STELLAR: Pacemaker Recognition Using 12-Lead ECG and Spatio-Temporal Harmonic Mechanism

Han Zhang[1†], Zeyuan Ding[1†], Leping Yang[1], Yu Lu[1], Jiatong Ding[1], Dian Ding[1*], Yiding Qi[2], Ruogu Li[2], Guanghui Gao[3] Yi-Chao Chen[1], Guangtao Xue[1]

[1] Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China

[2] Shanghai Chest Hostpital, Shanghai Jiao Tong University, China

[3] Department of Oncology, Shanghai Pulmonary Hospital & Thoracic Cancer Institute, Tongji University School of Medicine, China

Email:{ han_zhang, boris1013, y.jerry, yulu01, jerrysfls, dingdian94}@sjtu.edu.cn, zeeding@alumni.sjtu.edu.cn, lrg@sjtu.edu.cn, ghgao103@tongji.edu.cn, {yichao, gt_xue}@sjtu.edu.cn

***Abstract*—As cardiovascular diseases and arrhythmias rise globally, pacemakers have become a critical therapeutic option for managing cardiac rhythm disorders. Accurate identification of pacemaker implantation sites is essential for personalized pacing therapy and optimal clinical outcomes. While 12-lead electrocardiogram (ECG) signals provide a non-invasive means to infer implantation locations, they are susceptible to noise and morphological variability, posing challenges for high-accuracy localization. To advance data-driven solutions in this domain, we present PILDE, the first publicly available dataset specifically designed for pacemaker implantation site identification, comprising 12-lead ECG recordings from 733 patients across four distinct implantation locations. Based on this dataset, we propose STELLAR, a novel deep learning framework that integrates a Spatio-Temporal Lead-Harmonic Mechanism to model both the temporal dynamics of ECG waveforms and the spatial coherence across leads. Extensive experiments demonstrate that STELLAR outperforms conventional deep models—including CNN, LSTM, and Transformer baselines—on both the PILDE and PTB-XL datasets. Specifically, STELLAR achieves an average accuracy improvement of 10.45% on PILDE and 14.19% on PTB-XL, with significant gains in sensitivity and F1-score for minority classes. These results highlight the robustness and precision of STELLAR in automating implantation site identification, offering a promising tool for pre-procedural planning and clinical decision support. The source code and dataset access information will be made publicly available.**

***Index Terms*—Pacemaker Recognition, 12-lead ECG, Spatio-Temporal Lead-Harmonic Mechanism**

I. Introduction

Cardiac rhythm disorders, particularly those involving dysfunction of the cardiac conduction system—such as bradycardia, atrioventricular conduction block, and atrial fibrillation—are a growing global public health concern. According to the World Health Organization (WHO), cardiovascular diseases remain the leading cause of death worldwide, responsible for over 17 million deaths annually [1], with a significant proportion linked to arrhythmias like bradycardia and atrial fibrillation. To address these disorders, cardiac pacemakers have become a vital therapeutic tool. These small implantable devices deliver electrical impulses to regulate heart rate and rhythm, maintaining normal cardiac function. Over 10 million people worldwide have received pacemaker implants to date [2], [3], underscoring their critical role in improving quality of life and extending survival.

Fig. 1. STELLAR enables convenient and low-cost pacemaker recognition based on 12-lead ECG signals.

Pacemakers can be classified into four types based on the implantation location and the heart area they stimulate [4]. **Biventricular Pacing (BP)** stimulates both the left and right ventricles simultaneously and is suitable for heart failure patients, helping to improve the heart's pumping function. **Right Ventricular Septal Pacing (RVSP)** stimulates the right ventricular septum and is commonly used for the right bundle branch block, improving the coordination of heart contractions. **Right Ventricular Apex Pacing (RVAP)** is a common type of single-chamber pacing, suitable for conditions like bradycardia, and works by stimulating the apex of the right ventricle. **Left Bundle Branch Pacing (LBBP)** stimulates the left bundle branch and is used for bundle branch block, improving the heart's conduction function.

Accurately identifying the type of pacemaker required is

† Both authors contributed equally to the research.

* Dian Ding is Corresponding author.

crucial for formulating an appropriate surgical plan, as different pacing modalities involve distinct procedural approaches, technical demands, and levels of complexity. For instance, right ventricular apical or septal pacing (RVAP/RVSP) involves relatively simple implantation techniques suitable for routine procedures, whereas left bundle branch pacing (LBBP) requires precise lead placement guided by intraoperative electrophysiological monitoring, making it technically more demanding. Biventricular pacing (BP) entails the implantation of multiple leads and involves a more complex surgical pathway with higher procedural risks. Therefore, preoperative determination of the pacing strategy directly influences surgical duration, safety considerations, operator expertise requirements, equipment preparation, and perioperative management.

Identifying a pacemaker can be done in a number of ways, including scans, electrocardiograms (ECG), and X-rays. The following are specific applications of these three methods. **Programming Device Scan**: A specialized programming device can communicate directly with the pacemaker to read its internal information, such as make, model, serial number, and specific setup parameters [5]. **X-ray**: X-rays can show the location of the pacemaker device and the layout of the wires in the body [6]. **Electrocardiogram (ECG)**: The ECG shows patterns of electrical activity in the heart, which can be used to deduce the operating status of the pacemaker [7].

However, Programming Device Scan requires specialized equipment and medical personnel, making it difficult to implement in resource-limited healthcare settings. X-ray detection is costly and involves ionizing radiation, which carries certain safety risks. Additionally, identifying pacemaker types based on electrocardiograms (ECGs) is challenging due to the subtle differences in ECG features among various pacing modes, which are easily influenced by lead position, individual anatomical variations, and conduction status. More importantly, the current lack of large-scale, well-annotated public datasets limits the training and validation of algorithmic models, further increasing the difficulty of automatic classification.

