Interdisciplinary Sciences: Computational Life Sciences
https://doi.org/10.1007/s12539-025-00802-7

ORIGINAL RESEARCH ARTICLE

FHGNet: A Feature-Centric Hierarchical Network with Graph Attention Layer for Supraventricular Tachycardia Classification

Xiaolin Ju[1] · Tao Liu[1] · Bowen Luo[1] · Heling Cao[2] · Zhan Gao[1] · Haiyan Pan[3]

Received: 31 July 2025 / Revised: 18 November 2025 / Accepted: 22 November 2025

Abstract

Automated electrocardiogram (ECG) classification plays a critical role in arrhythmia diagnosis. However, current deep learning-based methodologies frequently fail to account for physiological rhythms and clinical diagnostic reasoning, thereby compromising their reliability and interpretability. This study proposes a clinically inspired multi-lead oscillatory Transformer framework, named FHGNet, to enhance the precision and interoperability of classifying ventricular tachycardia (VT) and supraventricular tachycardia (SVT). The proposed architecture integrates R-peak detection for heartbeat segmentation and adaptive-length patch extraction with R-wave positional encoding to enhance temporal awareness. It employs a convolutional neural network (CNN) to capture intra-beat morphological features (QRS morphology), a Transformer with FANLayer to model inter-beat rhythmic patterns, and a graph attention network (GAT) to fuse multi-lead dependencies. Additionally, a two-stage classifier is designed to enhance the detection of rare arrhythmia classes. Experimental evaluations on the MIT-BIH Supraventricular Arrhythmia dataset demonstrate FHGNet achieves a macro F1-score of 91.35% outperforming baselines. Ablation studies reveal that removing GAT reduces F1 by 2.42% in multi-lead scenarios, while the two-stage design improves minority class recall by 5.82%. Attention visualization confirms the model focuses on clinically relevant features, such as ST-T segment energy ratios and inter-lead phase differences, aligning with established diagnostic criteria. Additionally, the interpretability of FHGNet is further enhanced by two aspects: 1) Explicit integration of physiological priors (e.g., RR interval variability, intra-beat positional information) in dynamic feature engineering, which enables the model to align with clinicians' rhythm analysis logic; 2) The two-stage classifier strictly follows the clinical diagnostic workflow (first screening abnormalities, then subclassifying), making the decision-making process traceable. This work provides an interpretable, clinically adaptive framework for high-accuracy ECG classification, potentially reducing reliance on invasive electrophysiological studies.

Tao Liu and Bowen Luo have contributed equally to this work.

✉ Bowen Luo
cooing.code@gmail.com

✉ Zhan Gao
gaozhan@ntu.edu.cn

✉ Haiyan Pan
dr.phy@ntu.edu.cn

Xiaolin Ju
ju.xl@ntu.edu.cn

Tao Liu
Liu.tao@stmail.ntu.edu.cn

Heling Cao
caohl@haut.edu.cn

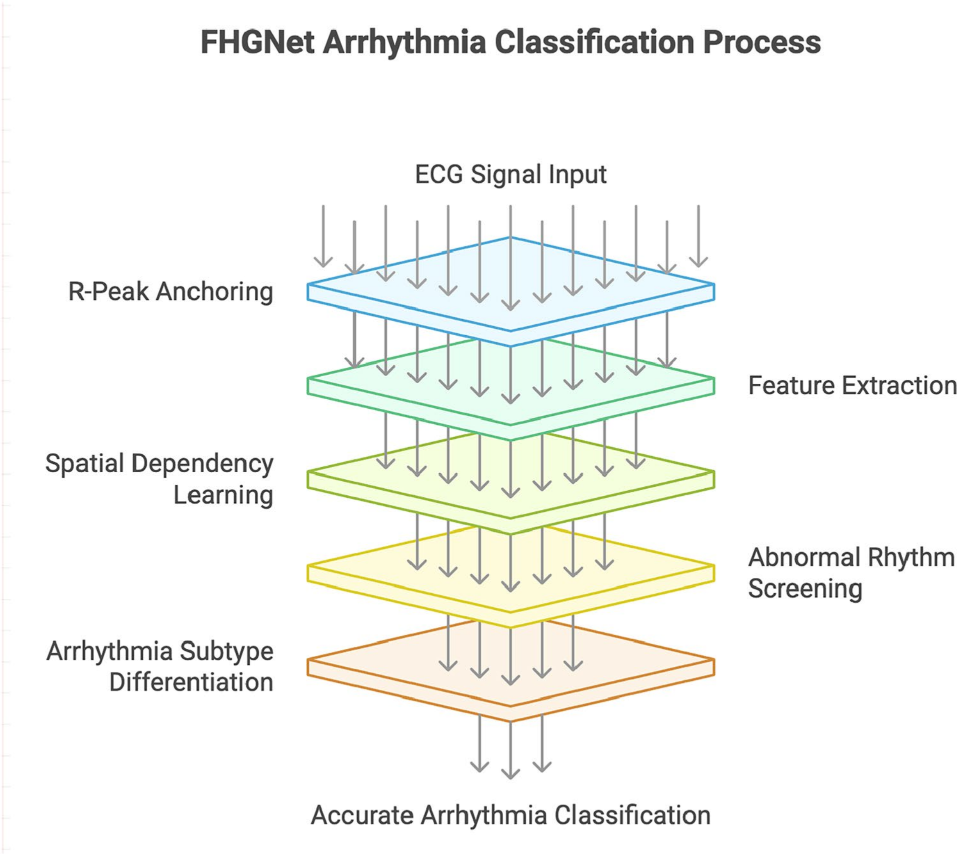
[1] School of Artificial Intelligence and Computing, Nantong University, Nantong 226019, China

[2] College of Information Science and Engineering, Henan University of Technology, Zhengzhou 450001, China

[3] Affiliated Hospital of Nantong University, Nantong 226019, China

Published online: 03 February 2026

Graphical Abstract


FHGNet Arrhythmia Classification Process
ECG Signal Input
R-Peak Anchoring
Feature Extraction
Spatial Dependency Learning
Abnormal Rhythm Screening
Arrhythmia Subtype Differentiation
Accurate Arrhythmia Classification

Keywords Electrocardiogram (ECG) classification · Deep learning · Transformer · Multi-lead analysis · Oscillatory feature

1 Introduction

The electrocardiogram (ECG) is a cornerstone [1] for cardiovascular diagnosis, and the automated classification of life-threatening arrhythmias [2], such as ventricular tachycardia (VT) and supraventricular tachycardia (SVT), is critical for timely intervention. Differentiating VT from SVT guides therapeutic decisions [3] (cardioversion, antiarrhythmic drugs, or radiofrequency ablation), but their morphological overlaps (e.g., wide-QRS SVT vs. VT) and spatiotemporal variations in multi-lead signals pose significant challenges for traditional rule-based methods.

Clinicians' interpretation of ECG follows a rigorous logical chain [4]: beginning with R-wave localization to capture complete P-QRS-T complexes, then assessing QRS width and ST-segment deviation [5, 6], analyzing RR interval patterns, integrating multi-lead features [7], and finally differentiating abnormal subtypes through staged reasoning [8].

Deep learning and clinical reasoning steps, such as “localization before analysis", can lead to poor generalization in scenarios with treatable morphological variations (e.g., ST-segment abnormalities) and quasiperiodic-to-noise ratios (e.g., baseline drift). Meanwhile, most existing models use fixed windows to extract heartbeats, failing to adapt to dynamic RR interval changes, which easily causes truncation of escape beats after long pauses or premature beats with short coupling intervals. Additionally, the lack of explicit encoding [9, 10] for the relative position of R-waves makes it challenging to simulate clinicians' sensitivity to the temporal relationship of waveforms.

Furthermore, traditional methods for multi-lead information integration [11, 12] often use simple concatenation or static weighting, ignoring dynamic dependencies between leads (e.g., spatial correlations between the frontal plane axis and the precordial transition zone). Recent advances in multi-lead ECG modeling have focused on attention and graph-based methods: Zhang et al. [13] proposed a pattern-guided attention Transformer that integrates physiological priors for interpretable ECG classification, but it does not explicitly address dynamic multi-lead spatial dependencies;

Li et al. [14] developed a multi-scale shifted window Transformer for 12-lead ECG classification, capturing temporal patterns at multiple resolutions, yet without combining graph attention to model inter-lead relationships. However, few studies have combined graph attention with Transformer to simultaneously capture dynamic spatial dependencies and long-range temporal rhythms, which is a key gap addressed by FHGNet.

Similar to recent advances in other cardiovascular studies that integrate computational modeling with clinical practice, our work emphasizes not only algorithmic accuracy but also clinical interpretability, aiming to facilitate practical diagnostic adoption.

To address these limitations, we introduce FHGNet (a feature-centric hierarchical network with graph attention layer for SVT classification), a framework designed to emulate the clinical diagnostic workflow for enhanced arrhythmia classification. The core contributions of this work are:

(1) *Clinically-Informed Input Representation*: A globally optimized R-peak anchoring strategy, combined with an adaptable patch length, ensures that each heartbeat is analyzed within a standardized and physiologically comprehensive context. This approach facilitates consistent and accurate analysis of cardiac signals while preserving the integrity of their temporal and morphological features.

(2) *Dual-Branch Feature Decoupling*: A parallel architecture is proposed, featuring a convolutional neural network (CNN) branch to capture intra-beat **morphological characteristics** and a Transformer branch to model inter-beat *rhythm patterns*. This structure emulates the diagnostic approach employed by clinicians, who distinctly analyze waveform morphology and rhythm to ensure a comprehensive assessment of cardiac signals.

(3) *Relational Multi-Lead Fusion*: The graph attention network (GAT) explicitly learns and assigns the dynamic spatial dependencies between ECG leads. By moving beyond conventional concatenation methods, GAT effectively models inter-lead relationships, enabling a more sophisticated representation of spatial interactions within multi-lead ECG signals.

(4) *Hierarchical Diagnostic Reasoning*: The two-stage classification framework is designed to first differentiate normal from abnormal heartbeats and then classify the specific subtypes of abnormal heartbeats. This strategy enhances the detection of infrequent classes while aligning with clinical decision-making, thereby improving overall classification accuracy and clinical relevance.

Experiments on the MIT-BIH Supraventricular Arrhythmia Database (SVDB) [15], which contains 16574 samples (N: 8131, VT: 3815, SVT: 4628), validate the necessity of biomimetic components through ablation studies and reveal the decision logic via attention visualization.

FHGNet is expected to outperform traditional methods in variant morphologies and multi-lead scenarios, providing an interpretable tool for clinical diagnosis and potentially reducing reliance on invasive electrophysiological studies.

