VARS offers interpretability by pinpointing the exact waveforms that lead to specific model outputs, thereby assisting clinicians in making informed decisions. These findings suggest that our VARS will likely emerge as an invaluable tool for comprehensive cardiac health assessment.

This research was supported by the National Natural Science Foundation of China under Grant No. 62476246, the Zhejiang Provincial Natural Science Foundation of China under Grant No. LY23F020019, and the Opening Foundation of the State Key Laboratory of Transvascular Implantation Devices under Grant No. SKLTID2024003. (Corresponding Authors: Haochao Ying and Jian Wu.)

Yue Wang and Yuyang Xu are with the College of Computer Science and Technology, Zhejiang University, Hangzhou 310012, China. They are also with the State Key Laboratory of Transvascular Implantation Devices of the Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou 310009, China, and Zhejiang Key Laboratory of Medical Imaging Artificial Intelligence, Hangzhou 310058, China. (E-mail: ywang2022@zju.edu.cn, xuyuyang@zju.edu.cn).

Renjun Hu is with the Alibaba Group, Hangzhou 310023, China.

Fanqi Shen and Hanyun Jiang are with the College of Computer Science and Technology, Zhejiang University, Hangzhou 310012, China.

Jun Wang is with the School of Computer and Computational Science, Hangzhou City University, Hangzhou 310015, China.

Jintai Chen is with the Information Hub, Hong Kong University of Science and Technology (Guangzhou), Guangzhou, 511458, China.

Danny Z. Chen is with the Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN 46556, USA.

Jian Wu is with the State Key Laboratory of Transvascular Implantation Devices of the Second Affiliated Hospital and School of Public Health, Zhejiang University School of Medicine, Hangzhou 310009, China. He is also with Zhejiang Key Laboratory of Medical Imaging Artificial Intelligence, Hangzhou 310058, China.

Haochao Ying is with the School of Public Health and Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou 310058, China. (E-mail: haochaoying@zju.edu.cn)

Index Terms—ECG Representation, Cardiac Diagnosis, Denoising Reconstruction, Contrastive Learning

I. INTRODUCTION

Electrocardiogram (ECG) stands as the most accessible and cost-effective technique for detecting a wide range of cardiac anomalies, such as atrial fibrillation, myocardial ischemia, and hypokalemia, among others [1]–[6]. This diagnostic method involves the attachment of electrodes at specified locations on the body skin near the heart to monitor the electrical activity of the heart muscle. In practice, ECG devices typically utilize between 2 to 12 leads, acquiring data at sampling frequencies ranging from 250 to 1000 Hz for varying durations. Clinicians then rely on these data to diagnose conditions such as arrhythmias and structural cardiac abnormalities by scrutinizing crucial yet frequently subtle aspects like heart rhythm, waveform morphology, and timing intervals between beats. Not surprisingly, manual diagnosis is time-consuming and requires high-level expertise. Consequently, the development of model-based approaches for automated cardiac diagnosis has recently emerged as a significant area of research.

However, the heterogeneity of ECG data, accompanied by patient-specific variables and a diverse array of monitoring equipment, introduces significant challenges to diagnosis. A spectrum of sophisticated models, especially those based on Deep Neural Networks (DNNs), has been proposed. Despite the effectiveness of DNNs for ECG diagnostics [7]–[12], their applications remain constrained to analysis within specific data types, owing to employing pre-defined feature extraction modules and classification labels. The fundamental differences in the number of leads, sampling frequencies, and durations of ECG recordings have systematically challenged the versatility of the representational capabilities of traditional DNN models [13]. It turns out that, while these models perform admirably on certain datasets, their ability to consistently recognize and differentiate intricate patterns across diverse ECG signals remains questionable [14]. Additionally, their performance is compromised in the face of class imbalance within ECG datasets, indicating a notable deficiency in detecting risk signals [15]. The over-reliance on labels further undermines the model performance and their capacity to generalize, impeding the learning of effective ECG signal representations. In summary, current ECG signal representation and processing approaches exhibit two critical limitations: the lack of versatility in processing ECG signals with diverse configurations, and the inadequate detection of risk signals due to sample imbalances.

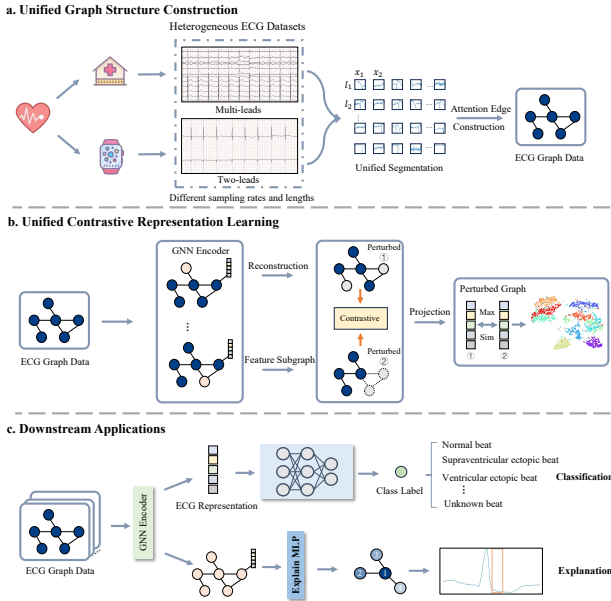


Fig. 1. **Illustrating our VARS approach.** The approach involves three parts: unified structure construction, representation learning, and downstream tasks. **a.** Heterogeneous ECG datasets, including multi-lead and two-lead configurations with varied sampling rates and lengths, are transformed into a unified signal structure to ensure consistency. **b.** A contrastive learning method with a Graph Neural Network (GNN) encoder learns a unified ECG representation, with feature subgraphs and perturbed graphs, enhancing robustness. **c.** These graph structures support downstream tasks like classification and interpretation, categorizing signals into normal beats, supraventricular ectopic beats, ventricular ectopic beats, and other types.

To overcome these challenges, we propose an innovative approach, Versatile and Risk-Sensitive cardiac diagnosis (VARS), which leverages a graph-based representation to handle the heterogeneity of ECG signals for versatile and risk-sensitive cardiac health assessment. We develop a new method for transforming ECG signals into a universal graph structure. To achieve the goal of unified data abstraction, we segment an ECG sequence with a specific time interval, and each time interval is taken as a graph node. Correspondingly, different leads are converted into different sets of nodes. This approach aligns well with cardiologists' clinical knowledge viewing the features of a time interval (PR interval) as a pattern unit to assess diseases. We then calculate the distances between the nodes to compose the ECG graphs so that semantic correlations between raw signal sequences are preserved. This graph-based abstraction has the potential to capture key diagnostic features while remaining unaffected by variations in lead count, sampling frequency, and signal duration.

ECG data often present the 'lost-in-the-middle' quandary, where critical diagnostic decisions hinge upon a minority of waveforms within the full recording. These small yet significant patterns are known to be difficult to detect, especially in the situations of class imbalances. As illustrated in Figure 1, our graph-centric perspective aims to increase diagnostic sensitivity, enabling precise localization and identification of those elusive abnormal ECG patterns that conventional methods struggle to discern. Specifically, VARS combines denoising reconstruction [16] and contrastive learning [17] for

this purpose. Denoising reconstruction can filter out irrelevant or misleading noise from raw signals without damaging the integrity of the original data, hence preserving the original ECG information while enhancing representational effectiveness. This step is crucial for maintaining the signals' purity and usability. Additionally, contrastive learning plays an essential role in enhancing the model's ability to differentiate among diagnostic patterns. By comparing different signal patterns and amplifying the defining features of each, this technique facilitates a more detailed and nuanced understanding of the data. However, classic contrastive learning methods cannot effectively capture subtle key information from the entire ECG graphs. To address this, we implement subgraph extraction and representation for contrastive learning. In this way, we enhance discrimination among signals and widen the distance between samples of different categories in the feature space, thereby improving the model's representation ability for classification.

We conduct extensive experiments on three distinct ECG datasets (*i.e.*, MITBIH, PTB-XL, and ST-T) that encompass a range of structural variations to evaluate the efficacy of VARS. We show that VARS consistently surpasses 12 baselines across these datasets in all the tested metrics. Moreover, VARS also exhibits more substantial improvement in identifying risk signals. For instance, in the anomaly categories, VARS outperforms the baselines by a factor of ten compared to its improvement in the overall classification. Additionally, VARS enhances the F1 score by $\sim 6.4\%$ across the three datasets for the risk categories. We present an ablation study to empirically justify the technical designs behind VARS. Finally, our case study and visual analysis results further confirm the advantages of VARS in interpretability (by pinpointing the exact waveforms that lead to specific model outputs) and representation (with better categorical separation).

The main contributions of our work are as follows:

- We study versatile and risk-sensitive cardiac diagnosis and propose a novel transformation from ECG signals to graphs, which is among the first in the literature.
- We develop the new VARS approach that integrates denoising reconstruction and contrastive learning to facilitate the representation transformation and signal analysis.
- We conduct extensive experiments to demonstrate the multi-dimensional superiority of VARS on versatility for signal diversity, risk-sensitivity, and interpretability.

II. RELATED WORK

A. Graph Structure Construction

When applying Graph Neural Networks (GNNs) to data that do not directly provide a graph structure, it is essential to construct a graph representation to enable effective GNN processing. Typically, there are two main types of approaches for constructing graphs. One type involves calculating the similarity between nodes based on their features within the dataset, while the other type constructs a graph by leveraging the explicit physical associations among the data elements. In the first type, Vision GNN [18] divided an image into multiple patches, converted each patch into a one-dimensional vector

to form a node representation of the image, and computed k -nearest neighbors of each node to form edges between the nodes. Time2Graph [19] determined the weights of edges between nodes by calculating the distribution probabilities between nodes to detect anomalies in time series data. Brain-Net [20] calculated cosine similarity between nodes and set a threshold to establish edges between nodes. GraphSleepNet [21] constructed the sleep stage network using an adjacency matrix learned from EEG signal sequences, and the sleep feature matrix was filled by differential entropy features at each frequency band. In the second type, GDN [22] computed the normalized dot product between the embedding vectors of sensors to learn the relationships between the sensors according to the realistic positions of the sensors. MST-GAT [23] used time series embedding to construct a flexible graph structure and computed cosine similarity to build an adjacency matrix to adequately represent the time series' inherent properties. In this paper, different from the prior work, we investigate the capability of graph construction to tackle the problem of handling variations in ECG signals effectively. Our graph structure preserves specific time-interval information and captures sequential correlations within the signals, thereby enabling its potential application in ECG disease diagnosis.