With the development of deep learning and other technologies, researchers have successfully implemented ECG-based atrial fibrillation prediction [8], as well as early diagnosis and monitoring of heart diseases such as myocardial infarction [9], heart failure [10], and QT interval abnormalities [11]. We aim to use deep learning models to classify pacemaker implantation locations based on ECG, as different implantation locations can alter the propagation of cardiac electrical activity, thereby affecting the shape and duration of the QRS complex [12]. Based on this, we propose a pacemaker implantation location classification system based on ECG signals. Our main contributions include:

- To the best of our knowledge, we present the first **P**acemaker **I**mplantation **L**ocalization **D**ataset based on 12-lead **E**CG signals (**PILDE**), containing samples from 733 patients. The data have been anonymized to ensure privacy and ethical approval has been obtained from the relevant ethics committee.
- We propose the **STELLAR**, the **S**patio-**T**emporal **E**CG-based pacemaker **L**ocalization via **L**ead-h**AR**monic model that models the spatial correlation between leads by graph convolutional network, and innovatively employs multi-scale time-frequency convolution to capture the autocorrelation and temporal dynamics of 12-lead ECG.
- Our **STELLAR** model demonstrates a significant accuracy improvement of up to 10.45% on the **PILDE** dataset, outperforming other methods. Additionally, it achieves a 14.19% improvement on the **PTB-XL** dataset, further showcasing its robustness and effectiveness across different ECG datasets.

II. Background

High-quality, well-annotated datasets are becoming increasingly critical in the development and validation of signal processing algorithms and machine learning models in cardiology. These datasets not only facilitate the understanding of the electrophysiological characteristics of paced rhythms but also provide essential support for designing automated detection and classification systems. Cardiac pacemakers have evolved from simple single-chamber devices to more sophisticated systems, such as biventricular pacing used in cardiac resynchronization therapy (CRT) for heart failure patients. This evolution has significantly improved the quality of life for patients with conduction system disorders. As these technologies advance, the accurate identification and classification of different types of pacemaker signals have become increasingly important.

Accurate signal classification is crucial for optimizing pacemaker functionality and enhancing diagnostic precision and therapeutic outcomes, thereby further improving patient health and quality of life. By analyzing and categorizing these signals, researchers can develop smarter and more personalized rhythm management strategies, ensuring optimal treatment for each individual. High-quality datasets enable the training of intelligent classification models that can dynamically adjust treatment protocols based on detailed signal analysis, leading to enhanced patient outcomes and more effective personalized care. Thus, precise signal classification plays a pivotal role in advancing both the technical capabilities and clinical efficacy of cardiac rhythm management.

A. Cardiac Pacemaker

Cardiac pacemakers are classified by implantation site, each inducing distinct ventricular activation patterns and QRS morphologies on ECG. We consider four common types:

a) Biventricular Pacing (BP): Also known as cardiac resynchronization therapy (CRT), BP stimulates both ventricles simultaneously via leads in the right ventricle and coronary sinus. It is used for heart failure patients with dyssynchrony, producing relatively narrow, fused QRS complexes.

b) Right Ventricular Septal Pacing (RVSP): This physiological pacing method targets the interventricular septum, yielding activation closer to natural conduction. It results in narrow QRS with minimal deviation, reducing dyssynchrony risk.

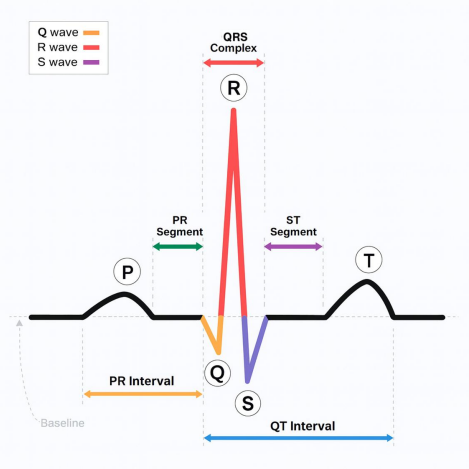

Fig. 2. Schematic representation of a typical ECG waveform.

c) Right Ventricular Apex Pacing (RVAP): A conventional approach, RVAP stimulates the right ventricular apex, leading to non-physiological apex-to-base activation. It produces wide QRS complexes (¿120 ms) with LBBB-like morphology and is associated with long-term cardiac dysfunction.

d) Left Bundle Branch Pacing (LBBP): An emerging physiological pacing technique, LBBP directly captures the left conduction system. It restores near-normal activation, generating narrow QRS (¡120 ms) that closely resembles intrinsic conduction, ideal for bundle branch block patients.

B. Principles of Pacemaker Recognition

The electrocardiogram (ECG) serves as a valuable tool for identifying cardiac pacing modes due to the distinct QRS morphological patterns generated by different lead placements and activation sequences. Each pacing mode alters ventricular depolarization in a characteristic way, resulting in reproducible changes in QRS width, axis, and waveform configuration. These ECG signatures reflect the underlying electrophysiological behavior of the paced heart and enable reliable differentiation between pacing types in clinical practice.