2 Related Work

As a core task in ECG analysis, the evolution of arrhythmia classification methodologies reflects the technological iteration from signal processing to data-driven approaches.

Existing research primarily revolves around single-lead feature extraction, multi-lead spatial correlation modeling, and temporal dynamics analysis, which can be categorized into three frameworks: digital signal processing-based [16, 17], traditional machine learning-based [18, 19], and deep learning-based [20, 21]. Each framework has distinct limitations in feature representation, model generalization, and clinical applicability.

2.1 Digital Signal Processing-Based Arrhythmia Classification Methods

Early studies on arrhythmia classification relied primarily on digital signal processing techniques to perform time-frequency analysis, noise filtering, or waveform enhancement on ECGs to extract discriminative features between VT and SVT.

The wavelet transform (WT) is widely used for its multi-resolution analysis capability. For example, Martinez et al. [22] employed single-lead WT to extract morphological characteristics (e.g., amplitude, duration) of P waves, QRS complexes, and T waves, combining these with statistical methods for arrhythmia classification. Rincon et al. [23] extended multi-lead WT algorithms to wireless body sensor networks for real-time dynamic ECG classification. However, these methods heavily depend on intensive mathematical computations and manual threshold setting (e.g., QRS amplitude thresholds, ST-segment offset frequency thresholds), lacking robustness to low-signal-to-noise ratio (SNR) signals or morphologically variant waveforms (e.g., VT with atrial fibrillation).

The switching Kalman filter (SKF) model [24] modeled ECG waveforms using Gaussian functions and baseline drift using autoregressive models, reducing mean error but being highly sensitive to manually set initial parameters, which limits its adaptability to diverse data acquired by different devices. The phasor transform (PT) simplifies feature extraction by enhancing the amplitude of P and T waves. Still, it focuses solely on single-lead waveform enhancement, ignoring the spatial potential correlations among 12 leads (e.g., inter-lead amplitude differences typical in VT), thus losing critical discriminative information.

In summary, prior knowledge and manual parameter tuning constrain methods based on digital signal processing, with limited generalization to complex arrhythmias.

2.2 Traditional Machine Learning-Based Arrhythmia Classification Methods

Traditional machine learning methods integrate ECGs' temporal, frequency-domain, and statistical features through manual feature engineering, with typical algorithms including artificial neural networks (ANN)] [25], genetic algorithms [26], Bayesian models [27], and *k*-nearest neighbors (KNN) [28].

Agrawal et al. [29] developed an ANN-based adaptive whitening filter to remove nonlinear noise from ECGs, improving QRS detection accuracy and providing clean signal inputs for subsequent classification. Nowostawski et al. [30] used genetic algorithms to optimize feature combinations (e.g., RR interval, QRS axis) via evolutionary strategies, but the method suffered from local optimum issues, failing to cover rare arrhythmia patterns. Bayesian models leverage the local correlation of ECG signals to model the probability distribution of P and T waves, achieving high accuracy in single-lead classification tasks. However, these models encounter significant high computational costs in multi-lead scenarios, limiting their scalability and efficiency in such contexts.

The KNN-based approaches detect fiducial points and waveform boundaries with simplicity and efficiency. However, their performance heavily depends on the local structure of the training data, resulting in poor generalization to cross-center ECGs (e.g., different lead configurations and noise levels). A standard limitation of these methods is their reliance on domain experts for feature design (e.g., heart rate variability, ST-segment slope, inter-lead amplitude differences), which leads to high-dimensional feature spaces that are vulnerable to noise and unable to capture latent discriminative patterns of VT/SVT automatically. Additionally, they do not explicitly model spatial dependencies among multi-leads (e.g., synergistic changes between limb and chest leads).

2.3 Deep Learning-Based Arrhythmia Classification Methods

The end-to-end feature learning capability of deep learning has advanced arrhythmia classification toward automation, with CNNs [31] and recurrent neural networks (e.g., LSTM) as mainstream architectures.

Li et al. [32] proposed a two-step CNN framework to first segment QRS complexes and then locate their boundaries via feature mapping, enabling the classification of single-lead QRS-related arrhythmias. Abrishami et al. [33] utilized bidirectional long short-term memory (BiLSTM) networks to capture dynamic changes in RR intervals, achieving higher classification accuracy on single-lead data compared to traditional models. To fuse multi-lead information, hybrid deep networks were developed: Semwal et al. [34] combined CNN-based spatial feature extraction with BiLSTM-based temporal dependency modeling to improve the detection of ST-segment abnormality-related arrhythmias. Nurmaini et al. [35] and Peimankar et al. [36] validated the performance gain of multi-modal feature fusion by stacking convolutional and recurrent modules.

Encoder-decoder architectures (e.g., U-Net, BiLSTM encoders) adapt to sequence-to-sequence tasks via hierarchical feature extraction, while Wang et al. [37] further integrated domain knowledge (e.g., rules of electrical axis deviation) to optimize model decisions, significantly improving the classification accuracy for complex cases.

Nevertheless, existing deep learning methods face critical challenges: most treat 12 leads as independent signals, failing to explicitly model inter-lead spatial correlations (e.g., dynamic learning of lead weights via GATs); temporal modeling relies on recurrent networks or simple attention mechanisms, insufficient for capturing long-range rhythmic patterns (e.g., periodic wide QRS complexes in VT); moreover, the time-frequency oscillatory features of ECGs (e.g., high-frequency regular oscillations in SVT vs. low-frequency chaotic oscillations in VT) are underutilized, limiting the discriminative ability for morphologically variant cases.

Transformer-based models, while effective in capturing long-range temporal dependencies, have several key limitations in ECG classification. These include high computational overhead due to the quadratic time complexity of self-attention ($\mathcal{O}(N^2)$, where N denotes the sequence length in self-attention computation), a lack of inductive bias for capturing local morphological features, and limited ability to model spatial dependencies between ECG leads. In response, FHGNet integrates CNNs and GATs to overcome these limitations, capturing both local and long-range features while modeling dynamic spatial relationships.

FHGNet proposed in this study explicitly captures time-frequency oscillatory features via a multi-branch feature extraction network, dynamically learns inter-lead spatial dependencies using a GAT, and integrates spatio-temporal features through Transformer encoders, thereby addressing these limitations. This model presents a novel framework for VT/SVT classification that fuses oscillatory feature representation, spatial correlation modeling, and temporal dynamic analysis.

Recent studies have also proposed advanced deep learning frameworks for multi-lead ECG analysis. For instance, IM-ECG employs dual-kernel residual blocks (DKR-blocks) combined with Grad-CAM visualization to improve feature extraction and interpretability across multiple leads. Meanwhile, ECGMamba [38] introduces a bidirectional state space model (BiSSM), demonstrating competitive performance with Transformer-based methods while significantly reducing computational cost and improving inference efficiency.

These methods highlight the trend toward efficient and interpretable ECG classification. However, unlike FHGNet, they do not explicitly integrate a hierarchical two-stage classifier or graph attention-based spatial dependency modeling, which are critical for minority class detection and dynamic inter-lead feature fusion.

Recent clinical studies in related cardiovascular diagnostic tasks further highlight the importance of model robustness and interpretability. For example, Long et al. [39] employed single-cell sequencing to uncover endothelial heterogeneity after myocardial infarction, while Li et al. [40] optimized diagnostic procedures in echocardiography. These studies inspire future extensions of FHGNet toward integrating multimodal and clinical workflow-aware features.

3 Methods

3.1 The Motivation of FHGNet

In ECG analysis, prior research has predominantly relied on single CNNs to capture local morphological features. However, these approaches demonstrate insufficient capacity to model long-range rhythmic patterns between heartbeats (such as RR interval variability) and dynamic spatial correlations between multiple leads (such as QRS concordance of the QRS of the precordial lead), thus limiting the discriminative precision for complex arrhythmias such as VT and SVT.

When fusing multi-lead information, conventional methods often employ simple concatenation or static weighting strategies, overlooking the synergistic roles of different leads in diagnosis. For instance, the critical correlation between precordial transition zone localization and limb lead P-wave morphology in clinical diagnosis remains inadequately learned by existing deep learning models. Furthermore, due to the high proportion of normal samples in clinical datasets (often exceeding 60%), traditional single-stage multiclassification models are prone to bias from dominant classes, resulting in a high rate of missed diagnoses for rare abnormalities such as SVT.

Moreover, these deep learning models lack the stepwise reasoning logic of clinical practice, which involves screening for abnormalities first and then subclassifying types. To address these challenges, this study proposes the FHGNet framework (Fig. 1), which is inspired by clinical diagnostic workflows. Through innovative feature decoupling, dynamic multi-lead modeling, and staged classification strategies, this framework systematically optimizes the spatio-temporal feature representation of complex ECG signals to enhance the clinical utility of automated diagnosis.

3.2 R-peak Anchoring and Adaptive Beat Construction

We employ the Pan-Tompkins algorithm to detect R-peaks, which serve as physiological anchors [41]. This algorithm achieves an R-peak detection accuracy of over 98% for normal sinus rhythms through adaptive thresholding and band-pass filtering, but its robustness degrades for abnormal waveforms: (1) For wide-QRS VT, QRS complex morphological distortion may cause the algorithm to misidentify secondary peaks as R-peaks. (2) For atrial fibrillation with premature ventricular contractions (PVCs), large RR interval variability and baseline drift increase detection errors. To address this, we introduce a two-step verification process after initial R-peak detection:

(1) Calculate the coefficient of variation (CV) of adjacent RR intervals; if CV > 0.2 (indicating irregular rhythm), trigger morphological verification.
(2) For suspected R-peak regions, check two morphological criteria: QRS wave amplitude (must be > 3 times the baseline noise level) and QRS width (normal range: 60–120 ms for SVT, > 120 ms for VT).

This verification reduces the R-peak detection error rate for abnormal waveforms from 8% to < 3%. Subsequently, we determine an optimal patch length L_{patch} by analyzing the RR-interval distribution across the entire training dataset to standardize inputs while respecting physiological context.