B. ECG Representation Learning

Many artificial intelligence-based methods have been developed to enhance the representations of ECG signals, improving both the accuracy and efficiency of detecting cardiac abnormalities [24]–[26]. ECG representation learning methods primarily fall into two categories: supervised learning based on class labels and unsupervised learning. In supervised representation learning, models focus on capturing the spatial and temporal features of ECG data by extracting patterns associated with class labels. For instance, RCNNs [27]–[30] developed a hybrid model combining CNN and RNN modules to learn such features, enabling accurate classification of one-lead fixed-length ECG beats. Similarly, 1D-CNN [31]–[34] was adapted for arrhythmia classification using long-duration ECG signals (e.g., ten seconds). An improved Residual Network [35], [36] was applied to standard short-duration 12-lead ECG data, outperforming cardiology residents in detecting six types of abnormalities. Zubair *et al.* [37] employed a re-sampling strategy with a translation loss function to enhance focus on relevant information for arrhythmic beat detection.

However, an overemphasis on class label information can fall short of providing a comprehensive characterization of ECG signals, leading to insufficient representations. In contrast, unsupervised representation learning seeks to address this issue by capturing the intrinsic nature of ECG signals. Nakamoto *et al.* [38] and Rabbani *et al.* [39] employed contrastive learning to bring ECG signals closer to their augmented samples in the feature space while pushing them further away from the other ECG signals and their augmented samples. Grabowski *et al.* [40], Rodriguez *et al.* [41], and Gedon *et al.* [42] applied random masking to ECG signals and trained models to predict the masked portions, extracting features for downstream tasks. Chen *et al.* [43] and Hu *et*

al. [44] proposed methods to integrate disease information into the generative process, yielding ECG signal representations enriched with disease-specific features. While previous methods often captured fundamental signal patterns, they still struggled with detecting risk signals in imbalanced ECG data, where crucial diagnostics may rely on a few critical waveforms. This is because standard augmentations in unsupervised learning, such as random dropping and noise addition, often miss small but significant patterns, making them harder to detect risk signals.

III. METHOD

A. Problem Definition

We first describe the notation and definition of the problem in this study. Suppose an ECG dataset $S = [S_1, S_2, \dots, S_r]$ is given (S contains r samples). Each sample $S_i = [s_1, s_2, \dots, s_n] \in \mathbb{R}^{n \times t}$ is recorded from n leads, with t timestamps per lead. We divide each lead of S_i into a sequence of time intervals, $s_l = [x_1, x_2, \dots, x_{\frac{t}{m}}]$, i.e., s_l is represented as $s_l \in \mathbb{R}^{\frac{t}{m} \times m}$, with m denoting the preset length of the time interval.

Arrhythmia Classification. Given the ECG dataset S and the corresponding label set $Y = [y_1, y_2, \dots, y_r]$, where $y_v \in [0, 1]$ is the value of the v -th label. The objective of arrhythmia classification is to predict the labels $\hat{Y}$ of an ECG test set S_{test} :

$$\hat{Y} = M(Y, R(S), R(S_{test})), \quad (1)$$

where R is a representation learning model used to extract features from ECG signals, and M is a linear classifier.

B. General Description

To develop a unified representation of ECG data and achieve versatile and risk-sensitive cardiac diagnosis, we propose a novel unified ECG representation approach, VARS. Figure 2 illustrates the structure of the VARS model, mainly comprising three primary modules: the graph construction module, the denoising reconstruction module, and the feature subgraph contrastive learning module. First, to address the issue of non-uniform ECG data formats, VARS transforms the heterogeneous ECG data with different leads into graph data. In the graph construction module, we use a specific time interval to define a graph node, and the signal patterns between nodes are computed with a self-attention mechanism. Second, to enhance risk-sensitive cardiac diagnosis, we propose a unified representation contrastive learning method. Specifically, we employ a GNN decoder to reconstruct masked graphs, aiming to preserve the signals' purity and usability while simultaneously filtering out irrelevant noise. On the other hand, by leveraging the feature subgraph contrastive learning module, we extract subgraphs to obtain sample pairs to enhance discrimination among signals in contrastive learning. Finally, the trained GNN encoder is employed for downstream tasks such as classification and interpretation of graph data.

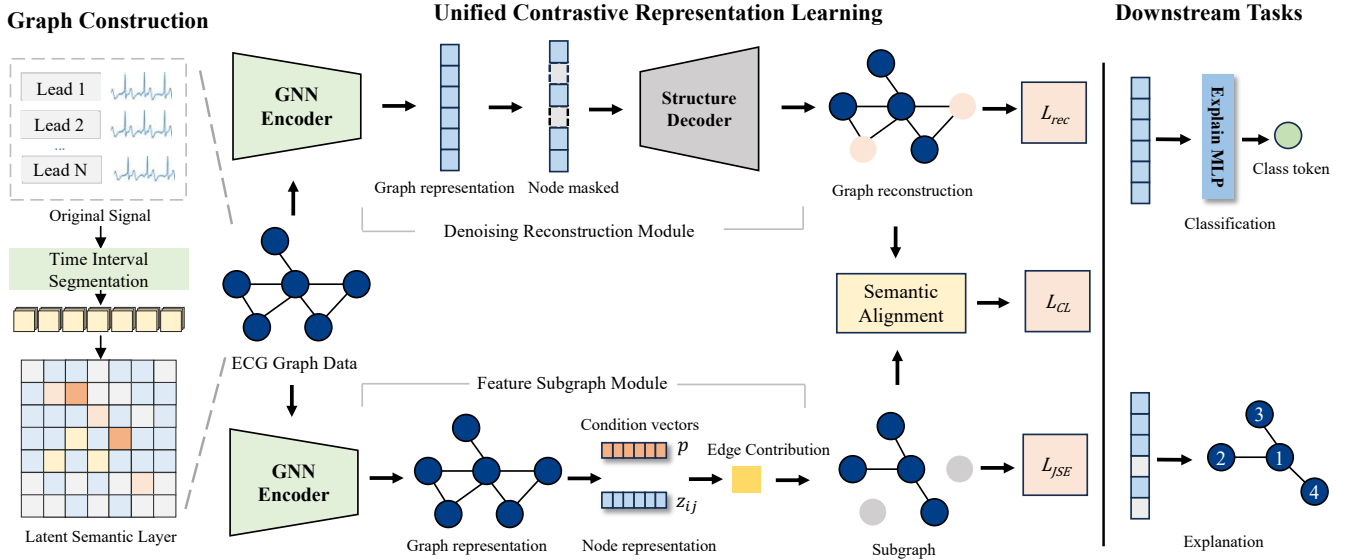


Fig. 2. **An overview of our VARS framework.** VARS consists of three main parts. The graph construction process transforms ECG data into graph data, using attention mechanisms to capture semantic relationships between graph nodes. Feature subgraph contrastive learning is employed to generate a unified ECG representation. Downstream tasks utilize the trained GNN encoder for classification and interpretation of graph data.

C. Graph Construction

To address the challenge of processing ECG signals with diverse configurations, we develop a method to convert ECG signals into a universal graph structure. By establishing nodes and edges, the graph structures allow us to preserve essential temporal interval details and accurately capture sequential relationships within the data. Moreover, considering the potential complex dependencies and diagnostic patterns among different time intervals in ECG signals, we introduce a self-attention mechanism to automatically learn these patterns. The self-attention mechanism enables the model to flexibly capture long-range dependencies in sequential data, hence more accurately uncovering significant patterns within ECG signals. In this method, we compute attention scores to enable the mechanism to focus on the most relevant aspects of the signals. Specifically, we apply multi-head self-attention to the set of nodes representing the time intervals. Each attention head computes its own attention score matrix, capturing different aspects of the inter-node relationships. For each attention head e , we compute the attention score matrix W_e using the query Q_e and key K_e matrices, as follows:

$$W_e = \text{Sigmoid} \left(\frac{Q_e K_e^T}{\sqrt{d_k}} \right), \quad (2)$$

where d_k is the hidden dimension size of Q_e and K_e . Finally, we take the mean of the attention score matrices, $\bar{W}$, which is used to compute our adjacency matrix A .

To construct the adjacency matrix A , we apply a threshold Θ to $\bar{W}$, converting scores below the threshold to zeros:

$$A_{ij} = \begin{cases} \bar{W}_{ij}, & \text{if } \bar{W}_{ij} \geq \Theta, \\ 0, & \text{otherwise,} \end{cases} \quad (3)$$

where A_{ij} is the element of the adjacency matrix A for the edge between nodes i and j . We use the threshold Θ to filter

out weaker connections, hence ensuring that only significant relationships are represented in the graph and yielding a sparser graph that retains only edges indicative of strong semantic correlations between nodes. This process not only simplifies the graph but also enhances the focus on semantic information.

We construct a graph $G = (\mathcal{V}, A, X)$, where $\mathcal{V}$ denotes the nodes corresponding to the time intervals, the adjacency matrix $A \in \{0, 1\}^{N \times N}$ represents edges between the nodes, $X \in R^{N \times d}$ is the feature matrix of the nodes, and $N = |\mathcal{V}|$.

D. Unified Contrastive Representation Learning

In traditional contrastive learning, graph enhancement often involves randomly masking nodes, which can potentially alter the original attributes and may lead to the creation of anomalous sample pairs. To address this issue, we propose Unified Contrastive Representation Learning, which preserves the original attributes of the data by removing irrelevant noise from normal nodes while highlighting risk signals (i.e., removing noise while maintaining the original graph structure). Initially, we perform masking operations on ECG graph data, followed by weak enhancement through the reconstruction of ECG data using the GNN decoder. Subsequently, we employ the feature subgraph contrastive learning module to strongly enhance the data, obtaining enhanced subgraphs while ignoring irrelevant noise, reinforcing signal differentiation, and amplifying the separation between samples of different categories in the feature space, thus ensuring the acquisition of more representative sample pairs.

Denoising Reconstruction Module: To preserve the original ECG information and mitigate the impact of noise during contrastive learning, we employ a GNN encoder and GNN decoder to reconstruct the structure of the ECG graph. By reconstructing the masked nodes, we produce a perturbed

graph G' that retains the semantic information of the original signals. In the GNN encoder stage, consider the graph $G = (\mathcal{V}, A, X)$ with $N = |\mathcal{V}|$. We seek to reconstruct the features and structure, as:

$$H = f_E(\mathcal{V}, A, X), \quad (4)$$

$$G' = f_D(A, H), \quad (5)$$

where f_E is a graph encoder function that encodes the features of the nodes to obtain the hidden feature H , and f_D is a graph decoder function that reconstructs from H a graph G' , so that G' well approximates the structure and features of the original graph G , ensuring essential information being preserved while reducing noise introduced during contrastive learning.

GNN-based Structure Decoder: In general, a decoder reconstructs the original input from the latent representation generated by the corresponding encoder. In traditional natural language processing (NLP), decoders aim to recover masked words with rich semantic information, making conventional decoders like Multi-Layer Perceptrons (MLPs) highly effective. However, in ECG analysis, the subtle differences in shape and relationships between various cardiac beat structures often lead to nearly identical features when using low-level decoders. Therefore, VARS employs a Graph Isomorphism Network (GIN) [45] as its decoder to discriminate more effectively among different beat structures. The decoder correlates the nodes' and neighbors' features to reconstruct nodes, rather than solely relying on node-specific features, thus enabling the encoder to learn higher-level latent semantic representations. We define $\tilde{\mathcal{V}} \subset \mathcal{V}$ as the subset of nodes selected for masking during training. We replace the features of these masked nodes with a learnable vector denoted as $h_{[M]}$. The masking method is as follows:

$$\tilde{h}_i = \begin{cases} h_{[M]}, & \text{if } v_i \in \tilde{\mathcal{V}}, \\ h_i, & \text{if } v_i \notin \tilde{\mathcal{V}}, \end{cases} \quad (6)$$

where $\tilde{h}_i$ is the hidden feature of node v_i after the hidden coding masking, h_i represents the feature vector of node v_i before masking is applied, and $h_{[M]}$ represents the hidden coding of the nodes in the masked subset $\tilde{\mathcal{V}}$.