As shown in Fig. 2, the QRS complex as is a fundamental component of the electrocardiogram that represents the electrical depolarization of the ventricular myocardium. It typically consists of three deflections. The Q wave corresponds to the initial downward deflection. The R wave represents the first upward deflection. The S wave denotes the subsequent negative deflection following the R wave. The morphology, duration and axis of the QRS complex vary depending on the underlying conduction pathway and ventricular activation sequence. In normal sinus rhythm, the QRS duration ranges from 80 to 110 milliseconds, reflecting rapid and synchronized ventricular depolarization via the His Purkinje system. Alterations in QRS morphology such as widening, notching or abnormal axis deviation are clinically significant and may indicate conduction abnormalities, bundle branch blocks, ventricular hypertrophy or the presence of artificial cardiac pacing. As such, analysis of the QRS complex is essential for diagnosing arrhythmias, assessing cardiac function and differentiating various physiological and pathological conditions.Based on the morphological characteristics of the QRS complex, the cardiac pacing modes discussed in the previous section manifest as four distinct electrocardiographic patterns.

Biventricular pacing, commonly used in cardiac resynchronization therapy (CRT), activates both ventricles simultaneously through separate leads in the right ventricle and coronary sinus. This synchronized stimulation produces a relatively narrow QRS complex with normal or near-normal morphology, reflecting coordinated ventricular activation. The ECG typically shows fused QRS patterns, preserved P waves (if atrial sensing is present), and normal T wave orientation, closely resembling physiological conduction.

Right ventricular septal pacing delivers electrical impulses to the interventricular septum, resulting in a more physiological activation sequence compared to apical pacing. It generates a narrow QRS complex with minimal deviation in morphology—often classified as normal or slightly abnormal. The ECG shows early activation of the septum, relatively synchronous ventricular contraction, and a clearly visible pacing spike.

Right ventricular apical pacing initiates depolarization at the apex of the right ventricle, leading to slow, abnormal conduction that spreads transmurally and from apex to base. This results in a wide, aberrant QRS complex (> 120 ms) with a left bundle branch block (LBBB) or right bundle branch block (RBBB)-like pattern. The ECG is characterized by a prominent pacing spike followed by a broad QRS, ventricular dyssynchrony, and secondary ST-T wave changes, all of which are hallmarks of non-physiological ventricular activation.

Left bundle branch pacing (LBBP) represents a more recent advancement in physiological pacing. By capturing the left conduction system directly, LBBP restores rapid and synchronized ventricular activation. The resulting QRS complex is narrow (< 120 ms) with a near-normal morphology, resembling intrinsic conduction. The ECG shows early leftward activation, synchronous ventricular depolarization, and preservation of normal sequence, distinguishing it clearly from traditional right ventricular pacing.

III. Related Work

A. Traditional Machine Learning Approaches

The most commonly used machine learning methods in cardiac health monitoring systems include Support Vector Machines (SVM), deep learning approaches, Linear Discriminant Analysis (LDA), and Random Forests (RF) [13], [14]. SVM performs classification of both linear and non-linear data by mapping input samples into an n-dimensional feature space (where n denotes the number of features) and identifying an optimal hyperplane that maximizes the margin between classes while minimizing classification errors, thereby achieving effective class separation [15]. Artificial Neural Networks (ANNs), which form the foundation of deep learning, consist of interconnected neurons organized into input, hidden, and output layers. The connections between neurons are associated

with adjustable weights, which are iteratively optimized during training through backpropagation, enabling the network to learn complex patterns and make accurate predictions on unseen data [15].

LDA aims to find a linear combination of features that maximizes the separation between different classes by reducing dimensionality while preserving discriminative information. This is achieved by minimizing within-class variance and maximizing between-class variance, leading to an optimized decision boundary [16]. Random Forest is an ensemble learning method composed of multiple decision trees, each trained on different subsets of the data. Each tree generates an individual classification result, and the final output is determined through majority voting, which enhances model stability and generalization performance [15].

From a clinical perspective, RF and linear methods are advantageous due to their interpretability, facilitating the identification of meaningful biomarkers and improving clinical decision-making [17].

B. Deep Learning Approaches

Recent advances in deep learning have significantly transformed the landscape of ECG signal analysis. Unlike traditional machine learning approaches that rely on handcrafted feature extraction, deep learning models are capable of automatically learning hierarchical representations directly from raw ECG signals. This end-to-end learning capability not only reduces the dependency on domain expertise but also enables the discovery of subtle, latent patterns that may be overlooked by human experts, thereby improving the accuracy and robustness of cardiac condition classification.

Several studies have demonstrated the effectiveness of deep learning in various ECG-based diagnostic tasks. Multi-Resolution Mutual Learning Network (MRMNet) [18], which leverages a dual-resolution attention architecture and a feature complementary mechanism to significantly enhance the performance of multi-label ECG classification. ECG-LLM [19], a framework that integrates large language models (LLMs) with ECG. It pioneers the application of LLMs for imputing missing signals in 12-lead ECGs by freezing the backbone Transformer layers and introducing textual time markers. CTRhythm [20], a method for single-lead ECG atrial fibrillation detection. It integrates CNN with a Transformer encoder, which effectively captures both local features and long-range dependencies. LVSD-ECG [21], a novel approach integrating 1D CNNs with a large-scale language model (LLM). It enables simultaneous analysis of sequential ECG data and non-sequential clinical metadata. ECGMamba [22], a model based on Mamba, a specialized State Space Model (SSM). It includes an ECG encoder using three 1D convolutions and a Mamba layer with a Multi-Path Mamba-based block for global context modeling, which offering an efficient and accurate ECG classification approach. Additional research efforts [23] further underscore the growing potential of deep learning in enabling reliable, real-time cardiovascular disease screening and monitoring.