This globally optimized length ensures that patches are, on average, best suited to capture a complete P-QRS-T

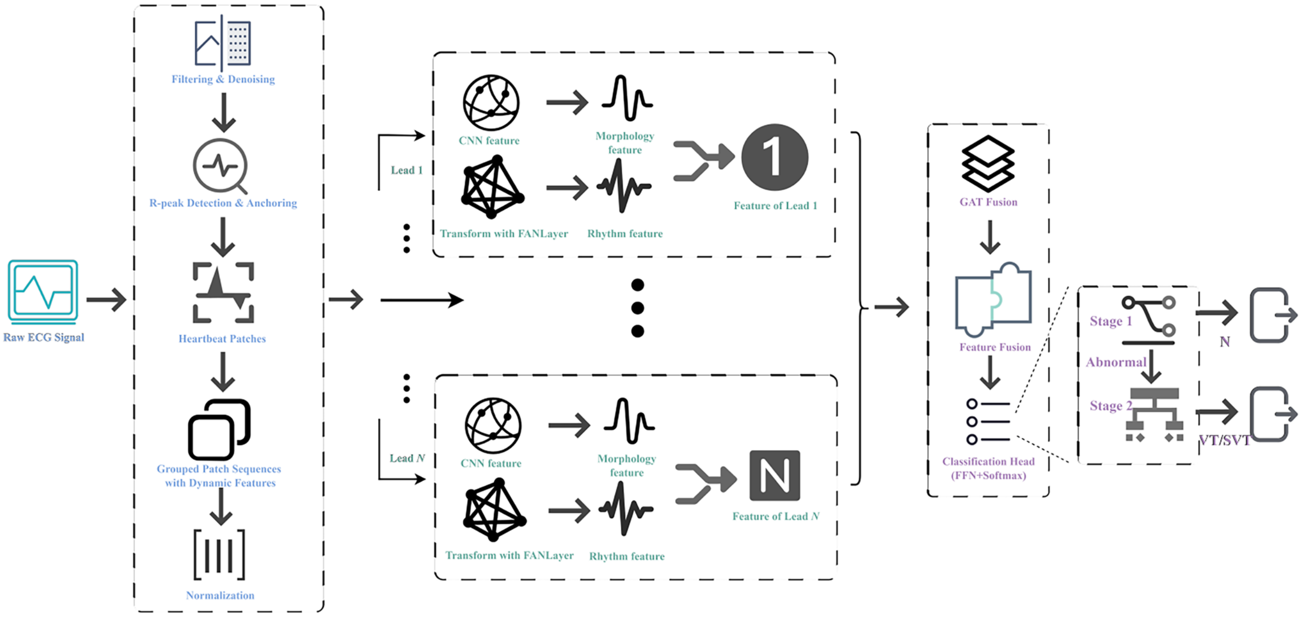


Fig. 1 Architecture of the FHGNet framework: a spatio-temporal feature modeling approach for ECG signals, integrating R-peak anchored adaptive beat extraction, CNN-Transformer dual-branch feature decoupling, GAT for multi-lead dependency modeling, and a two-stage classifier for complex arrhythmia classification

complex for the given dataset's heart rate characteristics. When processing, an asymmetric window (e.g., 1/3 pre-R, 2/3 post-R) is extracted around each R-peak using this fixed length. This strategy is adaptive when the model is applied to a new dataset. This analysis can be re-run to determine a new optimal patch length tailored to the new population's heart rate profile.

3.3 Dynamic Physiological Feature Engineering

To enrich the input representation with explicit physiological context, we engineer a set of dynamic features for each time step within a patch. These features, computed in real-time during data loading, are categorized as follows:

- *Rhythm and Variability Features*: These include the current and preceding RR-intervals normalized by the global average and the local heart rate variability (HRV) calculated over a small window of recent beats. This category provides the model with direct information about beat-to-beat timing and rhythm stability, which is crucial to distinguish regular from irregular tachycardias.
- *Intra-Beat Positional Features*: This group encodes the relative position of each sample within the cardiac cycle. It includes the distance to the central R-peak and phase information (encoded via sine and cosine functions) relative to the current beat's duration (RR interval). This helps the model understand if a feature occurs during the P-wave, QRS complex, or T-wave.
- *Sequential Context Features*: To provide context within a sequence of consecutive heartbeats (e.g., a group of 8), we include a feature indicating the patch's order in the sequence. This allows Transformer to model temporal evolution and patterns across multiple beats.

These dynamic features are first flattened and projected via a linear layer to match the embedding dimension of the ECG patch. The projected features are then combined with the patch embeddings through element-wise addition, integrating physiological priors (e.g., RR interval dynamics) with data-driven representations. This fusion enables the model to leverage explicit timing knowledge from the outset, thereby enhancing its sensitivity to clinically relevant patterns, such as ST-segment deviation and inter-lead phase relationships.

3.4 Dual-Branch Feature Extraction

Inspired by how clinicians separately evaluate waveform morphology and cardiac rhythm, FHGNet employs a dual-branch architecture for each lead to decouple these features.

- *Morphology Branch (CNN)*: A 1D-CNN is employed to capture intra-beat morphological features. A stack of convolutional layers with diverse kernel sizes (e.g., 3, 5, 7, 9) extracts hierarchical features, ranging from fine-grained details—such as QRS notching—to broader waveform configurations. Following convolution, an adaptive average pooling layer generates a fixed-dimensional feature vector for each kernel, which is then concatenated. This branch demonstrates excellence in learning spatially localized patterns within a single cardiac cycle.

- *Rhythm Branch (Transformer)*: A Transformer-based encoder with FANLayer (a frequency-aware feedforward network) is used to model long-range dependencies and rhythmic patterns across a sequence of consecutive heartbeats. By processing a sequence of patches (e.g., 8 consecutive beats), this branch is capable of learning patterns related to RR interval variability and other rhythmic markers, crucial for distinguishing arrhythmias such as VT and SVT.

3.5 Graph Attention Networks (GATs)

To explicitly model the spatial dependencies between N ECG leads, we treat the multi-lead system as a fully connected graph $G = (V, E)$, where each lead is a node $v_i \in V$ (here, G is a complete graph; v_i denotes an individual node, V is the node set, and E is the edge set). This enables the model to learn arbitrary inter-lead dependencies without relying on anatomical priors. The fused feature vector $\boldsymbol{F}_{\text{lead}_i}$ from the dual branch serves as the initial node representation $\boldsymbol{h}_i$. A GAT is then applied to learn the inter-lead relationships. The attention coefficient e_{ij} between lead i and lead j is computed as

$$e_{ij} = \text{LeakyReLU}(\boldsymbol{a}^{\text{T}}[\boldsymbol{W}\boldsymbol{h}_i || \boldsymbol{W}\boldsymbol{h}_j]) \tag{1}$$

where $\boldsymbol{W}$ is a learnable linear transformation and $\boldsymbol{a}$ is a learnable weight vector. These coefficients are then normalized across all neighbors of node i using the softmax function to obtain attention weights α_{ij}:

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k \in \text{N}_i} \exp(e_{ik})} \tag{2}$$

The updated feature representation for each node, $\boldsymbol{h}'_i$, is a weighted sum of its neighbors' features, allowing the model to dynamically prioritize the most relevant leads for a given diagnostic task:

$$\boldsymbol{h}'_i = \sigma \left(\sum_{j \in \text{N}_i} \alpha_{ij} \boldsymbol{W}\boldsymbol{h}_j \right) \tag{3}$$

In our implementation, the adjacency is initialized as a fully connected graph, ensuring that each lead can dynamically attend to all others. While this introduces $\mathcal{O}(N^2)$ complexity in attention computation, the overhead is acceptable because the number of ECG leads is relatively small (≤ 12). Furthermore, we apply sparse attention masks during training to prune low-weighted edges, which reduces redundant computation and ensures efficiency in practice.

Through this design, the GAT module allows FHGNet to flexibly capture dynamic inter-lead dependencies, complementing the temporal modeling of Transformer and the morphological extraction of the CNN branch.

3.6 Two-Stage Classifier and Confidence-Based Rejection

To tackle class imbalance and mimic clinical reasoning, a two-stage classifier is employed.

- *Stage 1: Anomaly Detection*: The fused global feature vector is first fed into a binary classifier to distinguish between the normal sinus rhythm (N) and any abnormal rhythm (S or V). We use a standard cross-entropy loss for this stage, with class weights derived from inverse frequency to handle the imbalance between normal and abnormal samples.
- *Stage 2: Subtype Classification*: If a beat is classified as abnormal in Stage 1, its feature vector is passed to a second, multi-class classifier. This stage is specifically trained to differentiate between the abnormal subtypes (e.g., S vs. V), also using cross-entropy loss.

The final loss is a weighted sum of the losses from both stages, with weights (w_1, w_2) determined via grid search on the validation set:

$$\mathcal{L}_{\text{total}} = w_1 \cdot \mathcal{L}_{\text{stage1}} + w_2 \cdot \mathcal{L}_{\text{stage2}} \tag{4}$$

To enhance clinical utility, a confidence-based rejection mechanism is implemented. The output probability of the predicted class from the model is employed as a confidence score to assess the reliability of predictions. An optimal confidence threshold is determined using the validation set by striking a balance between maximizing the accuracy of retained samples and maintaining a clinically acceptable rejection rate. During inference, predictions with confidence scores below this threshold are flagged for review by human experts.

4 Experimental Setup

4.1 Dataset

The SVDB dataset is employed as the primary dataset, comprising 48 12-lead ECG recordings (sampled at 360 Hz) with 15 classes, including normal sinus rhythm (NSR), VT, and SVT.

A “sample” is defined as a heartbeat segment extracted from the recordings at the patient level, whereas a “heartbeat segment” denotes a single cardiac cycle. The dataset encompasses 16,574 patient-level samples (derived from the 48 recordings), which are further processed into 156,145 heartbeat segments for model training and evaluation. The preprocessing pipeline emulates the logic of clinical diagnosis: R-peaks are detected using the Pan-Tompkins algorithm, which serve as temporal anchors for heartbeat segmentation.

Asymmetric patches (1/3 preceding and 2/3 following the R-peak, as illustrated in Fig. 2) are extracted with lengths dynamically adjusted to powers of two based on the average RR interval, thereby ensuring the complete capture of P-QRS-T waveforms.

By analyzing the 3D GAT attention weight distribution, we select the clinically commonly used Lead II and V1 (Fig. 3), apply 5–15 Hz bandpass filtering, and perform *Z*-score normalization to generate input sequences (dimension: time × leads). Dynamic features are encoded by converting RR intervals into sinusoidal phase encodings, which enhances the representation of rhythmic information. The dataset is split into training and test sets at an 8:2 ratio using a patient-level strategy to ensure that data from the same patient does not cross sets (thereby avoiding data leakage), and the splitting strategy is shown in Table 1.

4.2 Model Implementation

Dual Feature Extraction comprises a CNN branch and a Transformer branch. The CNN branch employs a 3-layer residual CNN (with kernel sizes 3, 5, 7, and 9) to extract morphological features, yielding 128-dimensional outputs. The Transformer branch utilizes a 4-layer encoder with FANLayer (a frequency-aware feedforward layer), which employs rotary positional encoding to capture rhythmic features.