Feature Subgraph Module: To address the limitations of classic contrastive learning in capturing subtle key information in ECG data, we use the embedded explainable module [46] as part of our feature subgraph module to enhance paired samples for contrastive learning. The feature subgraph module can discard unimportant edges while retaining important node information, preventing the loss of node attributes during the graph enhancement process. Simultaneously, the feature subgraph module can also interpret the classification results in downstream tasks. The formulas used by the feature subgraph module are:

$$G_{\text{sub}} = (\mathcal{V}, E_{\text{sub}}) = \mathcal{T}_\theta(p, G), \quad (7)$$

$$E_{\text{sub}} = \{(v_i, v_j) \mid (v_i, v_j) \in E \text{ and } w_{ij} \geq \delta\}, \quad (8)$$

where $\mathcal{V}$ is the node set of G , the model $\mathcal{T}_\theta$ determines whether to keep an edge by calculating the importance of the edge, θ denotes the parameters of the model $\mathcal{T}_\theta$, and p is a

conditional vector following the Laplace distribution which is used to highlight specific node features. The edge set E_{sub} is determined by applying a thresholding operation with a threshold δ to the edge contributions of the edge set E of G . For an edge $(v_i, v_j) \in E$, the edge contribution w_{ij} is calculated based on the features of nodes v_i and v_j in G .

When conducting subgraph extraction, the feature subgraph module takes the node representations z_i and z_j of nodes v_i and v_j and the conditional vector p as input, and calculates the score w_{ij} for an edge $(v_i, v_j) \in E$, as:

$$w_{ij} = \text{MLP}([z_i; z_j] \otimes \sigma(f_g(P))), \quad (9)$$

where $[\cdot; \cdot]$ denotes concatenation along the feature dimension, σ is the Sigmoid function, f_g is a linear projection used to match the dimensionality of the data, and P is the concatenation of the result of the classification prediction and the gradient in the inference stage of downstream tasks, which provides task-specific information. This process enables the feature subgraph module to learn edge scores related to both the graph structure and the specific requirements of the task, making the subgraph representation of ECG data more effective for downstream tasks such as classification and interpretation.

E. Loss Functions

Finally, to achieve the goal of discriminating between risk and non-risk patterns within diverse and complex ECG signals, we introduce a novel combination of three distinct loss functions, which are tailored to the denoising reconstruction, feature subgraph module, and unified contrastive representation.

In the denoising node reconstruction module, although ECG signals share similar structural types, their overall morphologies may vary significantly. This variability poses challenges to traditional reconstruction methods, in which Mean Squared Error (MSE) often struggles to capture the nuances of such diverse patterns due to its focus on minimizing differences of fixed values. To better capture these broader ECG patterns, we instead employ cosine error, which emphasizes the similarity in shape and overall morphology, rather than relying solely on precise value matching.

To better align the cosine similarity with the characteristics of the ECG signal reconstruction task, we scale the cosine error using a parameter $\gamma \geq 1$. Since predictions with high confidence typically result in cosine errors less than 1, this scaling reduces the weights of such high-confidence predictions in the overall loss, thus shifting the focus toward lower-confidence predictions. For a node $v_i \in \tilde{\mathcal{V}}$, given the original node feature x_i and the reconstructed output $\tilde{x}_i$ in $G' = f_D(A, H)$, the scaled cosine error is computed as:

$$L_{\text{rec}} = \frac{1}{|\tilde{\mathcal{V}}|} \sum_{v_i \in \tilde{\mathcal{V}}} \left(1 - \frac{x_i \tilde{x}_i^T}{|x_i| \cdot |\tilde{x}_i|}\right)^\gamma, \gamma \geq 1, \quad (10)$$

where the scale factor γ is a tunable hyperparameter that can be optimized for ECG datasets of varying sizes. This approach applies a global scaling to the reconstruction error, adjusting

the sensitivity of the loss function and improving the model's generalization across the entire dataset.

When calculating the subgraph loss in the feature subgraph module, we use the Jensen-Shannon Estimator (JSE) to compute the loss for subgraph generation. The corresponding formula is:

$$L_{JSE} = \min_{\theta} \left\{ \frac{1}{L} \sum_{i=1}^L \log \left[\sigma \left((p \otimes H_i) (p \otimes H_{i,\theta})^T \right) \right] \right. \\ \left. + \frac{1}{L^2 - L} \sum_{i \neq j} \log \left[1 - \sigma \left((p \otimes H_i) (p \otimes H_{j,\theta})^T \right) \right] \right\} \quad (11)$$

where L is the number of samples in a batch, σ is the Sigmoid function, and H_i and $H_{i,\theta}$ are the embeddings of the original graph G_i (which is the i -th graph in a batch) and its subgraph $T_{\theta}(p, G_i)$ respectively.

After the ECG graph data are processed by the denoising reconstruction module and the feature subgraph module, two augmented graphs are generated that retain the pathognomonic patterns of the ECG signals. These two augmented graphs are paired as positive samples, while contrastive learning is enforced by maximizing their distinction from negative sample pairs. To ensure consistency, we use the normalized temperature-scaled cross-entropy loss (NT-Xent) [47] to calculate the contrastive loss, which is given as:

$$L_{CL} = -\frac{1}{L} \sum_{i=1}^L \log \frac{\exp \left(\text{sim} \left(Z_i^{(r)}, Z_i^{(s)} \right) / \tau \right)}{\sum_{k=1}^{2L} \mathbf{1}_{k \neq i} \exp \left(\text{sim} \left(Z_i^{(r)}, Z_k \right) / \tau \right)}, \quad (12)$$

where τ is the temperature parameter, $Z_i^{(r)}$ and $Z_i^{(s)}$ denote the two augmented representations of the i -th graph in the batch (i.e., the 'r' = 'reconstruction' and 's' = 'subgraph' versions) respectively, and Z_k denotes the representation of the k -th graph in a set consisting of $2L$ representations, with L of $Z_i^{(r)}$'s and L of $Z_i^{(s)}$'s. The function $\text{sim}(\cdot, \cdot)$ calculates the cosine similarity between two representations, and the indicator function $\mathbf{1}_{k \neq i}$ ensures that the similarity of a representation with itself is excluded from the denominator.

Finally, the overall loss function of our VARS is:

$$L_{ALL} = \lambda_{\text{rec}} L_{\text{rec}} + \lambda_{\text{JSE}} L_{\text{JSE}} + \lambda_{\text{CL}} L_{\text{CL}}. \quad (13)$$

where λ_{rec} , λ_{JSE} , λ_{CL} are scalar weights for the reconstruction loss L_{rec} , subgraph loss L_{JSE} , and contrastive loss L_{CL} , respectively. Each loss plays a complementary role: L_{rec} preserves structural integrity under masking and denoising, L_{JSE} guides the interpreter to extract concise, task-aligned subgraphs, and L_{CL} sharpens class-discriminative geometry of the learned representations. To avoid introducing additional tuning, we fix $\lambda_{\text{rec}} = \lambda_{\text{JSE}} = \lambda_{\text{CL}} = 1$ throughout, and Appendix A demonstrates that this equal-weight setting is effective in our experiments.

IV. EXPERIMENTS

A. Experimental Setup

To evaluate model performance, we conduct experiments on four publicly available ECG datasets—MITBIH, PTB-XL, ST-T, and the contemporary 12-lead Chapman-Shaoxing dataset—covering diverse lead configurations, sampling rates,

TABLE I
STATISTICS OF THE ECG DATASETS USED IN THIS STUDY.

Dataset	Samples	Leads	Signal length	Frequency
MITBIH	100687	2	9–60 s	360 Hz
PTB-XL	21799	12	10 s	500 Hz
ST-T	791489	2	2 h	250 Hz
Chapman-Shaoxing	45152	12	10 s	500 Hz

and diagnostic targets. Using the criteria given by the American Association for the Advancement of Medical Instrumentation (AAMI), we divide ECG signals into five groups: non-ectopic (N), supraventricular (S), ventricular (V), fused (F), and unknown (Q). For the Chapman-Shaoxing dataset, the original label space comprises numerous expert diagnostic statements. The distribution is markedly long-tailed, with many labels appearing in only a single or a few records. To obtain reliable estimates and emphasize clinically risk-relevant abnormalities, we aggregate diagnostic statements into five multi-label disease superclasses—*Atrial Abnormalities*, *Junctional Arrhythmias*, *Ventricular Arrhythmias*, *Conduction Abnormalities*, and *Myocardial Ischemia/Infarction*—so that each recording may belong to multiple superclasses. Accordingly, on Chapman-Shaoxing we report results under the *Anomaly Risk Categories* setting. The dataset statistics are summarized in Table I.

MITBIH: The MITBIH dataset [48] contains 48 half-hour two-channel ECG recordings sampled at 360 Hz. Four ECG records of pacemaker users (numbers 102, 104, 107, and 217) were deleted following the standards proposed by AAMI. As in the previous work [49]–[51], we use the single lead and modified limb lead II (MLII) as the input of the baselines.

PTB-XL: The PTB-XL dataset [52] contains 21,837 12-lead ECG records sampled at 500 Hz for exactly 10 seconds in duration. We pre-process this dataset following the AAMI criteria, resulting in 21,799 ECG recordings with 12 leads, partitioned by PTB-XL superclass labels.

ST-T: The European ST-T dataset [53] contains 90 excerpts of annotated Holter recordings from 79 subjects. Each recording has a duration of two hours and consists of two signals, and each is sampled at 250 samples per second. We divide the two hours of recordings by heartbeat labels.

Chapman-Shaoxing: The Chapman-Shaoxing dataset [54] contains 45,152 10-second, 12-lead ECGs at 500 Hz, labeled by licensed physicians with diagnostic statements. Labels are mapped to Systematized Nomenclature of Medicine—Clinical Terms (SNOMED CT) for multi-label classification, and records are provided in WaveForm Database (WFDB) with per-record metadata, offering a large and contemporary dataset for arrhythmia modeling.

B. Implementation Details

We randomly allocate 70% of the ECG signals to the training set and retain the remaining 30% for the test set, ensuring robust evaluation of the model's performance on unseen data. Hyperparameters are heuristically set with a threshold parameter Θ filtering out 75% of graph connections, $\delta = 0.80$ to control edge retention, a masking rate $\rho =$

TABLE II

PERFORMANCE COMPARISON OF CLASSIFICATION IN THE OVERALL CATEGORIES ON THE THREE ECG DATASETS. THE BEST AND SECOND-BEST RESULTS IN EACH COLUMN ARE MARKED IN **bold** AND UNDERLINED, RESPECTIVELY. ‘–’ INDICATES THAT WE CANNOT FIND OR REPRODUCE THE RESULTS DUE TO PRIVATE IMPLEMENTATION OF THE ORIGINAL PAPERS OR INAPPLICABLE SETTINGS.