IV. Methods

A. Data Preprocessing

The raw 12-lead ECG signals are first standardized and normalized to eliminate inter-lead variations and reduce noise interference, ensuring consistent and stable inputs for subsequent heartbeat cycle segmentation and model training.

Next, the ECG data are divided into multiple heartbeat cycles based on pacing points to maintain temporal consistency across all 12 leads and increase the number of training samples. Specifically, we will first perform a difference operation on the ECG signal to highlight the changing features of the signal; further, we will calculate the mean and standard deviation (std) of the differentiated signal and find the locations of the feature points that are higher than $mean + 3 * std$, which will be considered as pacing pegs by us; finally, we will divide the signal into multiple cycles according to the pacing pegs, and we will trim or compensate the data within each cycle to ensure that the length of each cycle is 1000.

B. Model Structure

To enhance the performance of ECG diagnostic tasks, such as identifying different cardiac pacing modalities and various types of cardiac diseases, we propose a deep learning model, STELLAR, which fully leverages the spatial and temporal correlations in the 12-lead ECG signals. As shown in Fig. 3, the proposed model consists of two components: the Lead-harmonic Spatial Correlator, which captures the spatial correlation between leads, and the Multi-scale Temporal Module, which captures the autocorrelation and temporal dynamics of each lead's ECG signal.

Lead-harmonic Spatial Correlator. We innovatively integrate the ECG lead system's spatial constraints and harmonic resonance properties into a graph convolution network (GCN). Since there is a specific correlation between the 12 leads [24], we divided the 12 leads into three subgraphs [25]. **Standard Leads (I, II, III)**: Capturing the propagation pattern of frontal ECG vector loops. **Augmented Leads (aVR, aVL, aVF)**: Enhance signal-to-noise ratio of limb-end signals. **Precordial Leads (V1-V6)**: Mapping cross-sectional ventricular excitation wavefront propagation.

Given an adjacency matrix A and the feature matrix X, the GCN model constructs a filter in the Fourier domain. The filter, acting on the nodes of a graph, captures spatial features between the nodes by its first-order neighborhood, and then the GCN model can be built by stacking multiple convolutional layers, which can be expressed as:

$$H^{(l+1)} = \sigma\left(\widetilde{D}^{-\frac{1}{2}}\widetilde{A}\widetilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}\right) \tag{1}$$

where $\widetilde{A} = A + I_N$ is the matrix with added self-connections, I_N is the identity matrix, $\widetilde{D} = \sum_j \widetilde{A}_{ij}$ is the degree matrix, $H^{(l)}$ is the output of l layer, $W^{(l)}$ is a layer-specific trainable weight matrix, $H^{(l)}$ is the output of l layer, and $\sigma(.)$ denotes the ReLU function.

Thus, we fully explored the spatial correlation among the 12 leads of ECG data by the 3 GCN models. After that, we

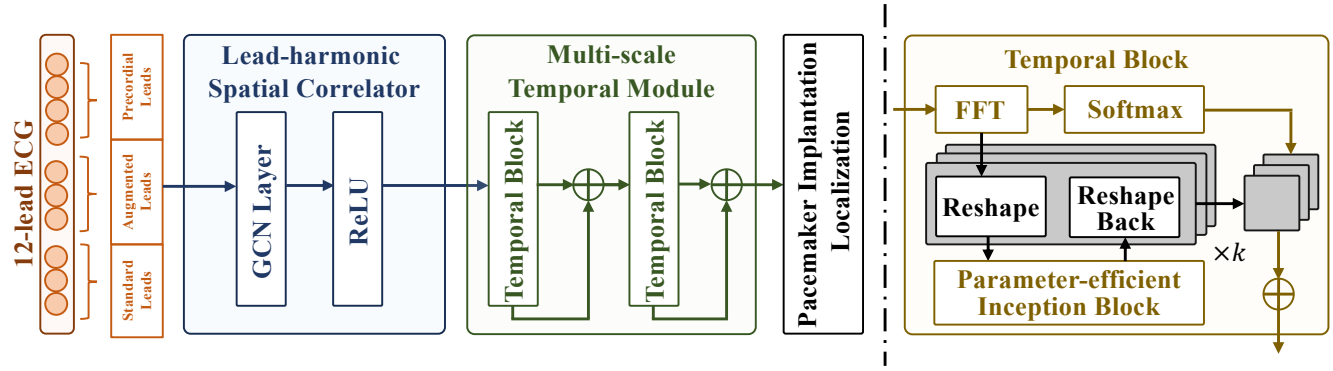

Fig. 3. Overall structure of our model.

concatenate the Embedding of the 3 GCN outputs and input the merged results as features to the Multi-scale Temporal Module to mine the temporal features of ECG signals.

Multi-scale Temporal Module. Inspired by effective methods for modeling time-series data [26]–[29], we propose the Multi-scale Temporal Module to enhance the analysis of ECG signals. Similar to approaches that transform traditional 1D time series into structured 2D tensors [26], [27], [29], our module leverages this transformation to capture intra-period and inter-period variations simultaneously. By reconstructing the 1D ECG signal into 2D tensors, the module takes advantage of the inherent multi-periodicity of the data, allowing for the application of 2D convolutional kernels. This enables the model to capture intricate temporal patterns, including short-term dynamics and long-term trends.