For multi-lead fusion, leads are modeled as graph nodes, with GAT learning dynamic inter-lead weights to capture spatial dependencies explicitly. The two-stage classifier comprises Stage 1, which involves binary classification (normal/abnormal) using CrossEntropyLoss, and Stage 2, which entails multiclass classification for abnormal subtypes using CrossEntropyLoss. The model was implemented using PyTorch. Key hyperparameters were determined via experimental validation on the validation set.

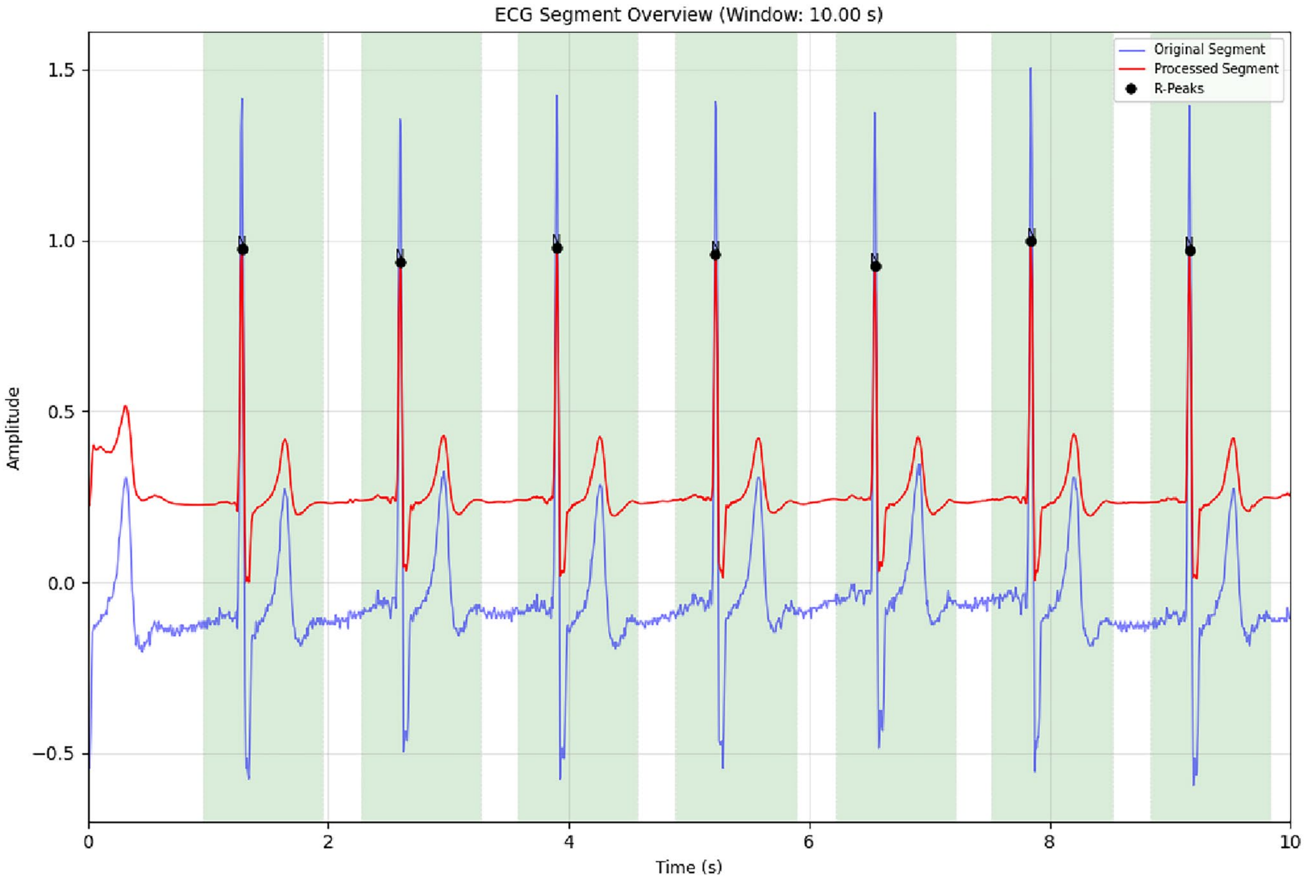


Fig. 2 RR interval windowing

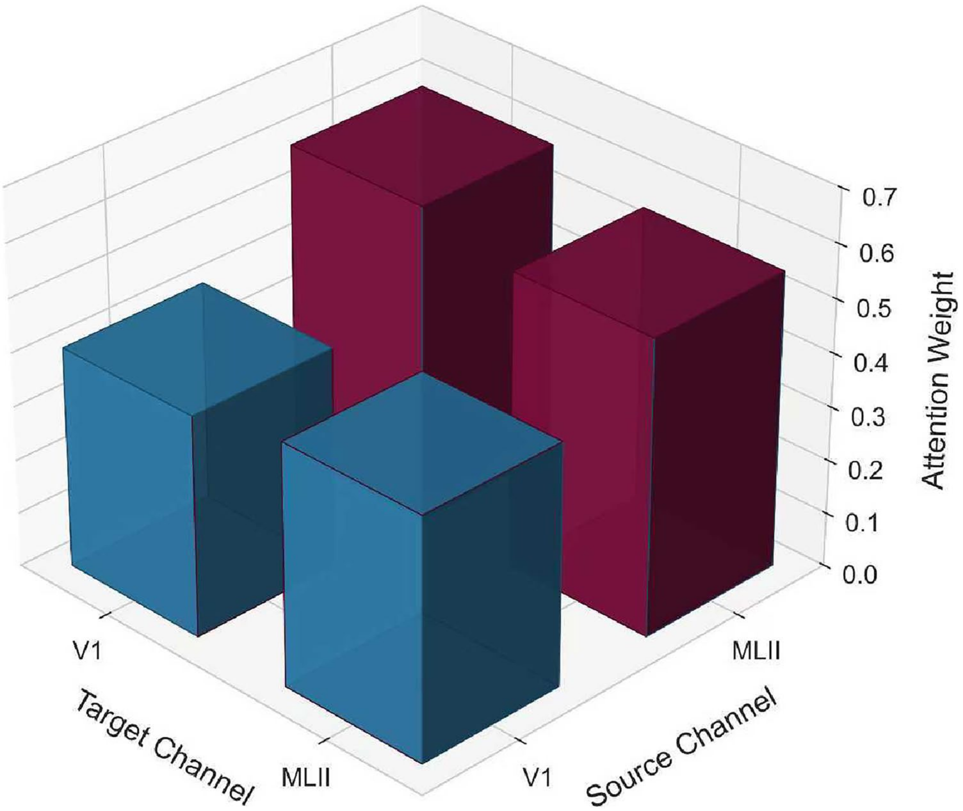


Fig. 3 3D GAT attention weight distribution

Table 1 SVDB dataset splitting strategy

Dataset	Number of patients	Number of heartbeat segments	Class distribution (NSR:VT:SVT)
Training set	38	124,916	7:2:1
Test set	10	31,229	7:2:1

Table 2 Key hyperparameters for the FHGNet model

Class	Parameter	Value
Architecture	Model dimension (d_{model})	128
	Transformer encoder layers	3
	Attention heads (n_{heads})	4
	GAT heads	2
	CNN kernel sizes	{3, 5, 7, 9}
	Pooling mode	Attention pooling
Training	Optimizer	AdamW
	Learning rate (LR)	0.0003
	Batch size	64
	Epochs	60
	LR scheduler	ReduceLROnPlateau
	Warmup epochs	10
Data & Preprocessing	Downsample factor	2
	Consecutive patches	8
	SMOTE oversampling	Enabled (on training set)

All test set results reported herein are the average of 5 independent runs with different random seeds to ensure statistical robustness. A summary of the final model configuration is provided in Table 2.

4.3 Implementation Details

All experiments were conducted on a server with a 13th Gen Intel(R) Core(TM) i5-13600K 3.50 GHz CPU and 1 NVIDIA GeForce RTX 4090 GPU. The source code and datasets will be made publicly available upon acceptance of this paper. Extensive experiments were conducted to optimize hyperparameters, including training epochs, batch size, and learning rate. Detailed results and analyses of these hyperparameter settings are presented in Sect. 5.1 (Ablation Experiments).

5 Results and Analysis

In this section, we present and discuss the experimental results and findings. First, ablation experiments were performed to investigate the role of core components in FHGNet, including the removal of the CNN branch, the GAT module, and the Transformer branch, thus validating the need for multimodal feature fusion.

Subsequently, comparative experiments were performed to evaluate FHGNet against traditional machine learning methods (SVM), classical deep learning models (1D/2D CNN), and Transformer-based architectures, demonstrating the effectiveness and superiority of FHGNet in terms of classification accuracy and generalization capability.

We also investigated the impact of sample balancing strategies on the SVDB test set, where Table 3 independently compares different imbalance handling methods. Compared with the unbalanced baseline (None, macro F1 = 0.8321), applying noise-based data augmentation alone (with_aug) provides a modest improvement (macro F1 increased by 3.44 percentage points, PR-AUC increased by 4.16 percentage points), indicating that augmentation enhances intra-class variability but has limited effect on shifting the decision boundary. Combining SMOTE with augmentation (with_smote_aug) further improves macro F1 to 0.8874. The best performance comes from using SMOTE alone (with_smote): accuracy 0.9232 and macro F1 0.9135 (an improvement of 8.14 percentage points over the baseline), along with the highest ROC-AUC (0.9492) and PR-AUC (0.9172).

This suggests that synthesizing minority class samples in feature space is the primary driver for improving recall and precision of rare classes (e.g., SVT), whereas combining with strong augmentation may introduce distributional noise that slightly offsets some of the gains.

5.1 Ablation Experiments

In this subsection, we conduct ablation experiments on the dataset to investigate the effect of each component of the FHGNet model on its final performance. The components in the study include the CNN branch, the GAT module, and the Transformer branch, the single-branch structure serving as a comparative configuration.

5.1.1 Ablation Study of the Multi-Branch Structure

We first explore the impact of the multi-branch structure on FHGNet. Specifically, we develop configurations with GAT and a two-stage classifier (GAT + Two-stage) as well as configurations with different branches removed.

As shown in Figs. 4 and 5, the GAT + Two-stage configuration achieves an overall accuracy of 89.25% and a macro F1 of 87.83%. Removing the CNN branch (Morph. CNN + Two-stage) leads to a decrease in macro F1 to 87.73%, and stage 2 precision (abnormal subtype classification, refer to Fig. 6) drops from 85.79% to 81.97%, highlighting the critical role of local morphological features in distinguishing abnormal subtypes.