Method	MITBIH				ST-T				PTB-XL			
	ACC	F1	Sen	Spe	ACC	F1	Sen	Spe	ACC	F1	Sen	Spe
GAT	0.9507	0.5667	0.5003	0.6538	0.9910	0.1991	0.2000	0.4989	0.6168	0.4472	0.4466	0.5756
GIN	0.9003	0.2663	0.2490	0.5355	0.9910	0.1991	0.2000	0.4990	0.4624	0.1265	0.2000	0.4355
SAGPool	0.9777	0.7236	0.6888	0.7050	0.9914	0.1991	0.2000	0.4990	0.6809	0.5643	0.5562	0.6171
MINIROCKET	<u>0.9913</u>	0.6601	0.7489	0.9274	0.9967	0.7264	0.5741	0.9274	0.7217	0.6179	0.6048	0.9201
ResNet50	0.9909	0.7629	0.7589	0.9886	0.9971	0.6757	0.6528	0.9568	0.6794	0.5631	0.5510	0.9085
LSTM-FCN	0.9684	0.7604	0.7524	0.7187	0.9790	0.2130	0.2139	0.8182	0.7289	0.6250	0.6016	0.9221
TCN	0.9008	0.2835	0.2549	0.5685	0.9913	0.1991	0.2000	0.4990	0.4635	0.2369	0.2524	0.4851
RCNNs	0.9846	0.7196	0.7233	0.9873	–	–	–	–	–	–	–	–
CascadeCNN	0.9907	<u>0.7665</u>	0.7548	0.9874	<u>0.9972</u>	0.7479	0.7029	0.9520	0.6697	0.5752	0.5803	0.9096
Transformer	0.9912	0.7633	<u>0.7600</u>	<u>0.9895</u>	0.9965	0.6333	0.6147	<u>0.9575</u>	0.7106	0.6028	0.5959	0.9206
BMIRC	0.9849	0.6743	0.7533	0.9891	0.9922	0.6450	0.6401	0.9015	0.7255	0.6231	<u>0.6102</u>	0.8942
WResHDual	0.9901	0.7622	0.7505	0.9875	0.9968	0.7403	0.6955	0.9490	0.7210	0.6190	0.5983	0.9195
TCL	0.9812	0.7280	0.7084	0.9736	0.9951	0.3674	0.3536	0.8990	0.4546	0.1250	0.2000	0.4214
CPC	0.5124	0.1651	0.1982	0.7989	0.8622	0.1874	0.1992	0.7993	0.4546	0.1250	0.2000	0.4214
VARS (ours)	0.9916	0.7712	0.7656	0.9910	0.9975	0.7719	0.7042	0.9592	0.7313	0.6400	0.6284	0.9248

TABLE III

PERFORMANCE COMPARISON OF CLASSIFICATION IN THE ANOMALY RISK CATEGORIES. THE BEST AND SECOND-BEST RESULTS IN EACH COLUMN ARE MARKED IN **bold** AND UNDERLINED, RESPECTIVELY. ‘–’ INDICATES THAT WE CANNOT FIND OR REPRODUCE THE RESULTS DUE TO PRIVATE IMPLEMENTATION OF THE ORIGINAL PAPERS OR INAPPLICABLE SETTINGS.

Method	MITBIH			ST-T			PTB-XL		
	ACC	F1	Sen	ACC	F1	Sen	ACC	F1	Sen
MINIROCKET	0.9230	<u>0.5702</u>	0.6863	0.6339	0.4605	0.4676	0.5940	0.4350	0.5325
LSTM-FCN	0.8441	0.4867	0.5792	0.0474	0.0100	0.0397	<u>0.6140</u>	<u>0.4475</u>	<u>0.5564</u>
CascadeCNN	0.9397	0.5653	0.6931	<u>0.8340</u>	0.4705	<u>0.6145</u>	0.5404	0.4142	0.5193
Transformer	<u>0.9434</u>	0.5642	<u>0.7008</u>	0.7552	0.4336	0.5187	0.5884	0.4289	0.5248
BMIRC	0.9385	0.5670	0.6955	0.8310	<u>0.4725</u>	0.6120	0.6110	0.4450	0.5530
WResHDual	0.9355	0.5633	0.6910	0.8290	0.4690	0.6080	0.6085	0.4421	0.5501
TCL	0.8682	0.5300	0.6369	0.4851	0.1679	0.1920	–	–	–
CPC	0.1025	0.0278	0.0233	0.0017	0.0021	0.0015	–	–	–
VARS (ours)	0.9554	0.5721	0.7080	0.8472	0.6241	0.6896	0.6179	0.4712	0.5688

70%, and a scaling factor γ of 2 to balance reconstruction errors and enhance generalization. The default batch size is set to 1024, and model training is conducted on an NVIDIA GeForce RTX 4090. In the cases where GPU memory limits are encountered, we appropriately reduce the batch size for training purposes. In the downstream tasks, we use an MLP to classify the ECG representation. For the interpretation tasks, we directly utilize the trained Feature Subgraph Module to perform subgraph extraction and provide interpretability. We use several unified metrics to assess the model performance, including accuracy, F1 score, specificity, and sensitivity. These metrics are computed using the macro method to identify the advantages and areas for improvement in cardiac arrhythmia detection and heartbeat classification methods.

C. Comparison with Baseline Methods

We evaluate the effectiveness of our proposed VARS and compare with 12 ECG classification baselines, including GNN-based models (GAT [55], GIN [45], and SAGPool [56]), CNN-based models (MINIROCKET [57], LSTM-FCN [58], ResNet50 [59], TCN [60], RCNNs [27], CascadeCNN [61], and Transformer [62]), and self-supervised learning models (TCL [63] and CPC [64]). To provide a contemporary compar-

ison in ECG-specific modeling, we additionally include two recent, ECG-tailored baselines: a multimodal fusion method that combines spatiotemporal and frequency representations (WResHDual) [65], and a bimodal masked autoencoder with internal representation connections (BMIRC) [66]. Since the baselines have various input dimension requirements, we re-sample the datasets in order to meet the input requirements of each model. We experiment with classification performance in both the overall categories and risk categories. The results show that our VARS model not only adapts to the diverse nature of ECG data but also excels in identifying the nuanced variations of risk categories, demonstrating the robustness and adaptability of our new method.

Classification Performance in the Overall Categories: As shown in Table II, **in the realm of GNN classification models**, GNNs exhibit commendable performance on the relatively simple MITBIH dataset. However, their effectiveness diminishes on the significantly imbalanced ST-T dataset, on which the message-passing technique struggles to differentiate among fewer ECG signal categories. **CNN-based models** demonstrate strong performance on the datasets aligned with their input structure but struggle with variability in ECG data, as they are unable to maintain high performance across different datasets.

RCNNs classify arrhythmias by transforming ECG signals into image data, and consequently face scalability issues when applied to large-scale ECG datasets. Due to the significant sizes and complexity of the ST-T and PTB-XL datasets, training the model on them is not feasible with such an approach. Therefore, results in certain metrics cannot be computed, and we use ‘—’ as a placeholder to indicate the missing values in Table II. **In self-supervised learning models**, it is observable from Table II that TCL achieves better performance on short ECG sequences but fails to represent effectively on the longer sequence PTB-XL dataset. CPC shows poor classification performance on the three ECG datasets. This can be attributed to the periodicity inherent in ECG signals, which poses a challenge to the intrinsic predictive feature learning method employed by CPC. Consequently, this approach struggles to effectively capture and represent key discriminative features necessary for accurate classification of ECG signals. From Table II, it is evident that no single known method dominates as the runner-up (to our VARS) across the different datasets. In comparison, our VARS model captures a unified representation of ECG signals across these datasets, avoiding classification performance degradation typically caused by dataset format variations. This demonstrates the robustness and adaptability of our model in achieving high classification accuracy, regardless of the dataset complexities.

Classification Performance in the Risk Categories: Due to the predominance of the normal categories in general ECG data, the results obtained for the overall classification performance can obscure the models’ ability to classify abnormal categories. To further accurately evaluate the models’ performance in the anomaly risk categories, we compare our VARS with those baselines yielding second-best performance and the self-supervised learning baselines. Table III shows the results. It is important to note that we omit the results of specificity metrics because our focus is on the performance of the abnormal categories. On the PTB-XL dataset, TCL and CPC cannot effectively distinguish between normal and abnormal samples, and hence we use ‘—’ to indicate this. The results reveal significant performance disparities between normal and abnormal categories across the baseline models, as demonstrated in Tables II and III. VARS consistently excels in classifying abnormal samples, demonstrating its ability to accurately identify abnormal patterns in ECG data. This superior performance underscores the effectiveness and reliability of VARS in detecting critical cardiac abnormalities.

Classification Performance on the Multi-Label Dataset: As summarized in Table IV, VARS attains the strongest accuracy, specificity, and AUC on Chapman-Shaoxing, while remaining competitive on F1 and precision. Relative to WResH-Dual, VARS accepts a slight trade-off in F1 and precision for clearer gains in separability and in true-negative control, yielding an operating profile that is robust to threshold selection—particularly desirable for multi-label clinical screening. CNN baselines such as ResNet and Transformer show strong specificity but lower separability and overall accuracy and F1 compared with VARS. GNN baselines (GAT, GIN, SAGPool) group at a lower performance tier across metrics, and methods like MINIROCKET and BMIRC underperform broadly. Over-

all, Table IV indicates that VARS combines high separability and specificity with competitive positive-call performance, offering a balanced choice for contemporary multi-label ECG analysis.

TABLE IV
DETAILED PERFORMANCE ON THE CHAPMAN-SHAOXING DATASET.

Method	Chapman-Shaoxing				
	ACC	F1	Precision	Spe	AUC
GAT	0.574	0.570	0.520	0.908	0.852
GIN	0.597	0.573	0.540	0.921	0.857
SAGPool	0.583	0.570	0.530	0.911	0.853
MINIROCKET	0.465	0.460	0.400	0.844	0.751
ResNet50	0.640	0.682	0.630	0.939	0.930
LSTM-FCN	0.597	0.584	0.530	0.916	0.857
CascadeCNN	0.632	0.657	0.621	0.923	0.901
Transformer	0.641	0.622	0.590	0.939	0.891
BMIRC	0.453	0.471	0.430	0.853	0.773
WResHDual	0.661	0.730	0.710	0.923	0.926
VARS (ours)	0.669	0.728	0.690	0.945	0.939

D. Model Interpretability

The interpretability of the VARS model enables it to provide meaningful explanations for its classification decisions. Traditional interpretability often offers only a characteristic explanation of the ECG signal, which can be perplexing. To address this issue, we implement an interpretable method that integrates both qualitative and quantitative approaches, aligning with experts’ practice perspectives.