While the Lead-harmonic Spatial Correlator effectively captures the spatial correlations between leads through graph structures in ECG signal analysis, the Multi-scale Temporal Module further complements this by thoroughly modeling the temporal correlations inherent in ECG signals. By reconstructing the 1D signal into 2D tensors based on multiple periodicities, the Multi-scale Temporal Module extracts intricate temporal patterns, capturing both short-term intra-period dynamics and long-term inter-period trends. The Multi-scale Temporal Module comprises Temporal Blocks, which form the core building units of the model. Each Temporal Block is responsible for:

1. Multi-Periodicity Discovery: Using Fast Fourier Transform (FFT), it adaptively identifies the most significant periodic components $\{f_1, f_2, ..., f_k\}$ from the input time series, where k is set to 2 in our experiments:

$$\begin{aligned} A &= \text{Avg}(\text{Amp}(\text{FFT}(X_{1\text{D}}))), \\ \{f_1, f_2, \ldots, f_k\} &= \text{argTopK}(A), \\ p_i &= \left\lfloor \frac{T}{f_i} \right\rfloor, \quad i \in \{1, \ldots, k\}. \end{aligned} \tag{2}$$

where $\text{FFT}(\cdot)$ denotes the fast Fourier transform, $\text{Amp}(\cdot)$ denotes the computation of amplitude values, A represents the amplitude spectrum, and p_i denotes the estimated period lengths.

2. 1D to 2D Transformation: The time series $X_{1D} \in \mathbb{R}^{T \times C}$ is reshaped into multiple 2D tensors $\{X_{2D}^i\}_{i=1}^k$ based on the detected periodicities:

$$X_{2D}^i = \text{Reshape}_{p_i, f_i}\left(\text{Padding}(X_{1\text{D}})\right) \tag{3}$$

where rows capture inter-period variations and columns represent intra-period variations.

3. Parameter-Efficient Representation Learning: The transformed 2D tensors are processed by a parameter-efficient inception block with multi-scale 2D convolution kernels:

$$\hat{X}_{2D}^i = \text{Inception}(X_{2D}^i) \tag{4}$$

which enables the model to extract features across multiple temporal scales simultaneously.

4. Feature Aggregation: Processed 2D tensors are reshaped back into 1D representations and aggregated using a learned weighting mechanism based on the normalized amplitudes of the selected periods:

$$H = \sum_{i=1}^{k} \alpha_i \cdot \text{Trunc}\left(\text{Reshape}^{-1}(\hat{X}_{2D}^i)\right) \tag{5}$$

where α_i is derived from amplitude-based weights.

C. Loss Function

To enhance the recognition accuracy of the model, in addition to effectively extracting the temporal and spatial features of the ECG 12-lead signals, special attention needs to be paid to the problem of imbalance in the distribution of the sample categories because the sample imbalance may lead to the model's insufficient ability to recognize the minority class samples during the training process. Therefore, we adopt the Focal Loss to replace the cross-entropy loss function in the traditional classification problem to dynamically adjust the loss weights of different categories of samples to enhance the model's ability to discriminate minority samples.

The Focal Loss is proposed to address the extreme imbalance between foreground and background classes during training in a one-stage object detection scenario [30]. In this paper, we use Focal Loss to solve the problem of imbalance

between the samples of each category, and Focal Loss is defined as:

$$\mathrm{FL}(p_t) = -\alpha_t(1-p_t)^{\gamma}\log(p_t) \tag{6}$$

The p_t is defined:

$$p_t = \begin{cases} p, & \text{if } y = 1, \\ 1-p, & \text{otherwise.} \end{cases} \tag{7}$$

where γ is the focusing factor that controls the attention given to hard and easy samples; t denotes the category; p_t is the model's probability of predicting the true category; α_t is the weight of the corresponding category t, indicating the importance of category t. We set γ to 2 and α_t to the inverse of the number of samples in category t in our experiments.

V. Experiments

A. Experimental Setup

Datasets. To be able to evaluate the performance of STELLAR comprehensively, we use two datasets:

PILDE The PILDE dataset consists of electrocardiogram waveform files stored in the WaveForm Database (WFDB) format with 16-bit precision. The data has a resolution of 1 V/LSB and a sampling frequency of 500 Hz. The samples were collected between August 2022 and August 2024. The PILDE dataset contains 12-lead ECG data from 733 patients, each with different pacing modalities, categorized into four groups: 93 patient samples based on Biventricular Pacing (BP), 53 samples based on Right Ventricular Septal Pacing (RVSP), 181 samples based on Right Ventricular Apical Pacing (RVAP), and 406 samples based on Left Bundle Branch Pacing (LBPP). The ECG data used in this study were provided by a collaborating institution (name anonymized for review purposes). The data have been anonymized to ensure privacy and ethical approval was granted by the relevant ethics committee. The data were used in accordance with ethical guidelines. To address class imbalance, we split the dataset in an 7:1:2 ratio.

PTB-XL The PTB-XL dataset [31] includes clinical 12-lead ECGs from over 21,000 patients. There are 5 categories, including 9528 patients with NORM type, 2538 with MI type, 3008 with STTC, 4492 with CD, and 1864 with HYP. PTB-XL is also split 7:1:2 to maintain consistency with prior works.

Experimental Details. All experiments were performed on an NVIDIA A800 GPU. For training, we employ an Adam optimizer with a learning rate of **1e-4**, **100** epochs, and a batch size set to **35**.

Test Benchmarks. To validate the performance of STELLAR, we have chosen five SOTA methods for comparison, which are SCDNN [32], convTran [33], ECG-CLF [34], TodyNet [35] and ShapeFormer [36].

B. Experimental Result

1) Comparison Experiment: Table I and Table II show the accuracy of different methods on PILDE and PTB-XL. STELLAR achieves the highest accuracy on two datasets. On the PILDE dateset, the accuracy of STELLAR improves from **5.87%** to **10.45%** compared to the existing SOTA method; on the PTB-XL Dataset, the accuracy of STELLAR improves from **11.51%** to **14.19%** compared to the existing SOTA method.