Turning to stage 1 accuracy (normal/abnormal classification, refer to Fig. 6), the GAT + Two-stage configuration reaches 91.92%, surpassing the 89.92% of the Morph. CNN + Two-stage model. This demonstrates the advantage of the multi-branch structure in synergistically integrating diverse features during the initial screening of abnormal categories.

5.1.2 Ablation Study of the GAT Module

To isolate the impact of GAT, we compared the model (FHGNet) with a configuration that retains the CNN branch, Transformer branch, and two-stage classifier but removes the GAT module (denoted “Morph. CNN + Two-stage").

As shown in Table 4, the complete model achieves a macro F1 of 91.35%, while the GAT removed configuration drops to 87.73% (a decrease of 3.62 percentage points), verifying GAT’s critical role in modeling interlead spatial dependencies. The rejection rate also decreases from 42.33% to 34.66%, indicating reduced discriminability for low-confidence samples when inter-lead dynamics are not considered.

5.1.3 Ablation Study of the Single-Branch Structure

A single-branch structure retaining only the GAT branch (GAT only) or the CNN branch (Morph. CNN only) is designed and compared with the complete model. Its overall accuracy (87.72%, 87.37% respectively) and macro F1 (85.83%, 85.44%) are significantly lower than those of the complete model (as shown in Fig. 5), verifying the necessity of multimodal feature fusion.

The performance of the two-stage pure classifier (two-stage only) (macro F1 = 86.84%) is comparable to that of the single-branch model, suggesting that a simple classifier design is challenging to improve accuracy without incorporating deep feature extraction. The Transformer-only configuration achieves the lowest performance, with an overall accuracy of 85.24% and a macro F1 score of 83.21% (refer to Fig. 5), further demonstrating the synergistic contributions

Table 3 Effect of sample balancing strategies on SVDB (test set)

Experiment	Accuracy	Macro precision	Macro recall	Macro F1	Weighted F1	ROC-AUC	PR-AUC
with_aug	0.8824	0.8721	0.8650	0.8665	0.8799	0.9241	0.8741
with_smote	**0.9232**	**0.9174**	**0.9103**	**0.9135**	**0.9224**	**0.9492**	**0.9172**
with_smote_aug	0.8992	0.8919	0.8840	0.8874	0.8980	0.9334	0.8987
None	0.8524	0.8391	0.8279	0.8321	0.8494	0.8977	0.8325

All models share the same architecture and training schedule; only the balancing strategy differs. We mark the best one of each metric in bold

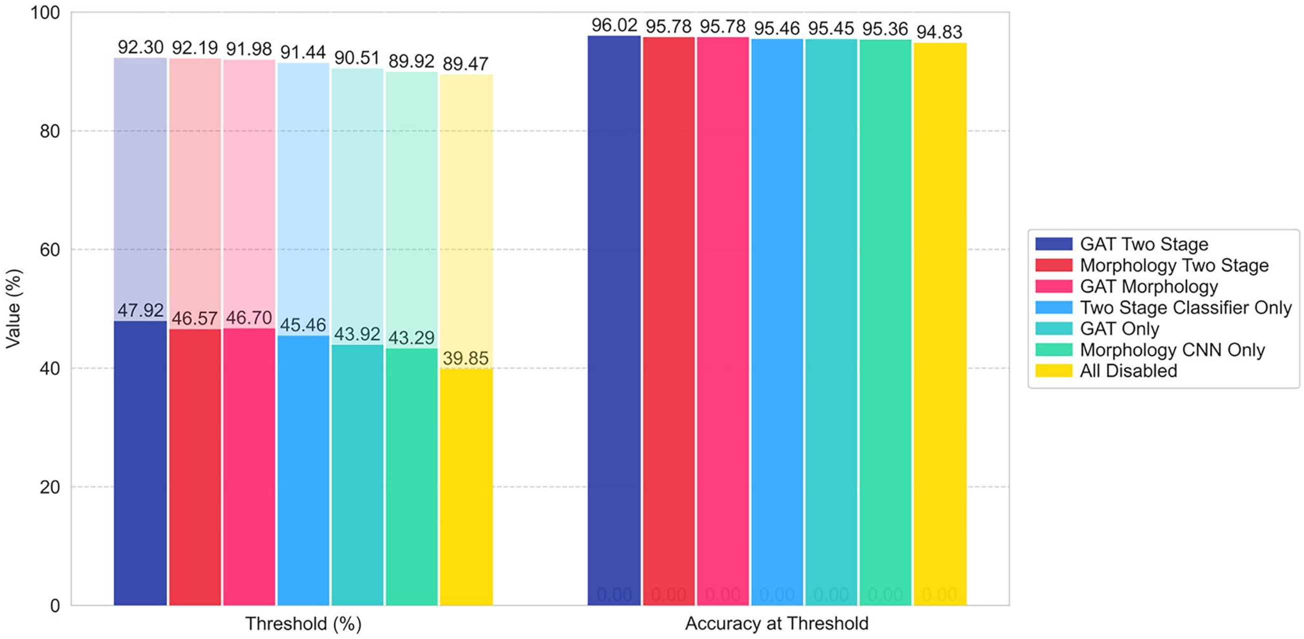


Fig. 4 Performance at recommendation threshold comparison

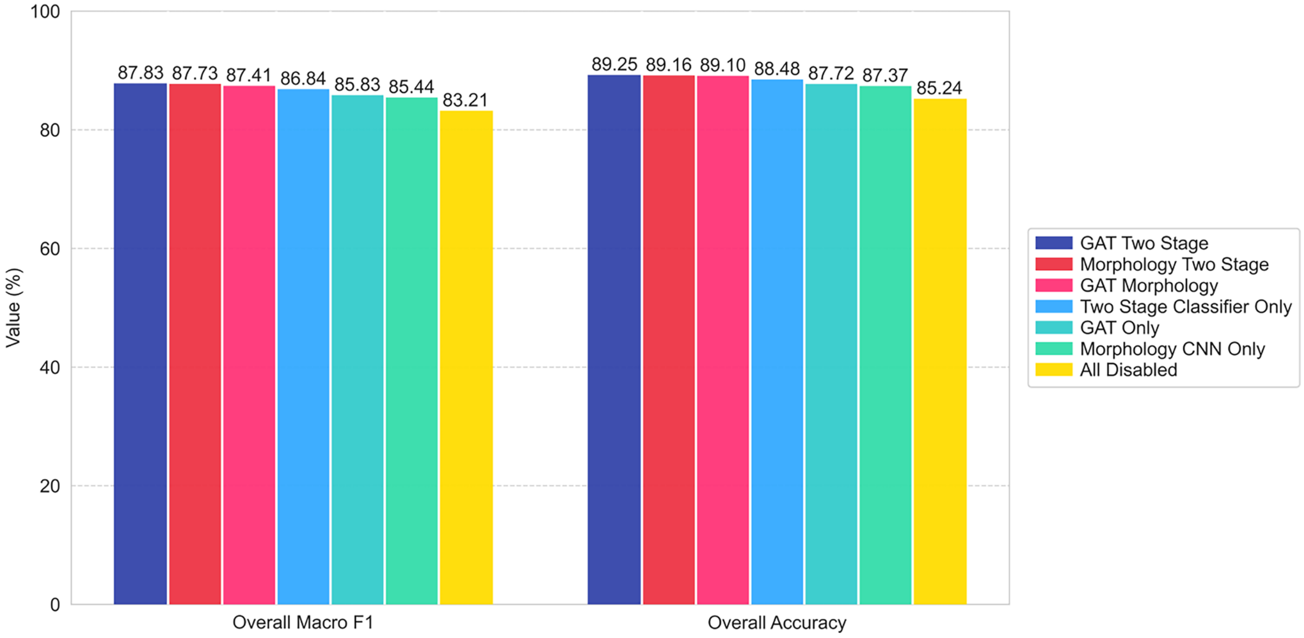


Fig. 5 Overall performance comparison

of each component. In summary, ablation experiments validate the necessity of multi-branch feature fusion and the two-stage classifier in FHGNet by comparing the performances of different module combinations.

As shown in Table 4, the whole model achieves 91.35% macro F1 and 92.32% overall accuracy through the synergy of CNN, GAT, and Transformer branches, significantly outperforming any single-branch configuration (e.g., GAT only: 85.83% macro F1; CNN only: 85.44% macro F1). Removing the GAT module (FHGNet without GAT) reduces the rejection rate from 42.33% to 25.67%, indicating its capability to model inter-lead spatial dependencies dynamically, thereby enhancing feature discriminability and reducing misjudgment of low-confidence samples.

The two-stage classifier design is critical for performance: retaining only the classifier (two-stage only) achieves 86.84% macro F1, lower than that of the entire model, highlighting that deep feature extraction modules (CNN/GAT/