Qualitative Analysis: We conduct granular model interpretability studies using data from the MITBIH and PTB-XL datasets. The results are presented in Figure 3. We investigate two different granularity levels of interpretability: the heartbeat structure level and heartbeat level. VARS allows medical professionals to make ECG-interpretable fine-grained choices based on their diagnostic needs to ensure that they can quickly locate abnormal ECG segments for ECG-aided diagnosis. In Figure 3, we see various labels such as N, S, F, and Q, which correspond to normal heartbeats, supraventricular premature beats, atrial fibrillation, and unknown rhythms, respectively. Their specific manifestations in the ECG data are clearly illustrated. In the heartbeat structure level, as shown in the top portion of Figure 3, the leftmost part illustrates a schematic of a normal heartbeat category, while the other three parts display schematics of ECG heartbeats for three different disease types. VARS can swiftly delineate the structure of an abnormal ECG segment to assist doctors in rapid diagnosis. We further consulted medical experts, who confirmed that the delineated abnormal regions are consistent with clinical diagnostic standards. Within the scope of the heartbeat level, the bottom portion of Figure 3 offers an interpretable analysis within the continuous ECG signal. The left part gives a normal ECG signal, and the right part illustrates the range of delineated abnormality classes. Doctors can select one or more heartbeats to allow the model to determine the range, thereby facilitating rapid diagnosis.

Quantitative Analysis: We quantify interpretability using model-internal signals produced by the Feature Subgraph

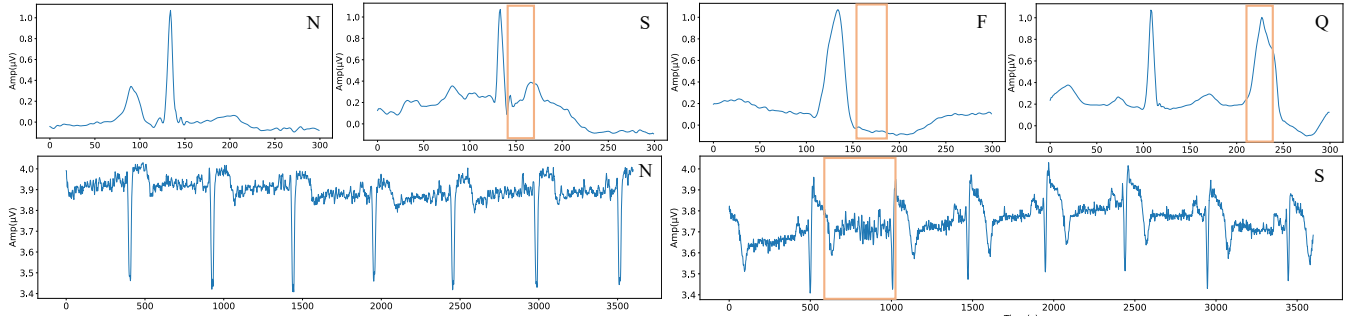


Fig. 3. **Visualization of ECG interpretability.** The top portion is an interpretable demonstration at the heartbeat structure level, and the bottom portion is an interpretable demonstration at the heartbeat level.

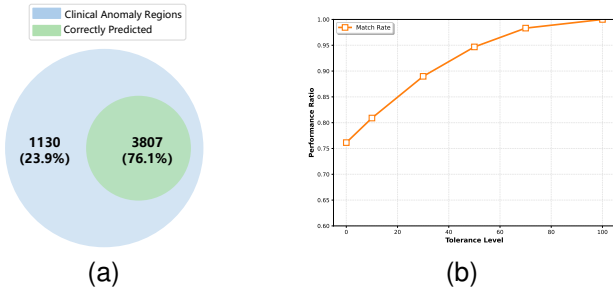


Fig. 4. **Match between clinical annotations and interpreter outputs.** (a) Venn diagram on the 5,000-record set showing where the interpreter’s top-1 segment overlaps the clinically annotated interval. (b) Match-rate curve as the window tolerance increases, showing a smooth rise that indicates stable agreement between the interpreter’s top-1 segment and the clinical markings.

Module. For each sample, the module yields node- and edge-level importance on the ECG graph and highlights the time-interval segments that constitute the explanatory subgraph. We report compact summaries of these outputs via per-sample dashboards to complement the qualitative analyses in Figure 3. Full visualization layouts and the machine-readable summary schema are provided in Appendix B.

To assess whether the interpreter’s evidence aligns with clinical reading, we curated a set of 5,000 ECG records and had clinical annotate a single diagnostically salient time-interval per record. Our interpreter then returns a *top-1* segment (the highest-importance patch) per record. We count a prediction as a match when the interpreter’s top-1 segment overlaps the reference interval derived under standard clinical interpretation principles. Figure 4(a) summarizes the overlap between the clinical reference annotations and the interpreter’s top-1 segments, while Figure 4(b) shows how the match rate varies as the tolerance window increases. Together, these plots indicate that the interpreter’s most salient window aligns well with clinical markings and that the match rate increases smoothly as the tolerance broadens. We also provide a full 12-lead visualization where clinician-marked segments and the interpreter’s top-1 segments. The panel shows both normal and abnormal examples and reports the segment indices and time ranges for clarity. The full figure is included in Appendix B.

E. Hyperparameter Sensitivity and Computational Efficiency

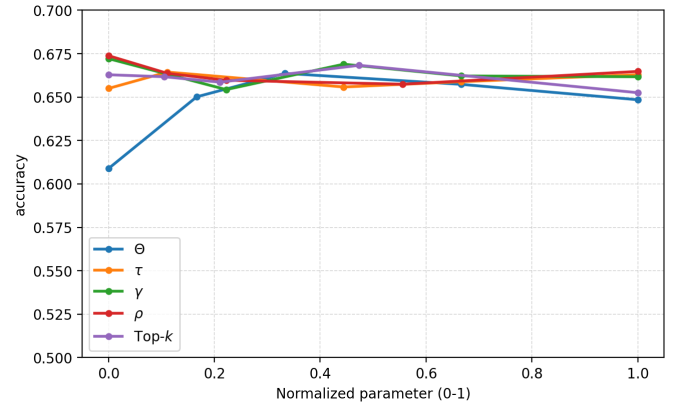


Fig. 5. Parameter sensitivity of VARS over five hyperparameters: Θ , Top- k , masking rate ρ , γ , and τ . Each curve uses a five-point grid centered at our default (the *third* tick), with two symmetric values on each side.

To assess robustness, we conduct a one-factor-at-a-time sensitivity study on five hyperparameters—graph threshold Θ , Top- k , masking rate ρ , reconstruction scaling γ , and contrastive temperature τ . For each hyperparameter, we evaluate five values by centering the grid at our model’s setting (the third point, median of the grid) and expanding symmetrically to both sides. The results show stable performance over broad ranges in Figure 5. The sensitivity curves remain stable around the default setting, and the observed variations across the five ablations are modest, indicating that the default configuration lies on a broad performance plateau and serves as a reliable choice without heavy per-dataset retuning.

In addition, we report the computational profile (parameter, FLOPs and latency). As summarized in Table V, VARS maintains a moderate parameter count among ECG-specific classifiers and achieves low FLOPs with competitive latency, and while smaller than large ECG models, it still delivers faithful, clinically meaningful explanations. Notably, although VARS is slightly larger than lightweight CNNs, it keeps inference times short enough for real-time use and sustains stable performance across heterogeneous settings.

F. Ablation Studies

We validate the efficacy of each module within our model through ablation experiments. Specifically, we remove the following components to examine their individual impact on

TABLE V

COMPUTATIONAL COMPLEXITY SUMMARY. METRICS REPORTED: PARAMETERS (MILLIONS), FLOPS (BILLIONS), AND END-TO-END LATENCY (MS PER SAMPLE) ON AN RTX 4090 (LOWER IS BETTER). BEST IN BOLD; SECOND BEST UNDERLINED.

Model	Model Size (M)	FLOPs (G)	Latency (ms)
BMIRC	19.38	<u>7.6</u>	5.25
WResHDual	33.08	16.53	13.11
VARS (ours)	<u>23.39</u>	5.15	<u>6.89</u>

TABLE VI

RESULTS OF ABLATION STUDY ON THREE DIFFERENT MODULES AND THREE LOSS FUNCTIONS OF OUR MODEL.

Dataset	Module	ACC	F1	Sen	Spe
MITBIH	w/o G-S	0.9828	0.7400	0.7290	0.9805
	w/o D-R	0.9775	0.7184	0.7061	0.9755
	w/o G-C	0.9818	0.7269	0.7059	0.9803
	w/o L_{rec}	0.9757	0.7096	0.6807	0.9672
	w/o L_{JSE}	0.9758	0.7102	0.6806	0.9663
	w/o L_{CIL}	0.9761	0.7113	0.6788	0.9659
	All	0.9916	0.7712	0.7656	0.9910

performance: the graph structure (G-S), denoising reconstruction module (D-R), and graph-based contrastive learning (G-C), as well as the three related loss functions. The results are presented in Table VI. It is evident that the removal of the corresponding modules led to a decline in performance metrics to varying extents. This confirms the critical importance of these three modules in attaining model performance. The G-S and D-R modules, in particular, showed a significant enhancement in performance, suggesting that the construction of ECG graph representation and noise reduction reconstruction within the model are crucial components. The D-R module gives the most obvious improvement of the three parts, and the denoising reconstruction occupies a large part of the role, which can remove ECG noise signals and retain the essential characteristics.

Additionally, we show the effects of the three loss functions in Table VI. The removal of any loss function resulted in a certain degree of performance decrease, indicating that all these loss functions positively impact the model’s performance. The complete model demonstrates superior performance across all the tests, illustrating the significance of our proposed modules and loss functions in enhancing ECG classification accuracy, enabling versatile and risk-sensitive cardiac diagnosis.

V. DISCUSSION

Motivated by clinical use, we develop a unified and interpretable ECG representation that handles diverse acquisition configurations while providing clinician-ready evidence. VARS converts heterogeneous waveforms into a graph representation so a single encoder can process different lead counts, sampling rates, and durations. The explanation pathway surfaces time-stamped intervals that support each prediction and can be overlaid in standard ECG viewers, enabling rapid triage and targeted review. At inference, the runtime path is lean: attention-based graph construction, a single encoder pass, and on-demand evidence visualization that reuses encoder embeddings with modest overhead. As summarized in Table V, VARS offers a favorable compute profile—moderate parameter

count, low FLOPs, and competitive end-to-end latency—while the interpretability results (Figures 3–4 and Appendix B) show that highlighted intervals align well with clinically informed annotations. Together, these properties suggest that VARS is a promising candidate for high-throughput ECG review in contemporary clinical workflows, and broader real-world validation remains an important next step.