TABLE I
Accuracy of Different Methods on PILDE. Bold font indicates the best accuracy.

Methods	BP	RVSP	RVAP	LBBP	Total
SCDNN-TS	47.69%	60.16%	88.55%	**94.71%**	84.91%
convTrans	61.23%	66.41%	89.19%	86.92%	82.95%
ECG-CLF	46.46%	78.90%	92.05%	84.50%	81.04%
TodyNet	49.85%	62.50%	87.60%	88.74%	81.79%
ShapeFormer	64.86%	80.00%	87.66%	90.82%	85.62%
STELLAR	**80.77%**	**80.47%**	**93.64%**	93.42%	**91.49%**

TABLE II
Accuracy of Different Methods on PTB-XL. Bold font indicates the best accuracy.

Methods	NORM	MI	STTC	CD	HYP	Total
SCDNN-TS	93.05%	33.98%	55.26%	62.50%	46.56%	70.18%
convTrans	91.18%	41.02%	61.18%	62.08%	61.70%	72.40%
ECG-CLF	95.33%	39.84%	43.42%	72.51%	21.81%	70.32%
TodyNet	90.98%	42.19%	62.17%	65.63%	43.62%	71.75%
ShapeFormer	93.05%	42.19%	59.87%	67.63%	44.68%	72.86%
STELLAR	**93.15%**	**71.09%**	**72.37%**	**81.82%**	**82.98%**	**84.37%**

On the PILDE dataset, STELLAR maintains classification accuracy above 80% on all categories, outperforming all baseline models across most diagnostic categories. This demonstrates STELLAR's ability to effectively capture both spatial and temporal dependencies in ECG signals, likely due to its integration of graph-based relational learning and dynamic time series modeling. While SCDNN-TS slightly surpasses STELLAR in LBBP, the margin is narrow and does not offset STELLAR's superior overall performance. Compared to transformer-based methods like convTrans and ShapeFormer, STELLAR shows significantly stronger classification capability, suggesting its hybrid structure better handles complex multivariate physiological patterns in ECG.

On the PTB-XL dataset, STELLAR demonstrates remarkable generalization by achieving the highest accuracy across all five classes and the total score. In particular, it significantly outperforms other models in the more challenging categories such as MI and STTC, indicating its robustness in detecting subtle waveform abnormalities. Competing models such as convTrans and ShapeFormer lag behind, especially in classes like HYP and CD, where STELLAR shows clear superiority. This consistent dominance highlights the advantage of combining temporal sequence modeling and relational structure learning, which likely enables STELLAR to learn richer, more discriminative representations from long ECG sequences compared to CNN- or transformer-only architectures.

TABLE III
RESULTS OF ABLATION EXPERIMENT. BOLD FONT INDICATES THE BEST ACCURACY. UNDERLINED FONT INDICATES THE SECOND BEST. W/O MEANS WITHOUT. LSC MEANS THE LEAD-HARMONIC SPATIAL CORRELATOR.

Methods	BP	RVSP	RVAP	LBBP	Total
w/o LSC **+**Cross Entropy	75.21%	42.97%	88.55%	93.20%	87.34%
STELLAR**+**Cross Entropy	71.37%	57.03%	**93.96%**	92.89%	89.02%
w/o LSC**+**Focal Loss	**82.05%**	78.125%	85.53%	**93.80%**	89.50%
STELLAR**+**Focal Loss	80.77%	**80.47%**	93.64%	93.42%	**91.49%**

2) Ablation Experiment: To validate the effectiveness of the Lead-harmonic Spatial Correlator (LSC) and Focal Loss, we conducted ablation experiments on the PILDE dataset. The results in Table 3 demonstrate clear performance improvements. Specifically, when using Focal Loss alone (without LSC), we observe significant gains in BP and RVSP categories compared to cross-entropy, indicating that Focal Loss is effective in addressing class imbalance by emphasizing harder samples and reducing the bias toward dominant classes.

Furthermore, when integrating the Lead-harmonic Spatial Correlator into the STELLAR model, performance further improves across most classes. This shows that LSC plays a crucial role in extracting inter-lead spatial dependencies that are physiologically meaningful. Traditional convolution or graph-based models may overlook the harmonic correlations embedded in ECG signals across leads. By explicitly modeling these lead-wise harmonic relationships, LSC enhances the spatial coherence of learned features, leading to better generalization on underrepresented pacing types like RVSP and BP.

Notably, the combination of LSC and Focal Loss achieves the highest overall accuracy (91.49%), outperforming all baselines. This synergy confirms that LSC enhances spatial feature representation, while Focal Loss improves label-level optimization. Together, they yield a more robust and class-aware classification system. These improvements are particularly important for real-world clinical deployment, where data imbalance and spatially entangled signals are common.

3) Effect of ECG sampling rate and the number of leads: To verify the effect of the ECG sampling rate and the number of leads on the accuracy, we performed the experiments on both datasets. The results in Tables IV and V reveal that both higher sampling rates and the inclusion of more ECG leads significantly contribute to improved model performance. Increasing the sampling rate from 250Hz to 1000Hz leads to a consistent boost in classification accuracy on both PILDE and PTB-XL datasets. This suggests that finer temporal resolution allows the model to capture more nuanced signal morphologies, such as pacing spikes or fragmented QRS complexes, which are vital for accurate diagnosis.