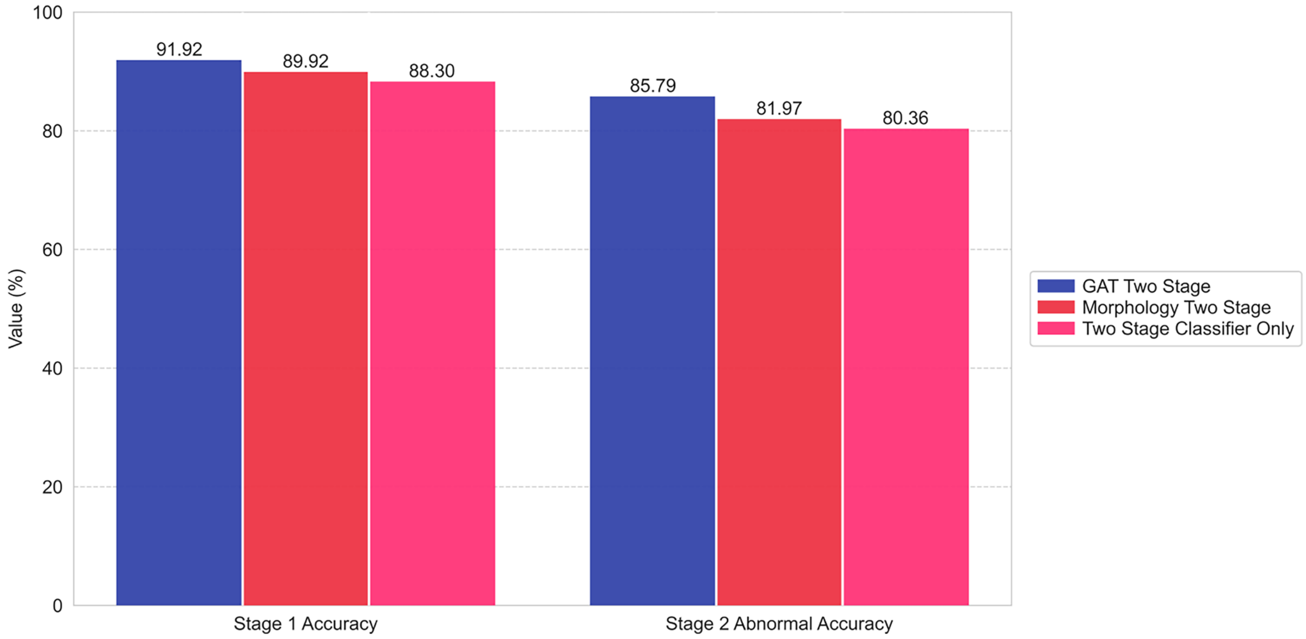


Fig. 6 Stage specific accuracy comparison

Table 4 Ablation study of proposed modules under different configurations

Configuration	Macro F1 (%)	RQR (%)	Acc@Thresh (%)	Overall Acc (%)	Stage 1 Acc (%)	Stage 2 Acc (%)
FHGNet(Full Model)	91.35	93.37	96.17	92.32	94.04	89.75
GAT + Two-stage	87.83	92.30	96.02	89.25	91.92	85.79
Morph. CNN + Two-stage	92.12	34.66	95.78	89.16	89.92	81.97
GAT + Morph. CNN	87.41	91.98	95.46	89.10	N/A	N/A
Two-stage only	86.84	91.44	95.45	88.48	88.30	80.36
GAT only	85.83	28.59	90.51	87.72	N/A	N/A
Morph. CNN only	85.44	89.92	95.36	87.37	N/A	N/A
Transformer only	83.21	89.47	94.83	85.24	N/A	N/A

Note: In implementation, we define $\mathrm{RQR} = \frac{n_{\mathrm{TR}}}{n_{\mathrm{TR}}+n_{\mathrm{FR}}}$, i.e., the proportion of truly "should-be-rejected" (incorrect) samples (n_{TR}) among all rejected samples ($n_{\mathrm{TR}} + n_{\mathrm{FR}}$), which measures the "purity" of the rejection strategy. A higher RQR indicates that the rejection set is more concentrated on high-risk/low-confidence erroneous samples. It can be regarded as the "precision" of the rejection set under the definition of "error" as the positive class

Transformer) are prerequisites for classifier effectiveness. Additionally, the "All modules off" configuration yields the lowest performance (83.21% macro F1), further confirming the irreplaceability of each component.

These results demonstrate that FHGNet's high performance stems from the complementarity of multi-branch features (morphological + rhythmic + spatial dependencies) and the hierarchical discriminative ability of the two-stage architecture.

5.2 Confidence-Based Rejection Analysis

To further analyze the trade-off between accuracy and rejection rate, we conducted a threshold sweep over confidence scores ranging from 0.5 to 0.99. Figure 7 illustrates the trade-off curve: as the rejection rate increases, the accuracy of retained samples steadily improves, achieving 98.7% accuracy when approximately 45% of uncertain samples are rejected. This result indicates that the rejection mechanism allows clinicians to adjust the balance between automation and manual review based on clinical needs. In practice, hospitals with higher tolerance for manual review can set lower thresholds (favoring higher recall), while resource-constrained settings may set higher thresholds to maximize precision. This flexibility enhances the clinical applicability of FHGNet.

5.3 Performance Comparison with State-of-the-art approaches

This study compares the proposed FHGNet model with traditional machine learning methods (SVM) [42], single-modality deep learning models (1D/2D CNN) [43, 44], and the Transformer-based model [45] to validate its

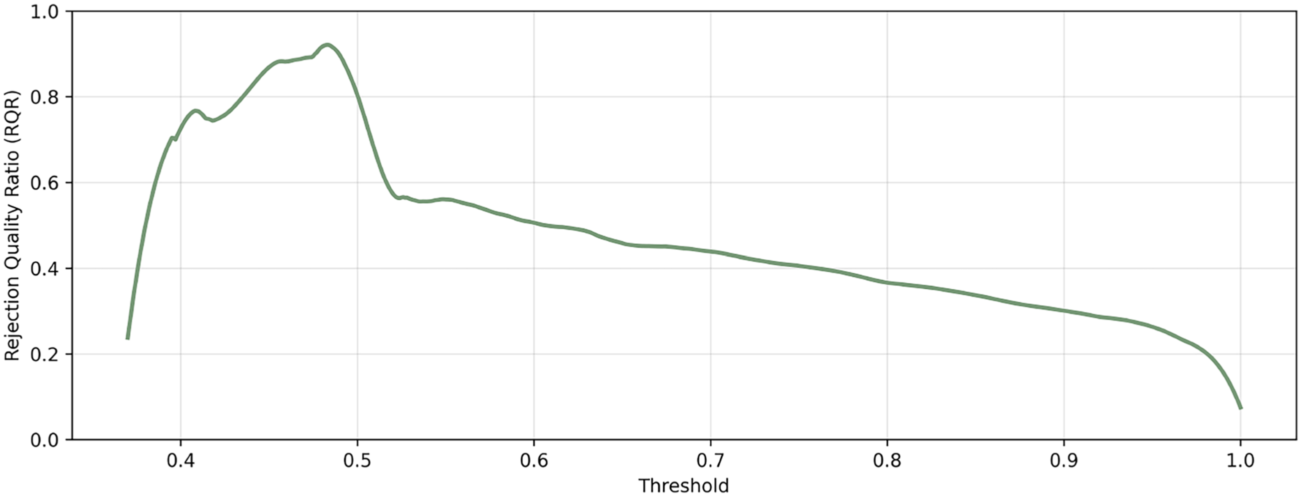


Fig. 7 Trade-off between accuracy and rejection rate under different confidence thresholds

Table 5 Comparison of performance metrics with baseline models

Model	Accuracy	Macro precision	Macro recall	Macro F1	Weighted F1
Ours	0.9232	0.9174	0.9103	0.9135	0.9224
IM-ECG	0.8993	0.7968	0.8297	0.7129	0.8617
ECGMamba	0.8931	0.7735	0.8165	0.6893	0.8384
Transformer-based model	0.7960	0.4990	0.7634	0.5536	0.8594
SVM	0.6528	0.4111	0.5676	0.4023	0.7412
1D CNN	0.8358	0.4792	0.5961	0.5111	0.8675
2D CNN	0.8625	0.5211	0.6328	0.5716	0.8679

Table 6 Classification report for high-confidence samples

Metric/Class	Precision	Recall	F1	Support
Class N	0.96	0.98	0.97	7309
Class S	0.93	0.86	0.89	2515
Class V	0.98	0.99	0.98	4326
Accuracy	N/A	N/A	0.96	14150
Macro average	0.95	0.94	0.95	14150
Weighted average	0.96	0.96	0.96	14150

performance in the classification of arrhythmias. In addition, we include two recent methods as baselines: IM-ECG, which leverages dual-kernel residual blocks and Grad-CAM for interpretable multi-lead classification, and ECGMamba, which applies a bidirectional state space model for efficient long-range temporal modeling. The analysis is based on ablation experiment data, focusing on overall classification accuracy, multi-dimensional feature effectiveness, and temporal modeling capabilities.

Key metrics are compared to evaluate FHGNet against traditional and single-modality models (Table 5): FHGNet achieves an accuracy of 92.32% and a macro F1 of 91.35%, significantly outperforming the Transformer-based model (accuracy 79.60%, macro F1 55.36%), IM-ECG (accuracy 80.93%, macro F1 71.29%), ECGMamba (accuracy 80.31%, macro F1 68.93%), SVM (macro F1 40.23%), and 1D/2D CNN (maximum macro F1 57.16%). Traditional methods (SVM) suffer from poor generalization in complex arrhythmias (e.g., VT with morphological variation) due to reliance on manual feature design.

Single-modality CNNs (1D CNN, 2D CNN) fail to model multi-lead spatial correlations, thereby limiting their performance. FHGNet's weighted F1 score (92.24%) is close to its overall accuracy, indicating balanced reliability across classes, especially for minority classes (SVT).

After screening low-confidence samples through the rejection mechanism, the high-confidence data classification report shows (see Table 6): the F1 scores of FHGNet for normal (N), supraventricular (S) and ventricular (V) classes reach 97%, 89% and 98%, respectively, among which the recall rate of the ventricular class (V) is as high as 99%, indicating the model's outstanding sensitivity to high-risk arrhythmias. Compared with the baseline model (see "Transformer-based model" in Table 5), FHGNet's macro average index (95% vs. 55%) has significantly improved, verifying the discriminative ability of multi-branch feature fusion for complex samples.

In addition to macro F1, we further report ROC-AUC and PR-AUC across three datasets (MIT-BIH-AR, SVDB, and LUDB [46]), as summarized in Table 7. The results show that FHGNet achieves consistently high performance across all metrics. On the MIT-BIH-AR dataset, the model attains an accuracy of 98.79% with macro F1 of 98.70%, and ROC-AUC/PR-AUC above 0.99, demonstrating excellent diagnostic reliability in a standard benchmark. On the SVDB dataset, where minority classes such as SVT are challenging, FHGNet maintains a macro F1 of 91.35% and PR-AUC of 0.9172, highlighting its ability to reduce false negatives while preserving precision. On the LUDB dataset, although

Table 7 Comprehensive evaluation metrics of FHGNet across three datasets

Dataset	Accuracy	Macro precision	Macro recall	Macro F1	Weighted F1	ROC-AUC/PR-AUC
MIT-BIH-AR	0.9879	0.9862	0.9879	0.9870	0.9880	0.9923/0.9891
SVDB	0.9232	0.9174	0.9103	0.9135	0.9224	0.9492/0.9172
LUDB	0.8578	0.8854	0.8333	0.8452	0.8527	0.8389/0.8112

Table 8 Cross-dataset comparison

Dataset	Accuracy	Macro precision	Macro recall	Macro F1	Weighted F1
MIT-BIH-AR	0.9676	0.9699	0.9716	0.9708	0.9717
LUDB	0.8041	0.8410	0.7916	0.8028	0.8102

the overall performance is lower due to severe class imbalance and its primary use as a waveform delineation benchmark, FHGNet still achieves a macro F1 of 84.52% with ROC-AUC/PR-AUC above 0.81, demonstrating robustness under distributional shifts.

