symmetry

MDPI

Article

A Systematic Method Combining Rotated Convolution and State Space Augmented Transformer for Digitizing and Classifying Paper ECGs

Xiang Wang and Jie Yang *

Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, Shanghai 200240, China
* Correspondence: jieyang@sjtu.edu.cn

Abstract: Billions of paper Electrocardiograms (ECGs) are recorded annually worldwide, particularly in the Global South. Manual review of this massive dataset is time-consuming and inefficient. Accurate digital reconstruction of these records is essential for efficient cardiac disease diagnosis. This paper proposes a systematic framework for digitizing paper ECGs with 12 symmetrically distributed leads and identifying abnormal samples. This method consists of three main components. First, we introduce an adaptive rotated convolution network to detect the positions of lead waveforms. By exploiting the symmetric distribution of 12 leads, a novel loss is proposed to improve the detection model's performance. Second, image processing techniques, including denoising and connected component analysis, are employed to digitize ECG waveforms. Finally, we propose a transformer-based classification method combined with a state space model. Our process is evaluated on a large synthetic dataset, including ECG images characterized by rotations, noise, and creases. The results demonstrate that the proposed detection method can effectively reconstruct paper ECGs, achieving an 11% improvement in SNR compared to the baseline. Moreover, our classification model exhibits slightly higher performance than other counterparts. The proposed approach offers a promising solution for the automated analysis of paper ECGs, supporting clinical decision-making.

Keywords: paper ECGs; digitization; rotated object detection; state space model; transformer

Academic Editors: Rumen Mironov, Roumiana Kountcheva and Hsien-Chung Wu

Received: 17 October 2024
Revised: 12 December 2024
Accepted: 20 December 2024
Published: 14 January 2025

Citation: Wang, X. and Yang, J. A Systematic Method Combining Rotated Convolution and State Space Augmented Transformer for Digitizing and Classifying Paper Electrocardiograms. *Symmetry* **2025**, *17*, 120. https://doi.org/10.3390/sym17010120

1. Introduction

The electrocardiogram (ECG) is the most common pre-screening tool for diagnosing cardiovascular diseases (CVDs) [1,2]. Currently, devices for detecting ECG signals include standard 12-lead electrocardiographs and portable or wearable ECG devices. Compared to other devices, 12-lead electrocardiographs can provide comprehensive, high-quality ECG signals, enabling more accurate classification and diagnosis of CVDs. Recently, researchers have developed many algorithmic approaches to interpreting the digital representations of the ECG waveforms. Although digital ECG methods offer the potential for increased access to ECG-based diagnoses and cardiac care, physical or paper ECGs have been a cornerstone of cardiac care for nearly a century and continue to be widely used, particularly in the Global South. There are likely billions of paper ECGs recorded each year globally [3]. This legacy embodies the variation and evolution of CVDs across different populations, regions, and time. However, proprietary systems with limited interoperability artificially exacerbate barriers to data analysis. Consequently, the digitization of ECGs and access to low-cost analytical tools are essential for capturing the full spectrum of ECG data and enhancing global accessibility of cardiac care [4]. The diagnosis of heart diseases using paper ECGs

consists of two stages: digitization of paper ECGs and classification of the resulting digital signals. Currently, several models have been explored to process paper ECGs.

(1) Digitization

There are two main paths to digitizing paper-based ECGs. One path combines traditional image processing techniques such as filtering, edge detection, and binary image segmentation with human expertise. Sibel et al. [5] introduced a method to digitize paper ECGs using 2D median filtering and simple image segmentation steps. Then, empirical mode decomposition and SVM were applied to detect abnormalities in the ECG signal. Sun et al. [6] developed an automated algorithm using edge detection and connected component analysis to separate ECG signals from scanned 12-lead paper ECGs. Wu et al. [7] presented a novel approach to digitizing paper ECGs employing automated horizontal and vertical anchor point detection and a dynamic morphological algorithm. Randazzo et al. [8] proposed a conversion algorithm from paper ECGs to digital ECGs by cropping images manually and binary thresholding segmentation. Ref. [9] reported a MATLAB-based tool that digitized paper ECGs through grayscale thresholding, column-wise pixel scanning, and template-based optical character recognition. Ref. [10] presented a MATLAB-based tool and algorithm that digitized printed or scanned ECG signals through image processing and serial steps, achieving high validation accuracy on a dataset of 30 scanned ECG images. Ref. [11] proposed a method that extracts features of varying grayscale levels from binarized paper ECGs rather than directly digitizing the ECG curves. These features are subsequently fed into a classifier for anomaly detection.