Furthermore, as the number of leads increases from 3 to 12, the classification accuracy also steadily improves due to two main factors:

TABLE IV
EFFECT OF ECG SAMPLING RATE ON CLASSIFICATION ACCURACY. BOLD FONT INDICATES THE BEST ACCURACY WITHIN EACH DATASET.

Dataset	Sampling Frequency (Hz)			
	250	500	750	1000
PILDE	84.30%	86.56%	88.59%	**91.49%**
PTB-XL	78.69%	81.43%	83.05%	**84.37%**

TABLE V
EFFECT OF NUMBER OF ECG LEADS ON CLASSIFICATION ACCURACY. BOLD FONT INDICATES THE BEST ACCURACY WITHIN EACH DATASET.

Dataset	Number of Leads			
	3	6	9	12
PILDE	84.52%	85.57%	88.03%	**91.49%**
PTB-XL	76.34%	79.06%	82.62%	**84.37%**

1) **Broader Information Capture**—each lead offers a unique perspective on the heart's electrical activity. More leads allow the model to capture additional spatial correlations and signal variations, thereby enhancing its understanding of the ECG and improving accuracy.
2) **Stronger Spatial Correlation**—as the number of leads increases, the spatial correlation in the ECG signals becomes more pronounced, enabling the model to more effectively capture the relationships between leads, further enhancing prediction performance.

VI. CONCLUSION

In conclusion, this study makes a significant contribution to the interdisciplinary field of artificial intelligence and cardiac electrophysiology through the introduction of the PILDE dataset—the first longitudinal 12-lead ECG dataset specifically designed for pacemaker implantation research. PILDE provides large-scale, high-quality electrocardiographic recordings, along with comprehensive clinical annotations.

Building upon PILDE, we propose STELLAR, a novel deep learning framework incorporating a Spatio-Temporal Lead-Harmonized Mechanism. Experimental results demonstrate that STELLAR achieves superior performance in arrhythmia classification tasks compared to conventional machine learning methods and state-of-the-art deep learning models, exhibiting both high accuracy and enhanced interpretability.

ACKNOWLEDGMENT

This work is supported in part by National Natural Science Foundation of China (No. 61936015), Natural Science Foundation of Shanghai (No. 24ZR1430600) and Shanghai Key Laboratory of Trusted Data Circulation and Governance, and Web3.

REFERENCES

[1] "Cardiovascular diseases (cvds)," Website, 2021, https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds).

[2] M. Brunner, M. Olschewski, A. Geibel, C. Bode, and M. Zehender, "Long-term survival after pacemaker implantation: prognostic importance of gender and baseline patient characteristics," *European Heart Journal*, vol. 25, no. 1, pp. 88–95, 2004.

[3] J. N. Koneru, P. W. Jones, E. F. Hammill, N. Wold, and K. A. Ellenbogen, "Risk factors and temporal trends of complications associated with transvenous implantable cardiac defibrillator leads," *Journal of the American Heart Association*, vol. 7, no. 10, p. e007691, 2018.

[4] V. Tsibulko, I. Iliev, and I. Jekova, "A review on pacemakers: Device types, operating modes and pacing pulses. problems related to the pacing pulses detection." *International Journal Bioautomation*, vol. 18, no. 2, 2014.

[5] B. Boyle, C. J. Love, J. E. Marine, J. Chrispin, A. S. Barth, J. W. Rickard, D. D. Spragg, R. Berger, H. Calkins, and S. K. Sinha, "Radiographic identification of cardiac implantable electronic device manufacturer: smartphone pacemaker-id application versus x-ray logo," *The Journal of Innovations in Cardiac Rhythm Management*, vol. 13, no. 8, p. 5104, 2022.

[6] S. Jacob, M. A. Shahzad, R. Maheshwari, S. S. Panaich, and R. Aravindhakshan, "Cardiac rhythm device identification algorithm using x-rays: Cardia-x," *Heart Rhythm*, vol. 8, no. 6, pp. 915–922, 2011. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1547527111000142

[7] S. Ploux, M. Strik, N. Varma, R. Eschalier, and P. Bordachar, "Remote monitoring of pacemakers," *Archives of Cardiovascular Diseases*, vol. 114, no. 8, pp. 588–597, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1875213621001285

[8] R. S. Andersen, A. Peimankar, and S. Puthusserypady, "A deep learning approach for real-time detection of atrial fibrillation," *Expert Systems with Applications*, vol. 115, pp. 465–473, 2019.

[9] S. Gustafsson, D. Gedon, E. Lampa, A. H. Ribeiro, M. J. Holzmann, T. B. Schön, and J. Sundström, "Development and validation of deep learning ecg-based prediction of myocardial infarction in emergency department patients," *Scientific Reports*, vol. 12, no. 1, p. 19615, 2022.

[10] J. Choi, S. Lee, M. Chang, Y. Lee, G. C. Oh, and H.-Y. Lee, "Deep learning of ecg waveforms for diagnosis of heart failure with a reduced left ventricular ejection fraction," *Scientific Reports*, vol. 12, no. 1, p. 14235, 2022.

[11] E. Prifti, A. Fall, G. Davogustto, A. Pulini, I. Denjoy, C. Funck-Brentano, Y. Khan, A. Durand-Salmon, F. Badilini, Q. S. Wells *et al.*, "Deep learning analysis of electrocardiogram for risk prediction of drug-induced arrhythmias and diagnosis of long qt syndrome," *European Heart Journal*, vol. 42, no. 38, pp. 3948–3961, 2021.