VI. CONCLUSION

In this work, we introduce the innovative VARS model for a unified ECG signal representation, which effectively tackles the challenges posed by the heterogeneity of ECG signals and the deficiencies in risk signal detection due to sample imbalances. VARS transforms ECG signals into versatile graph structures that capture essential diagnostic features, regardless of variations in lead count, sampling frequency, and duration. This graph-based approach enhances diagnostic sensitivity by enabling precise localization and identification of anomalous ECG patterns that typically elude conventional analysis methods. By integrating denoising reconstruction and contrastive learning, VARS not only preserves the raw ECG information but also accentuates pathognomonic patterns. The experimental results demonstrate that VARS consistently excels in performance across three widely-used datasets and shows significant improvements in detecting risk signals. Additionally, VARS enhances the interpretability of classification results through various granular levels of analysis (discussed below), thereby assisting doctors in making informed decisions.

While this work establishes a strong foundation, the path to clinical deployment requires further investigation. Future work will focus on two key areas. First, a more extensive evaluation on real-world, noisy data from continuous monitoring devices is necessary to confirm the model’s robustness outside of curated datasets. Finally, prospective clinical trials are essential to validate the clinical efficacy of VARS and to quantify its impact on diagnostic accuracy, workflow efficiency, and patient outcomes. We believe that addressing these challenges will pave the way for VARS to become an invaluable tool for comprehensive cardiac health assessment.

REFERENCES

- [1] K. C. Siontis, P. A. Noseworthy, Z. I. Attia, and P. A. Friedman, “Artificial intelligence-enhanced electrocardiography in cardiovascular disease management,” *Nature Reviews Cardiology*, vol. 18, no. 7, pp. 465–478, 2021.
- [2] I. Fernández-Ruiz, “Artificial intelligence to improve the diagnosis of cardiovascular diseases,” *Nature Reviews Cardiology*, vol. 16, no. 3, pp. 133–133, 2019.
- [3] M. Gadaleta, P. Harrington, E. Barnhill, E. Hytopoulos, M. P. Turakhia, S. R. Steinhubl, and G. Quer, “Prediction of atrial fibrillation from at-home single-lead ecg signals without arrhythmias,” *NPJ Digital Medicine*, vol. 6, no. 1, p. 229, 2023.
- [4] S. V. Kalmady, A. Salimi, W. Sun, N. Sepehrvand, Y. Nademi, K. Baaney, J. Ezekowitz, A. Hindle, F. McAlister, R. Greiner *et al.*, “Development and validation of machine learning algorithms based on electrocardiograms for cardiovascular diagnoses at the population level,” *NPJ Digital Medicine*, vol. 7, no. 1, p. 133, 2024.
- [5] E. Lee, S. Ito, W. R. Miranda, F. Lopez-Jimenez, G. C. Kane, S. J. Asirvatham, P. A. Noseworthy, P. A. Friedman, R. E. Carter, B. A. Borlaug *et al.*, “Artificial intelligence-enabled ecg for left ventricular diastolic function and filling pressure,” *npj Digital Medicine*, vol. 7, no. 1, p. 4, 2024.

- [6] H. Wang, Z. Luo, J. W. Yip, C. Ye, and M. Zhang, "Ecggan: A framework for effective and interpretable electrocardiogram anomaly detection," in *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2023, pp. 5071–5081.
- [7] X. Mei, N. Rao, Q. Li, C. Luo, B. F. Kipkurui, and H. Jiang, "Detecting atrial fibrillation from single-lead ECG using unbalanced multi-classification support vector machine," in *Proceedings of the 2019 4th International Conference on Machine Learning Technologies*, 2019, pp. 66–69.
- [8] B.-H. Kung, P.-Y. Hu, C.-C. Huang, C.-C. Lee, C.-Y. Yao, and C.-H. Kuan, "An efficient ECG classification system using resource-saving architecture and random forest," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 6, pp. 1904–1914, 2021.
- [9] X. Liu, H. Wang, Z. Li, and L. Qin, "Deep learning in ECG diagnosis: A review," *Knowledge-Based Systems*, vol. 227, p. 107187, 2021.
- [10] S. S. Al-Zaiti, C. Martin-Gill, J. K. Zègre-Hemsey, Z. Bouzid, Z. Faramand, M. O. Alrawashdeh, R. E. Gregg, S. Helman, N. T. Riek, K. Kraevsky-Phillips *et al.*, "Machine learning for ECG diagnosis and risk stratification of occlusion myocardial infarction," *Nature Medicine*, vol. 29, no. 7, pp. 1804–1813, 2023.
- [11] H. V. Denysyuk, R. J. Pinto, P. M. Silva, R. P. Duarte, F. A. Marinho, L. Pimenta, A. J. Gouveia, N. J. Gonçalves, P. J. Coelho, E. Zdravetski *et al.*, "Algorithms for automated diagnosis of cardiovascular diseases based on ECG data: A comprehensive systematic review," *Heliyon*, vol. 9, no. 2, 2023.
- [12] S. Zhang, Q. Sun, L. Yao, A. Zhang, and R. Zhang, "Feature detection algorithm combined with machine learning applied to abnormal ECG diagnosis system," in *Proceedings of the World Conference on Intelligent and 3-D Technologies (WC3DIT 2022) Methods, Algorithms and Applications*. Springer, 2023, pp. 207–216.
- [13] S. Hong, Y. Zhou, J. Shang, C. Xiao, and J. Sun, "Opportunities and challenges of deep learning methods for electrocardiogram data: A systematic review," *Computers in Biology and Medicine*, vol. 122, p. 103801, 2020.
- [14] Y. Zhou, S. Hong, J. Shang, M. Wu, Q. Wang, H. Li, and J. Xie, "K-margin-based residual-convolution-recurrent neural network for atrial fibrillation detection," *arXiv preprint arXiv:1908.06857*, 2019.
- [15] A. Rath, D. Mishra, G. Panda, and S. C. Satapathy, "Heart disease detection using deep learning methods from imbalanced ECG samples," *Biomedical Signal Processing and Control*, vol. 68, p. 102820, 2021.
- [16] Z. Hou, X. Liu, Y. Cen, Y. Dong, H. Yang, C. Wang, and J. Tang, "Graphmae: Self-supervised masked graph autoencoders," in *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2022, pp. 594–604.
- [17] S. Wu, Y. Tang, Y. Zhu, L. Wang, X. Xie, and T. Tan, "Session-based recommendation with graph neural networks," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 346–353.
- [18] K. Han, Y. Wang, J. Guo, Y. Tang, and E. Wu, "Vision GNN: An image is worth graph of nodes," *arXiv preprint arXiv:2206.00272*, 2022.
- [19] Z. Cheng, Y. Yang, W. Wang, W. Hu, Y. Zhuang, and G. Song, "Time2Graph: Revisiting time series modeling with dynamic shapelets," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, 2020, pp. 3617–3624.
- [20] J. Chen, Y. Yang, T. Yu, Y. Fan, X. Mo, and C. Yang, "BrainNet epileptic wave detection from SEEG with hierarchical graph diffusion learning," in *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2022, pp. 2741–2751.
- [21] Z. Jia, Y. Lin, J. Wang, R. Zhou, X. Ning, Y. He, and Y. Zhao, "Graph-SleepNet: Adaptive spatial-temporal graph convolutional networks for sleep stage classification," in *IJCAI*, vol. 2021, 2020, pp. 1324–1330.
- [22] A. Deng and B. Hooi, "Graph neural network-based anomaly detection in multivariate time series," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, 2021, pp. 4027–4035.
- [23] C. Ding, S. Sun, and J. Zhao, "MST-GAT: A multimodal spatial-temporal graph attention network for time series anomaly detection," *Information Fusion*, vol. 89, pp. 527–536, 2023.
- [24] S. Biton, M. Aldhfeeri, E. Marcusohn, K. Tsutsui, T. Szwagier, A. Elias, J. Oster, J. M. Sellal, M. Suleiman, and J. A. Behar, "Generalizable and robust deep learning algorithm for atrial fibrillation diagnosis across geography, ages and sexes," *NPJ Digital Medicine*, vol. 6, no. 1, p. 44, 2023.
- [25] Y. Yehuda, D. Freedman, and K. Radinsky, "Self-supervised classification of clinical multivariate time series using time series dynamics," in *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2023, pp. 5416–5427.
- [26] B. T. Lee, Y.-Y. Jo, S.-Y. Lim, Y. Song, and J.-m. Kwon, "Efficient data augmentation policy for electrocardiograms," in *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, 2022, pp. 4153–4157.
- [27] M. D. Le, V. S. Rathour, Q. S. Truong, Q. Mai, P. Brijesh, and N. Le, "Multi-module recurrent convolutional neural network with Transformer encoder for ECG arrhythmia classification," in *2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI)*. IEEE, 2021, pp. 1–5.
- [28] M. Wang, S. Rahardja, P. Fränti, and S. Rahardja, "Single-lead ECG recordings modeling for end-to-end recognition of atrial fibrillation with dual-path RNN," *Biomedical Signal Processing and Control*, vol. 79, p. 104067, 2023.
- [29] M. S. Islam, K. F. Hasan, S. Sultana, S. Uddin, J. M. Quinn, M. A. Moni *et al.*, "HARDC: A novel ECG-based heartbeat classification method to detect arrhythmia using hierarchical attention based dual structured RNN with dilated CNN," *Neural Networks*, vol. 162, pp. 271–287, 2023.
- [30] F. Mason, A. C. Pandey, M. Gadaleta, E. J. Topol, E. D. Muse, and G. Quer, "Ai-enhanced reconstruction of the 12-lead electrocardiogram via 3-leads with accurate clinical assessment," *NPJ Digital Medicine*, vol. 7, no. 1, p. 201, 2024.
- [31] Ö. Yıldırım, P. Plawiak, R.-S. Tan, and U. R. Acharya, "Arrhythmia detection using deep convolutional neural network with long duration ECG signals," *Computers in Biology and Medicine*, vol. 102, pp. 411–420, 2018.
- [32] S. D. Goodfellow, A. Goodwin, R. Greer, P. C. Laussen, M. Mazwi, and D. Eytan, "Towards understanding ECG rhythm classification using convolutional neural networks and attention mappings," in *Machine Learning for Healthcare Conference*. PMLR, 2018, pp. 83–101.
- [33] J. Niu, Y. Tang, Z. Sun, and W. Zhang, "Inter-patient ECG classification with symbolic representations and multi-perspective convolutional neural networks," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 5, pp. 1321–1332, 2019.
- [34] M. Z. Kolk, D. M. Frodi, J. Langford, T. O. Andersen, P. K. Jacobsen, N. Risum, H. L. Tan, J. H. Svendsen, R. E. Knops, S. Z. Diederichsen *et al.*, "Deep behavioural representation learning reveals risk profiles for malignant ventricular arrhythmias," *NPJ digital medicine*, vol. 7, no. 1, p. 250, 2024.
- [35] A. H. Ribeiro, M. H. Ribeiro, G. M. Paixão, D. M. Oliveira, P. R. Gomes, J. A. Canazart, M. P. Ferreira, C. R. Andersson, P. W. Macfarlane, W. Meira Jr *et al.*, "Automatic diagnosis of the 12-lead ECG using a deep neural network," *Nature Communications*, vol. 11, no. 1, p. 1760, 2020.
- [36] J. P. Allam, S. Samantray, and S. Ari, "SpEC: A system for patient specific ECG beat classification using deep residual network," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 4, pp. 1446–1457, 2020.
- [37] M. Zubair, S. Woo, S. Lim, and D. Kim, "Deep representation learning with sample generation and augmented attention module for imbalanced ECG classification," *IEEE Journal of Biomedical and Health Informatics*, 2023.
- [38] M. Nakamoto, S. Kodera, H. Takeuchi, S. Sawano, S. Katsushika, K. Ninomiya, H. Akazawa, and I. Komuro, "Self-supervised contrastive learning for electrocardiograms to detect left ventricular systolic dysfunction," in *Proceedings of the Annual Conference of JSAI*, vol. JSAI2022, 2022, pp. 1–7.
- [39] S. Rabbani and N. Khan, "Contrastive self-supervised learning for stress detection from ECG data," *Bioengineering*, vol. 9, no. 8, p. 374, 2022.
- [40] B. Grabowski, P. Glomb, W. Masarczyk, P. Plawiak, Ö. Yıldırım, U. R. Acharya, and R.-S. Tan, "Classification and self-supervised regression of arrhythmic ECG signals using convolutional neural networks," *arXiv preprint arXiv:2210.14253*, 2022.
- [41] J. Vazquez-Rodriguez, G. Lefebvre, J. Cumin, and J. L. Crowley, "Transformer-based self-supervised learning for emotion recognition," in *2022 26th International Conference on Pattern Recognition (ICPR)*. IEEE, 2022, pp. 2605–2612.
- [42] D. Gedon, A. H. Ribeiro, N. Wahlström, and T. B. Schön, "First steps towards self-supervised pretraining of the 12-lead ECG," in *2021 Computing in Cardiology (CinC)*, vol. 48. IEEE, 2021, pp. 1–4.
- [43] J. Chen, K. Liao, K. Wei, H. Ying, D. Z. Chen, and J. Wu, "ME-GAN: Learning photonic electrocardio representations for multi-view ECG synthesis conditioned on heart diseases," in *International Conference on Machine Learning*. PMLR, 2022, pp. 3360–3370.
- [44] Y. Hu, J. Chen, L. Hu, D. Li, J. Yan, H. Ying, H. Liang, and J. Wu, "Personalized heart disease detection via ECG digital twin generation," *arXiv preprint arXiv:2404.11171*, 2024.
- [45] K. Xu, W. Hu, J. Leskovec, and S. Jegelka, "How powerful are graph neural networks?" *arXiv preprint arXiv:1810.00826*, 2018.