5.4 Cross-Database Validation

To further assess the generalizability of FHGNet, we conducted cross-database validation experiments. Specifically, the model was trained on the SVDB dataset and directly evaluated on two external datasets without retraining: LUDB and MIT-BIH-AR. The results are presented in Table 8. Note that the LUDB dataset is primarily intended for ECG waveform delineation with highly imbalanced rhythm/pathology categories; therefore, in this work LUDB is used to validate physiological event detection (R-peak and beat boundary delineation) rather than as a multi-class classification benchmark.

FHGNet achieved an accuracy of 85.78% and a macro F1-score of 0.8452 on LUDB, and 84.62% accuracy with a macro F1-score of 0.8317 on MIT-BIH-AR. These results demonstrate that FHGNet maintains strong performance across datasets with varying patient demographics, recording devices, and noise conditions, highlighting its robustness to domain shift.

6 Conclusion

This study introduces FHGNet, a Transformer-based framework for VT/SVT classification, which integrates multi-lead spatio-temporal features and rhythmic dynamics. Specifically, a CNN is employed for extracting morphological features, while a Transformer models interbeat rhythmic patterns. Additionally, a GAT dynamically fuses multilead spatial dependencies. By leveraging a Transformer encoder, the framework captures long-range temporal correlations of oscillatory features. The classification process is executed through a two-stage approach, encompassing anomaly screening and subtype identification.

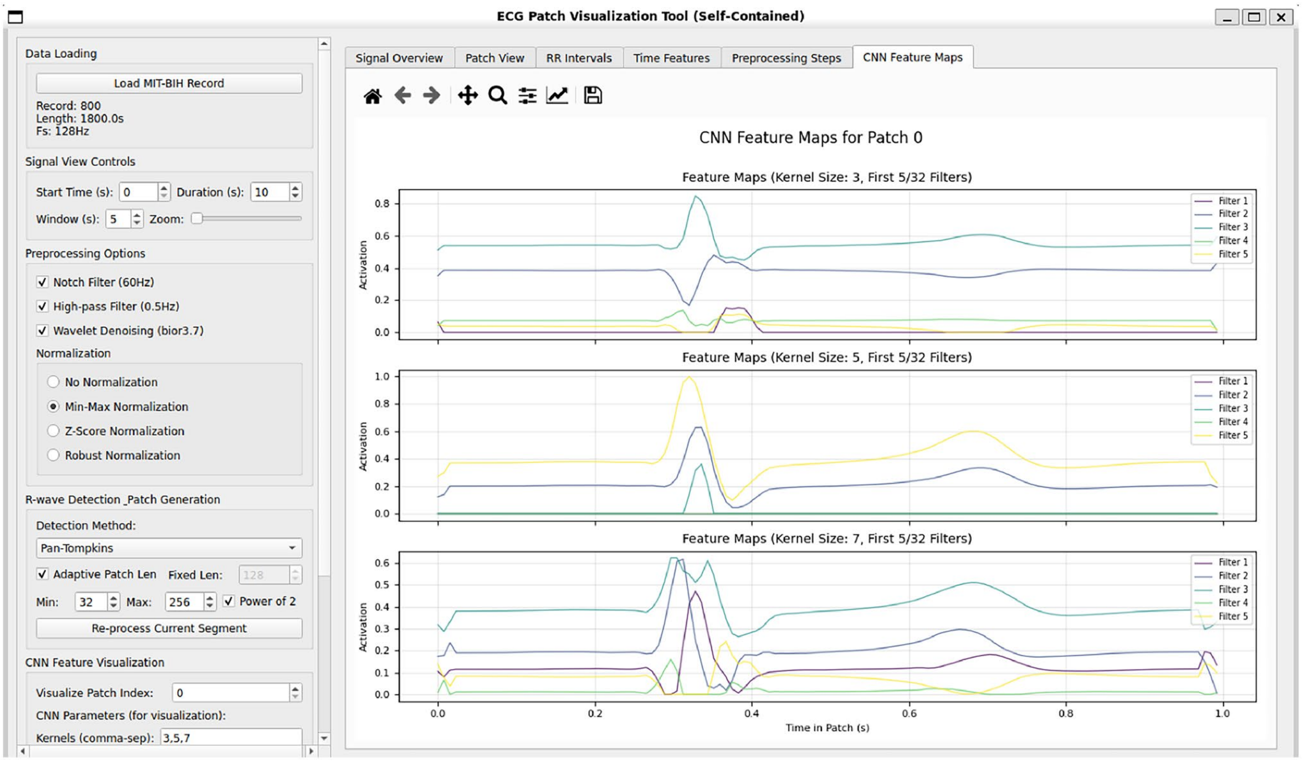
Attention visualization highlights the model's focus on QRS morphological variations and ST-segment energy distributions, which aligns with clinical diagnostic criteria. Across multi-center datasets, FHGNet achieves an overall accuracy of 92.32%, significantly outperforming traditional signal processing methods (e.g., STFT+SVM, 82.15%) and single-modality deep learning models (e.g., CNN, 87.37%).

Compared with IM-ECG and ECGMamba, FHGNet achieves higher macro F1 and recall for the minority SVT class, demonstrating its effectiveness in handling class imbalance and capturing dynamic multi-lead dependencies. Moreover, while ECGMamba is computationally efficient, it lacks explicit clinical interpretability, which is addressed in FHGNet through attention visualization and graph-based lead interaction modeling.

It demonstrates enhanced noise robustness in noise-related conditions (accuracy: 88.48%). Ablation experiments validate the necessity of multi-branch feature fusion, showing that the GAT module improves the rejection rate by 9.49% through spatial feature modeling, while Transformer enhances subtype classification accuracy (Stage 2) to 85.79% by capturing temporal dynamics.

Attention visualization aligns with clinical diagnostic logic, focusing on QRS complex morphological variations and ST-segment oscillation patterns, thereby demonstrating interpretability. While current attention visualization results confirm that the model emphasizes clinically relevant features such as ST–T segment energy ratios and inter-lead phase differences, this approach is essentially static and limited to heatmap-like representations. Such visualizations, although informative, may not fully convey the decision-making process to clinicians, particularly in complex multi-lead interactions.

To further improve interpretability, future work could integrate more sophisticated visualization and interactive techniques. For example, lead-wise dynamic saliency maps or temporal attention trajectories could highlight how the model's focus evolves across different cardiac cycles, while counterfactual explanations (e.g., "how would the prediction change if QRS width were reduced by 20 ms") could make the decision boundaries more transparent. Additionally, developing clinician-in-the-loop systems that allow doctors to adjust attention weights or explore model rationale interactively may bridge the gap between black-box AI and practical diagnostic reasoning. Such tools will not only improve physician trust in automated systems but also


ECG Patch Visualization Tool (Self-Contained)
Data Loading
Load MIT-BIH Record
Signal View Controls
Preprocessing Options
Normalization
R-wave Detection_Patch Generation
Re-process Current Segment
CNN Feature Visualization
Signal Overview
Patch View
RR Intervals
Time Features
Preprocessing Steps
CNN Feature Maps
CNN Feature Maps for Patch 0
Feature Maps (Kernel Size: 3, First 5/32 Filters)
Feature Maps (Kernel Size: 5, First 5/32 Filters)
Feature Maps (Kernel Size: 7, First 5/32 Filters)
Activation
Filter 1
Filter 2
Filter 3
Filter 4
Filter 5
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.8
1.0
Time in Patch (s)

Record: 800
Length: 1800.0s
Fs: 128Hz

Start Time (s): 0 Duration (s): 10
Window (s): 5 Zoom:

☑ Notch Filter (60Hz)
☑ High-pass Filter (0.5Hz)
☑ Wavelet Denoising (bior3.7)

○ No Normalization
● Min-Max Normalization
○ Z-Score Normalization
○ Robust Normalization

Detection Method: Pan-Tompkins
☑ Adaptive Patch Len Fixed Len: 128
Min: 32 Max: 256 ☑ Power of 2

Visualize Patch Index: 0
CNN Parameters (for visualization):
Kernels (comma-sep): 3,5,7

Fig. 8 Prototype of our interactive ECG interpretability tool. The interface allows clinicians to load MIT-BIH records and apply preprocessing (e.g., filtering, normalization, wavelet denoising), segment ECG signals via Pan-Tompkins R-peak detection with adaptive patch length, and visualize CNN feature maps with different kernel sizes. By enabling real-time inspection of signal transformations and model features, this tool provides a more transparent and clinically actionable interpretability framework

create opportunities for collaborative human–AI decision-making in arrhythmia diagnosis. While FHGNet performs effectively with fixed-length inputs, its limitation in handling variable input sizes motivates future research directions, including the development of dynamic input support, optimization of real-time computational efficiency, and exploration of multimodal fusion with clinical metadata to enhance generalization and clinical applicability.

Future work will explore the integration of clinical metadata (such as age, gender, comorbidities, and medical history) and imaging data (such as echocardiography and cardiac MRI) with ECG signals to improve the model's performance and clinical applicability. The inclusion of these additional modalities will provide a more comprehensive understanding of cardiac conditions, improve detection accuracy, and enhance the robustness of FHGNet in diverse clinical settings.

In addition to multimodal fusion, another important future direction is to develop more advanced interpretability tools. Current attention visualization provides static heatmaps that highlight relevant leads or intervals, but lacks dynamic and interactive features. To go beyond static attention maps, we have developed a prototype interactive visualization interface (Fig. 8). The tool integrates preprocessing configuration, R-peak-based patch generation, and CNN feature map inspection in a unified interface, enabling clinicians to dynamically explore how signal transformations influence model reasoning. Unlike static saliency visualizations, this system allows real-time validation of model decisions and has the potential to become a practical tool for clinical ECG interpretation.

In future work, we plan to design more sophisticated visualization techniques, such as time-varying saliency trajectories and counterfactual explanations (e.g., how predictions change if QRS width shortens), as well as interactive interfaces that allow clinicians to intuitively explore model reasoning. These improvements will further bridge the gap between algorithmic decisions and clinical diagnostic practice.

Acknowledgements This work is partially supported by the National Key R&D Plan of China (No. 2024YFE0202700) and the National Natural Science Foundation of China (No. 61872263).

Code Availability The code supporting the findings of this study is available at https://github.com/Cooing-code/MORTIS.