[12] M. Jastrzebski, J. Wiliński, K. Fijorek, T. Sondej, and D. Czarnecka, "Mortality and morbidity in cardiac resynchronization patients: impact of lead position, paced left ventricular qrs morphology and other characteristics on long-term outcome," *Europace*, vol. 15, no. 2, pp. 258–265, 2013.

[13] S. H. Kwon and L. Dong, "Flexible sensors and machine learning for heart monitoring," *Nano Energy*, vol. 102, p. 107632, 2022.

[14] K. Rjoob, R. Bond, D. Finlay, V. McGilligan, S. J. Leslie, A. Rababah, A. Iftikhar, D. Guldenring, C. Knoery, A. McShane, A. Peace, and P. W. Macfarlane, "Machine learning and the electrocardiogram over two decades: Time series and meta-analysis of the algorithms, evaluation metrics and applications," *Artificial Intelligence in Medicine*, vol. 132, p. 102381, 2022.

[15] S. Uddin, A. Khan, M. E. Hossain, and M. A. Moni, "Comparing different supervised machine learning algorithms for disease prediction," *BMC medical informatics and decision making*, vol. 19, no. 1, p. 281, 2019.

[16] T. M. C. Pereira, R. C. Conceição, and R. Sebastião, "Initial study using electrocardiogram for authentication and identification," *Sensors*, vol. 22, no. 6, p. 2202, 2022.

[17] A. Lyon, A. Mincholé, J. P. Martínez, P. Laguna, and B. Rodriguez, "Computational techniques for ecg analysis and interpretation in light of their contribution to medical advances," *Journal of the Royal Society Interface*, vol. 15, no. 138, p. 20170821, 2018.

[18] W. Huang, N. Wang, P. Feng, H. Wang, Z. Wang, and B. Zhou, "A multi-resolution mutual learning network for multi-label ecg classification," in *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2024, pp. 3303–3306.

[19] L. Liu, G. Cui, C. Wan, D. Wu, and Y. Li, "Ecg-llm: Leveraging large language models for low-quality ecg signal restoration," in *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2024, pp. 3537–3542.

[20] Z. Liang, C. Yang, Z. Yu, Y. Fu, B. Ren, M. Lin, Q. Li, X. Liu, Y. Chen, and L. C. Xia, "Ctrhythm: Accurate atrial fibrillation detection from single-lead ecg by convolutional neural network and transformer integration," in *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2024, pp. 4452–4458.

[21] W. Shim, N. Park, D. Ko, H. Gwag, Y. Park, S. Park, and J. Kim, "Lmcg-net: Integrating llms and ecg for lvef-based classification for pacing patients," in *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2024, pp. 5309–5314.

[22] Y. Qiang, X. Dong, X. Liu, Y. Yang, F. Hu, and R. Wang, "Ecgmamba: Towards ecg classification with state space models," in *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2024, pp. 6498–6505.

[23] Yildirim, "A novel wavelet sequence based on deep bidirectional lstm network model for ecg signal classification," *Computers in Biology and Medicine*, vol. 96, pp. 189–202, 2018.

[24] C. Ji, L. Wang, J. Qin, L. Liu, Y. Han, and Z. Wang, "Msgformer: A multi-scale grid transformer network for 12-lead ecg arrhythmia detection," *Biomedical Signal Processing and Control*, vol. 87, p. 105499, 2024.

[25] M. Holderith and T. Schanze, "Cross-correlation based comparison between the conventional 12-lead ecg and an easi derived 12-lead ecg," *Current Directions in Biomedical Engineering*, vol. 4, no. 1, pp. 621–624, 2018.

[26] H. Wu, T. Hu, Y. Liu, H. Zhou, J. Wang, and M. Long, "Timesnet: Temporal 2d-variation modeling for general time series analysis," *arXiv preprint arXiv:2210.02186*, 2022.

[27] F. Zhao, W. Lin, S. Lin, S. Tang, and K. Li, "Mscnet: Multi-scale network with convolutions for long-term cloud workload prediction," *IEEE Transactions on Services Computing*, 2025.

[28] Y. Wang, H. Wu, J. Dong, G. Qin, H. Zhang, Y. Liu, Y. Qiu, J. Wang, and M. Long, "Timexer: Empowering transformers for time series forecasting with exogenous variables," *Advances in Neural Information Processing Systems*, vol. 37, pp. 469–498, 2025.

[29] C. Si, W. Yu, P. Zhou, Y. Zhou, X. Wang, and S. Yan, "Inception transformer," *Advances in Neural Information Processing Systems*, vol. 35, pp. 23 495–23 509, 2022.

[30] T.-Y. Ross and G. Dollár, "Focal loss for dense object detection," in *proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2980–2988.

[31] 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, pp. 1–15, 2020.

[32] C. Liu, S. Cheng, W. Ding, and R. Arcucci, "Spectral cross-domain neural network with soft-adaptive threshold spectral enhancement," *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

[33] N. M. Foumani, C. W. Tan, G. I. Webb, and M. Salehi, "Improving position encoding of transformers for multivariate time series classification," *Data Mining and Knowledge Discovery*, vol. 38, no. 1, pp. 22–48, 2024.

[34] 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*, vol. 28, no. 5, pp. 2461–2472, 2024.

[35] H. Liu, D. Yang, X. Liu, X. Chen, Z. Liang, H. Wang, Y. Cui, and J. Gu, "Todynet: temporal dynamic graph neural network for multivariate time series classification," *Information Sciences*, p. 120914, 2024.

[36] X.-M. Le, L. Luo, U. Aickelin, and M.-T. Tran, "Shapeformer: Shapelet transformer for multivariate time series classification," in *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2024, pp. 1484–1494.