- [46] Y. Xie, S. Katariya, X. Tang, E. Huang, N. Rao, K. Subbian, and S. Ji, "Task-agnostic graph explanations," *Advances in Neural Information Processing Systems*, vol. 35, pp. 12 027–12 039, 2022.
- [47] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *International Conference on Machine Learning*. PMLR, 2020, pp. 1597–1607.
- [48] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH arrhythmia database," *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, no. 3, pp. 45–50, 2001.
- [49] A. Sellami and H. Hwang, "A robust deep convolutional neural network with batch-weighted loss for heartbeat classification," *Expert Systems with Applications*, vol. 122, pp. 75–84, 2019.
- [50] D. K. Atal and M. Singh, "Arrhythmia classification with ECG signals based on the optimization-enabled deep convolutional neural network," *Computer Methods and Programs in Biomedicine*, vol. 196, p. 105607, 2020.
- [51] M. Hammad, S. Zhang, and K. Wang, "A novel two-dimensional ECG feature extraction and classification algorithm based on convolution neural network for human authentication," *Future Generation Computer Systems*, vol. 101, pp. 180–196, 2019.
- [52] P. Wagner, N. Strodthoff, R.-D. Bousseljot, D. Kreiseler, F. I. Lunze, W. Samek, and T. Schaeffter, "PTB-XL, a large publicly available electrocardiography dataset," *Scientific Data*, vol. 7, no. 1, p. 154, 2020.
- [53] A. Taddei, G. Distanti, M. Emdin, P. Pisani, G. Moody, C. Zeelenberg, and C. Marchesi, "The European ST-T database: Standard for evaluating systems for the analysis of ST-T changes in ambulatory electrocardiography," *European Heart Journal*, vol. 13, no. 9, pp. 1164–1172, 1992.
- [54] J. Zheng, H. Guo, and H. Chu, "A large scale 12-lead electrocardiogram database for arrhythmia study (version 1.0. 0)," *PhysioNet 2022 Available online http://physionet.org/content/ecg_arrhythmia10* Accessed on, vol. 23, p. 7, 2022.
- [55] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," in *International Conference on Learning Representations*, 2018.
- [56] J. Lee, I. Lee, and J. Kang, "Self-attention graph pooling," in *International Conference on Machine Learning*. PMLR, 2019, pp. 3734–3743.
- [57] A. Dempster, D. F. Schmidt, and G. I. Webb, "MINIROCKET: A very fast (almost) deterministic transform for time series classification," in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 2021, pp. 248–257.
- [58] F. Karim, S. Majumdar, H. Darabi, and S. Chen, "LSTM fully convolutional networks for time series classification," *IEEE Access*, vol. 6, pp. 1662–1669, 2017.
- [59] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.
- [60] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," *arXiv preprint arXiv:1803.01271*, 2018.
- [61] H. Li, Z. Lin, X. Shen, J. Brandt, and G. Hua, "A convolutional neural network cascade for face detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 5325–5334.
- [62] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, 2017.
- [63] J. Pöppelbaum, G. S. Chadha, and A. Schwung, "Contrastive learning based self-supervised time-series analysis," *Applied Soft Computing*, vol. 117, p. 108397, 2022.
- [64] A. v. d. Oord, Y. Li, and O. Vinyals, "Representation learning with contrastive predictive coding," *arXiv preprint arXiv:1807.03748*, 2018.
- [65] C.-L. Liu, B. Xiao, and C.-H. Hsieh, "Multimodal fusion of spatial-temporal and frequency representations for enhanced ecg classification," *Information Fusion*, vol. 118, p. 102999, 2025.
- [66] Y. Wei, C. Lian, B. Xu, P. Zhao, H. Yang, and Z. Zeng, "Bimodal masked autoencoders with internal representation connections for electrocardiogram classification," *Pattern Recognition*, vol. 161, p. 111311, 2025.

APPENDIX A: ANALYSIS OF LOSS COMPONENTS

We examine how the three objectives—reconstruction (L_{rec}), JSE (L_{JSE}), and contrastive loss (L_{CL})—interact during training and how their relative weights influence downstream performance. To make the analysis transparent, we adopt a one-factor-at-a-time design: in each sweep we vary a single weight over $[0.1, 1.0]$ while holding the other two weights at 1. Concretely, we sweep λ_{rec} with $\lambda_{\text{JSE}}=1$ and $\lambda_{\text{CL}}=1$, sweep λ_{JSE} with $\lambda_{\text{rec}}=1$ and $\lambda_{\text{CL}}=1$, and sweep λ_{CL} with $\lambda_{\text{rec}}=1$ and $\lambda_{\text{JSE}}=1$. All three sweeps are conducted on the Chapman–Shaoxing 12-lead dataset (multi-label) under the same pre-processing, train/test split, and training protocol described in the Experimental Setup, to reflect contemporary acquisition conditions. For each setting we report accuracy, macro-F1, and macro-AUC.

As shown in Figure 1, two general patterns are apparent across the three sweeps: (i) the metric curves remain fairly flat over a broad range, with no consistent monotonic improvement when any single weight departs substantially from 1, and (ii) the vicinity of the equal-weight setting ($\lambda_{\text{rec}} = \lambda_{\text{JSE}} = \lambda_{\text{CL}} = 1$) forms a wide performance plateau, where accuracy, macro-F1, and macro-AUC are near their maxima. Overall, these results indicate limited sensitivity to moderate reweighting of the three objectives and support the use of fixed equal weights for simplicity.

APPENDIX B: EXPLAINABILITY ANALYSIS AND VISUALIZATION DETAILS

Consistent with the quantitative analysis in the main text, we quantify interpretability using model-internal signals emitted by the Feature Subgraph Module, without introducing an external post-hoc explainer. For each sample, the module produces node- and edge-level importance on the ECG graph and highlights the time-interval segments that constitute the explanatory subgraph. The composite Figure 2 below gathers these outputs into a single, per-sample dashboard to complement the qualitative plots in the main paper.

Operationally, ECG signals are split into fixed-size patches, and the module scores each patch to obtain patch importance. Saliency rendering uses a configurable threshold (default $\tau=0.3$), and top- K or quantile selection can be used to control display sparsity. The per-sample dashboard can be produced for any correctly classified case (e.g., the highest-confidence instance within a class).

The dashboard comprises four coordinated panels generated directly from the Feature Subgraph Module: (i) A multi-lead ECG with salient intervals, where the waveform is uniformly partitioned into time patches and segments exceeding a user-set threshold are shaded. (ii) A node-importance heatmap, obtained by zero-padding and reshaping the node vector into a near-square matrix for inspection. (iii) An edge-importance histogram, providing a distributional view of relational weights with reference lines for quick calibration, indicating whether evidence concentrates on a few high-weight connections or is more diffuse, and (iv) a prediction summary showing class probabilities.

Figure 3 provides a lead-wise view for one normal case (left) and one abnormal case (right), each showing the full 10 s waveform per lead as a complementary view. Each is labeled by its patch index and time range. In the normal example, no pathologic interval is marked, which is consistent with routine readings. In the abnormal example, the top-1 window selected by the interpreter aligns with the clinically annotated interval across multiple leads. This illustrates concordance between the model’s most salient evidence and clinical labeling. This full-signal overlay complements the per-sample dashboard by revealing how highlighted windows cohere across leads and by making disagreements—when present—visually apparent.