The other path is based on end-to-end deep learning models. Ref. [12] proposed a novel method combining U-Net and ResNet architectures to digitize and classify relatively clean paper-based ECGs. Digitizing paper ECGs with high-level noise currently remains a challenging task. Ref. [13] addressed this issue by a U-Net-based deep learning approach incorporating grid removal and connected component analysis. Although this method can handle images with diverse lead layouts, it struggles with rotated images.

Generally, current methods for digitizing paper ECGs have some shortcomings. First, deep learning models for digitizing 12-lead paper ECGs with symmetrical distributions need further investigation. Second, large-scale labeled datasets of paper ECGs are difficult to obtain, and existing methods are typically evaluated on small datasets of a few hundred samples [7,8,13]. Third, most methods achieve high performance only on high-quality paper ECG images. Little effort addresses low-quality paper ECG images, such as those with rotations, wrinkles, high-level noise, or missing signal leads.

(2) Classification

A growing body of research has highlighted the potential of deep learning models for accurate classification in various tasks [14–16]. Ref. [17] reported a wide and deep transformer neural network to classify 12-lead ECG sequences into 27 cardiac abnormality classes, combining handcrafted ECG features. Ref. [18] explored the application of structured state space models for ECG classification (SSM_ECG), demonstrating significant improvements over convolutional architectures in capturing long-term dependencies within time series data. Ref. [19] developed a multi-view and multi-scale deep neural network for ECG classification (MVMS), which treats different leads as distinct views and uses a multi-scale convolutional neural network to capture temporal features at various scales.

Although classification models for ECGs have been extensively investigated [20], the classification of digitized ECG signals warrants further research. A comprehensive evaluation of existing methods in this domain is necessary.

This paper proposes a deep learning-based approach for systematically processing large-scale paper ECG images to address these challenges:

(1) We generated a PTB-XL dataset containing 21,837 labeled paper ECG samples using the simulation tool provided by the 2024 PhysioNet/CinC (CinC2024) challenge. These images include 12-lead ECG signals and contain various distortions such as rotations, wrinkles, and high-level noise.
(2) The proposed model employs oriented R-CNN with adaptive rotated convolutions for 12-lead object detection. A new paper-ECG loss is introduced by leveraging the symmetrical distribution of leads to improve the model accuracy of detecting different leads. Each lead curve is segmented and digitized using a set of rectifiers.
(3) This work trains a state space augmented transformer model combining handcrafted features to identify abnormal ECG signals.

The following contents are structured as follows. Section 2 reviewed some related work in this work, covering essential backgrounds of rotated object detection, state space model, and transformers. In Section 3, we elaborated on the details of our model. Section 4 showed the experimental results and discussion. Section 5 is the conclusion.

2. Related Work

2.1. Rotated Object Detection

The accurate rotated object detection is quite significant in various tasks, including scene text detection [21], face detection [22], and aerial image recognition [23]. Recent research has yielded significant advancements in rotated object detection, particularly in the development of rotated object representations [24–26] and their associated loss functions [27–29]. Studies have also extensively explored the structure of detection networks, including the network's neck [30,31], the detection head [32], and rotated region proposal networks [33].