References

1. Smital L, Haider CR, Vitek M et al (2020) Real-time quality assessment of long-term ECG signals recorded by wearables in free-living conditions. IEEE Trans Biomed Eng 67(10):2721–2734. https://doi.org/10.1109/TBME.2020.2969719
2. Jiang Z, Almeida TP, Schlindwein FS et al (2020) Diagnostic of multiple cardiac disorders from 12-lead ECGs using graph convolutional network based multi-label classification. In: Computing in Cardiology , pp 1–4. https://doi.org/10.22489/CinC.2020.135
3. Kotadia ID, Williams SE, O'Neill M (2020) Supraventricular tachycardia: an overview of diagnosis and management. Clin Med (Lond) 20(1):43–47. https://doi.org/10.7861/clinmed.cme.20.1.3
4. Dotsinsky I, Clifford GD, Azuaje F et al (2007) Advanced methods and tools for ECG analysis. Biomed Eng Online 6:18. https://doi.org/10.1186/1475-925X-6-18
5. Chen PC, Lee S, Kuo CD (2006) Delineation of T-wave in ECG by wavelet transform using multiscale differential operator. IEEE Trans Biomed Eng 53(7):1429–1433. https://doi.org/10.1109/TBME.2006.875719
6. Starobin OE (2002) Book review: *Chou's electrocardiography in clinical practice: adult and pediatric*, 5th edition by Borys Surawicz and Timothy K. Knilans WB Saunders, 2001. J Intensive Care Med 17(4):204–204. https://doi.org/10.1177/0885066602017004011
7. Al-Khatib SM, Stevenson WG, Ackerman MJ et al (2018) 2017 AHA/ACC/HRS guideline for management of patients with ventricular arrhythmias and the prevention of sudden cardiac death. J Am Coll Cardiol 72(14):e91–e220. https://doi.org/10.1016/j.jacc.2017.10.054
8. Wellens HJJ (2001) Ventricular tachycardia: diagnosis of broad QRS complex tachycardia. Heart 86(5):579–585. https://doi.org/10.1136/heart.86.5.579
9. Ruttimann UE, Pipberger HV (1979) Compression of the ECG by prediction or interpolation and entropy encoding. IEEE Trans Biomed Eng 26(11):613–623. https://doi.org/10.1109/TBME.1979.326430
10. Hou B, Yang J, Wang P et al (2020) LSTM-based auto-encoder model for ECG arrhythmias classification. IEEE Trans Instrum Meas 69(4):1232–1240. https://doi.org/10.1109/TIM.2019.2910342
11. Tao R, Wang L, Xiong Y et al (2024) IM-ECG: an interpretable framework for arrhythmia detection using multi-lead ECG. Expert Syst Appl 237:121497. https://doi.org/10.1016/j.eswa.2023.121497
12. Daponte P, De Vito L, Iadarola G et al (2021) ECG monitoring based on dynamic compressed sensing of multi-lead signals. Sensors 21(21):7003. https://doi.org/10.3390/s21217003
13. Uğraş BK, Gerek ÖN, Saygı İT (2025) CardioPatternFormer: pattern-guided attention for interpretable ECG classification with transformer architecture. arXiv. https://doi.org/10.48550/arXiv.2505.20481
14. Cheng R, Zhuang Z, Zhuang S et al (2023) MSW-Transformer: multi-scale shifted windows transformer networks for 12-lead ECG classification. arXiv. https://doi.org/10.48550/arXiv.2306.12098
15. Moody GB, Mark RG (2001) The impact of the MIT-BIH arrhythmia database. IEEE Eng Med Biol Mag 20(3):45–50. https://doi.org/10.1109/51.932724
16. de Lannoy G, Frénay B, Verleysen M (2009) Supervised ECG delineation using the wavelet transform and hidden Markov models. In: 4th European Conferenceof the International Federation for Medical and Biological Engineering, pp 22–25. https://doi.org/10.1007/978-3-540-89208-3_7
17. Graja S, Boucher JM (2005) Hidden markov tree model applied to ECG delineation. IEEE Trans Instrum Meas 54(6):2163–2168. https://doi.org/10.1109/TIM.2005.858568
18. Silipo R, Marchesi C (1998) Artificial neural networks for automatic ECG analysis. IEEE Trans Signal Process 46(5):1417–1425. https://doi.org/10.1109/78.668803
19. Venkatesan C, Karthigaikumar P, Varatharajan R (2018) A novel LMS algorithm for ECG signal preprocessing and KNN classifier based abnormality detection. Multimed Tools Appl 77:10365–10374. https://doi.org/10.1007/s11042-018-5762-6
20. Saadatnejad S, Oveisi M, Hashemi M (2020) LSTM-based ECG classification for continuous monitoring on personal wearable devices. IEEE J Biomed Health Inform 24(2):515–523. https://doi.org/10.1109/JBHI.2019.2911367
21. Tran TD, Tran NQ, Dang TTK et al (2024) ECG captioning with prior-knowledge transformer and diffusion probabilistic model. J Healthc Inform Res. https://doi.org/10.1007/s41666-024-00176-3
22. Martínez A, Alcaraz R, Rieta JJ (2010) A new method for automatic delineation of ECG fiducial points based on the phasor transform. In: 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology, pp 4586–4589. https://doi.org/10.1109/IEMBS.2010.5626498
23. Rincón F, Recas J, Khaled N et al (2011) Development and evaluation of multilead wavelet-based ECG delineation algorithms for embedded wireless sensor nodes. IEEE Trans Inf Technol Biomed 15(6):854–863. https://doi.org/10.1109/TITB.2011.2163943
24. Oster J, Behar J, Sayadi O et al (2015) Semisupervised ECG ventricular beat classification with novelty detection based on switching Kalman filters. IEEE Trans Biomed Eng 62(9):2125–2134. https://doi.org/10.1109/TBME.2015.2402236
25. Vijaya G, Kumar V, Verma HK (1998) ANN-based QRS-complex analysis of ECG. J Med Eng Technol 22(4):160–167. https://doi.org/10.3109/03091909809032534
26. Li H, Yuan D, Ma X et al (2017) Genetic algorithm for the optimization of features and neural networks in ECG signals classification. Sci Rep 7(1):41011. https://doi.org/10.1038/srep41011
27. Sameni R, Shamsollahi MB, Jutten C et al (2007) A nonlinear Bayesian filtering framework for ECG denoising. IEEE Trans Biomed Eng 54(12):2172–2185. https://doi.org/10.1109/TBME.2007.897817
28. Saini I, Singh D, Khosla A (2013) QRS detection using K-nearest neighbor algorithm (KNN) and evaluation on standard ECG databases. J Adv Res 4(4):331–344. https://doi.org/10.1016/j.jare.2012.05.007
29. Agrawal P, Arun V, Basu A (2025) Artificial neural network based ECG feature extraction using wavelet transform. In: Emerging Wireless Technologies and Sciences, pp 8–22. https://doi.org/10.1007/978-3-031-87886-2_2
30. Nowostawski M, Poli R (1999) Parallel genetic algorithm taxonomy. In: 1999 Third International Conference on Knowledge-Based Intelligent Information Engineering Systems. Proceedings, pp 88–92. https://doi.org/10.1109/KES.1999.820127
31. Wei L, Li Y (2025) A multi-scale CNN-Transformer parallel network for 12-lead ECG signal classification. Signal Image Video Process 19(8):611. https://doi.org/10.1007/s11760-025-04215-3
32. Li L, Camps J, Wang ZJ et al (2024) Toward enabling cardiac digital twins of myocardial infarction using deep computational models for inverse inference. IEEE Trans Med Imaging 43(7):2466–2478. https://doi.org/10.1109/TMI.2024.3370631
33. Abrishami H, Han C, Zhou X et al (2018) Supervised ECG interval segmentation using LSTM neural network. In: Proceedings of the International Conference on Bioinformatics & Computational

Biology (BIOCOMP), pp 71–77. https://www.proquest.com/conference-papers-proceedings/supervised-ecg-interval-segmentation-using-lstm/docview/2139457108/se-2

34. Semwal A, Londhe ND (2021) Computer aided pain detection and intensity estimation using compact CNN-based fusion network. Appl Soft Comput 112:107780. https://doi.org/10.1016/j.asoc.2021.107780
35. Nurmaini S, Darmawahyuni A, Rachmatullah MN et al (2021) Beat-to-beat electrocardiogram waveform classification based on a stacked convolutional and bidirectional long short-term memory. IEEE Access 9:92600–92613. https://doi.org/10.1109/ACCESS.2021.3092631
36. Peimankar A, Puthusserypady S (2021) DENS-ECG: a deep learning approach for ECG signal delineation. Expert Syst Appl 165:113911. https://doi.org/10.1016/j.eswa.2020.113911
37. Wang J, Li R, Li R et al (2020) A knowledge-based deep learning method for ECG signal delineation. Future Gener Comput Syst 109:56–66. https://doi.org/10.1016/j.future.2020.02.068
38. Qiang Y, Dong X, Liu X et al (2024) ECGMamba: towards efficient ECG classification with BiSSM. arXiv. https://doi.org/10.48550/arXiv.2406.10098
39. Long X, Yan B, Mo Z (2023) Uncovering the heterogeneity and cell fate decisions of endothelial cells after myocardial infarction by single-cell sequencing. Med Adv 1(3):234–245. https://doi.org/10.1002/med4.34
40. Li H, Huang Y, Luo D et al (2024) Optimizing agitated saline volume for contrast echocardiography: balancing diagnostic performance and operator fatigue. Med Adv 2(3):254–261. https://doi.org/10.1002/med4.73
41. Pan J, Tompkins WJ (1985) A real-time QRS detection algorithm. IEEE Trans Biomed Eng BME 32(3):230–236
42. Venkatesan C, Karthigaikumar P, Paul A et al (2018) ECG signal preprocessing and SVM classifier-based abnormality detection in remote healthcare applications. IEEE Access 6:9767–9773. https://doi.org/10.1109/ACCESS.2018.2794346
43. Ahmed AA, Ali W, Abdullah TAA et al (2023) Classifying cardiac arrhythmia from ECG signal using 1D CNN deep learning model. Mathematics 11(3):562. https://doi.org/10.3390/math11030562
44. Mewada H (2023) 2D-wavelet encoded deep CNN for image-based ECG classification. Multimed Tools Appl 82(13):20553–20569. https://doi.org/10.1007/s11042-022-14302-z
45. Yan G, Liang S, Zhang Y et al (2019) Fusing transformer model with temporal features for ECG heartbeat classification. In: 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp 898–905. https://doi.org/10.1109/BIBM47256.2019.8983326
46. Kalyakulina AI, Yusipov II, Moskalenko VA et al (2020) LUDB: a new open-access validation tool for electrocardiogram delineation algorithms. IEEE Access 8:186181–186190. https://doi.org/10.1109/ACCESS.2020.3029211