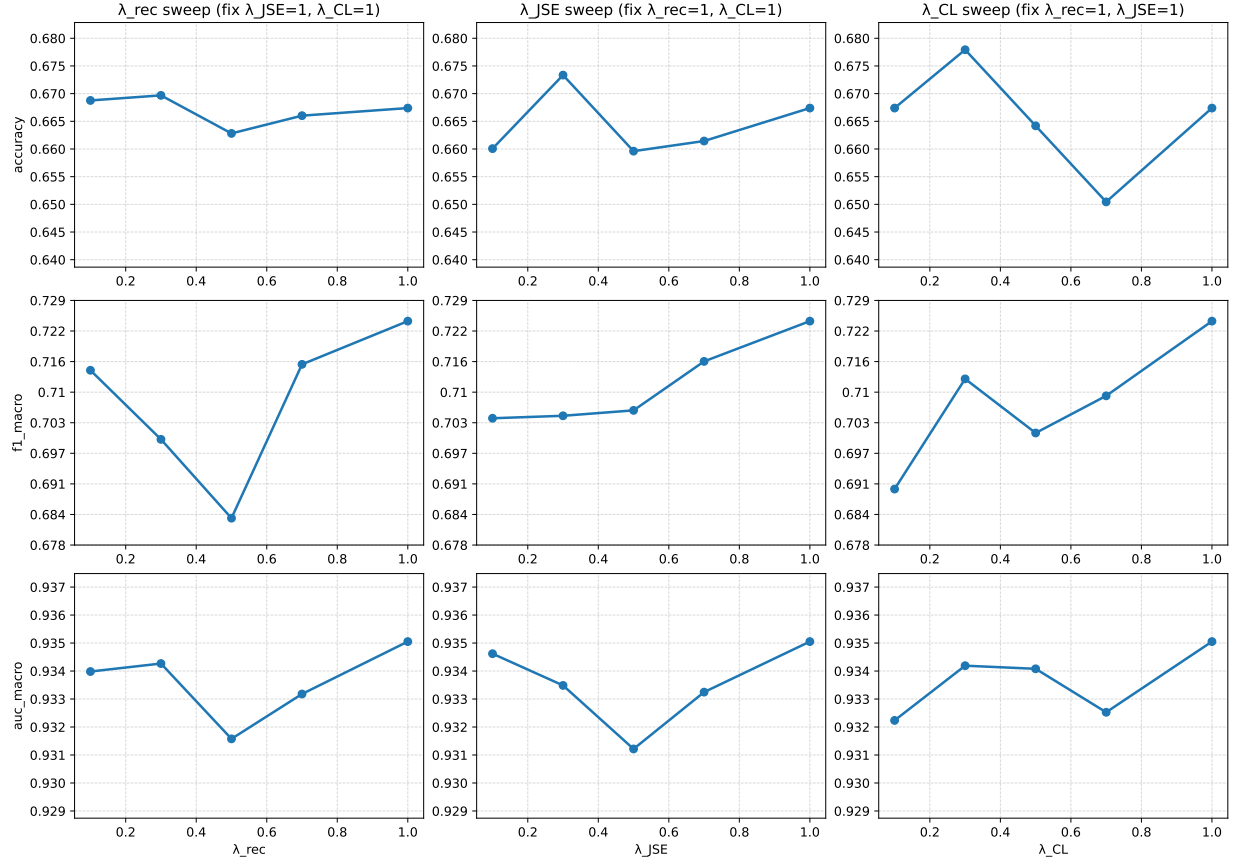


Fig. 1. One-factor-at-a-time sensitivity of loss weights. Columns sweep a single weight (λ_{rec} , λ_{JSE} , λ_{CL}) over $[0.1, 1.0]$ with the other two fixed at 1. Rows report accuracy, macro-F1, and macro-AUC. Curves are stable across the range, and the equal-weight point ($\lambda=1$) lies on a broad plateau across metrics.

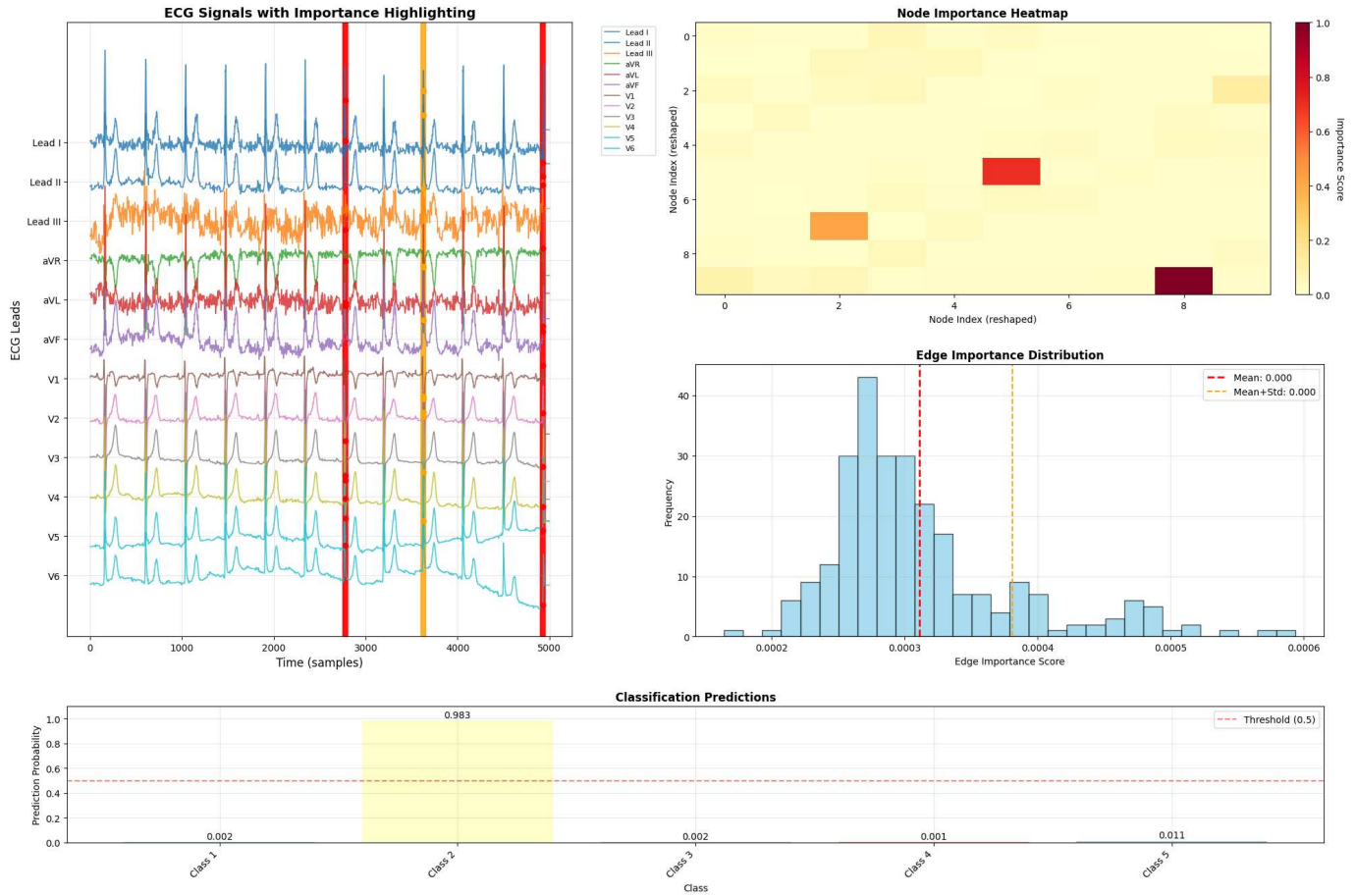


Fig. 2. Per-sample explainability dashboard generated from the Feature Subgraph Module. Panels: multi-lead ECG with time highlights (left), node-importance heatmap and edge-importance histogram (top/right), and prediction summary (bottom). Salient time windows appear coherently across leads, node importance is concentrated within a small subset of patches, and the edge distribution frequently exhibits a long tail, indicating sparse but critical relational evidence supporting the decision.

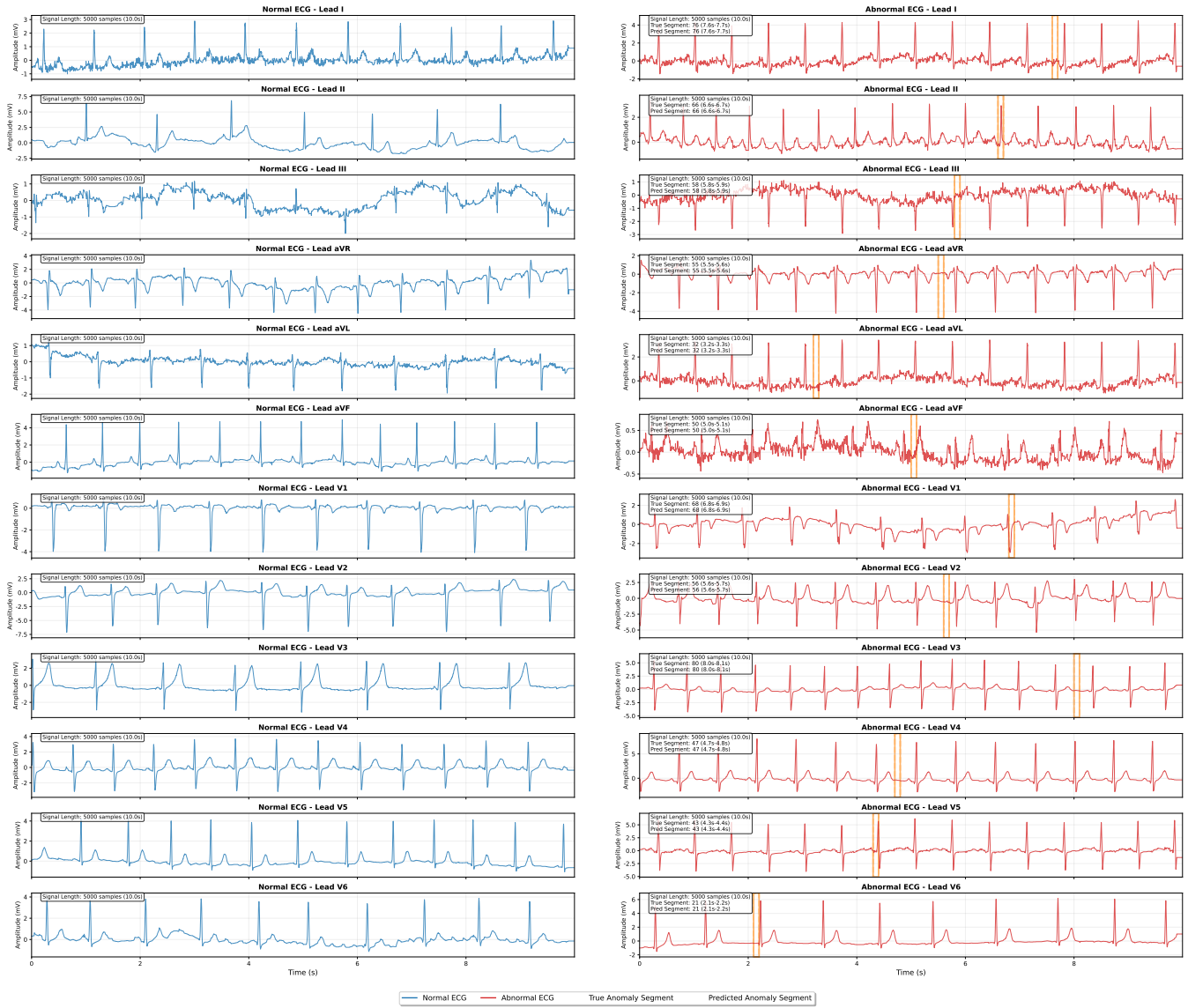


Fig. 3. Lead-wise overlay of clinically annotated and model-interpreted segments. **Left:** a normal ECG example (10s per lead). **Right:** an abnormal ECG example, where the interpreter highlights the most salient abnormal segment(s). The top-1 segment is indicated together with its index and time range.