Pu et al. (2023) [34] proposed an adaptively rotated convolution (ARC) module to construct a backbone model for the rotated object detection task. This module employs adaptive rotation of the convolution kernel according to the input feature maps. By employing a conditional computation mechanism, it can dynamically adjust its operations to handle multi-oriented objects. An ARC module comprises n kernels $(\mathbf{W_1}, \cdots, \mathbf{W_n})$, each possessing a shape of $[C_{out}, C_{in}, k, k]$. Using the input feature $\boldsymbol{x}$, the routing function f calculates the rotation angles $\boldsymbol{\theta}$ and the corresponding combination weights $\boldsymbol{\lambda}$:

$$\boldsymbol{\theta}, \boldsymbol{\lambda} = f(\boldsymbol{x}).$$

Each of the n kernels is first rotated according to its predicted rotation angle $\boldsymbol{\theta} = [\theta_1, \theta_2, \cdots, \theta_n]$,

$$\mathbf{W_i'} = \text{Rotate}(\mathbf{W_i}; \theta_i),\ i = 1, 2, \cdots, n.$$

Here, θ_i represents the rotation angle for $\mathbf{W_i}$, $\mathbf{W_i'}$ is the rotated kernel, and Rotate$(\cdot)$ denotes the rotation procedure for a $k \times k$ convolution kernel. Applying conditional parameterization, the output features $\boldsymbol{y}$ can be represented as

$$\boldsymbol{y} = (\lambda_1 \mathbf{W_1'} + \lambda_2 \mathbf{W_2'} + \cdots + \lambda_n \mathbf{W_n'}) * \boldsymbol{x}.$$

The ARC module can increase the network's representation ability by capturing the features of multiple-oriented objects. It can be readily integrated into any backbone network featuring convolutional layers.

Convolutional neural networks (CNNs), the most prevalent deep learning framework, are fundamental to image object detection. Variants of CNNs have surpassed traditional machine learning approaches in various tasks [35]. In this work, we build an ARC-based ResNet-50-FPN [36] to detect different leads from paper ECGs with various orientations.

2.2. Structured State Space Sequence (S4) Model

The structured state space model (SSSM) leverages a linear state space transition equation to link input and output sequences through a hidden state. Specifically, given a one-dimensional sequence $\boldsymbol{u}(t)$ (input) and a one-dimensional sequence $\boldsymbol{y}(t)$ (output), the transition equation can be defined as:

$$\begin{aligned} \boldsymbol{x}'(t) &= \mathbf{A}\boldsymbol{x}(t) + \mathbf{B}\boldsymbol{u}(t), \\ \boldsymbol{y}(t) &= \mathbf{C}\boldsymbol{x}(t) + \mathbf{D}\boldsymbol{u}(t), \end{aligned} \tag{1}$$

where $\boldsymbol{x}(t)$ represents an N-dimensional hidden state vector, and $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$ denote the transition matrices.

Given a step size Δ, the continuous-time parameters can be mapped to discrete-time parameter $\bar{\mathbf{A}}$, $\bar{\mathbf{B}}$, and $\bar{\mathbf{C}}$. These discrete parameters form the state-space model convolutional kernel $\mathcal{K}(\bar{\mathbf{A}}, \bar{\mathbf{B}}, \bar{\mathbf{C}})$, enabling the calculation of the output $\boldsymbol{y}$ using convolution: $\boldsymbol{y} = \mathcal{K} * \boldsymbol{u}$. A significant contribution of [37] lies in developing a stable and efficient method for evaluating the kernel $\mathcal{K}$. Building upon their previous work [38], they propose a specific initialization strategy for the matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$, inspired by HiPPO theory, to facilitate the capture of long-range interactions. By concatenating and fusing H copies of these layers, each performing a mapping from $\mathbb{R}$ to $\mathbb{R}$, an S4 layer is formed, capable of mapping from $\mathbb{R}^H$ to $\mathbb{R}^H$.

To further enhance SSSM capabilities, an S4 model can be built by stacking multiple S4 layers with normalization and point-wise fully connected layers. This architecture has demonstrated impressive performance on various long-range sequence tasks, including 12-lead ECG classification and generation [18,39], object detection [40], and probabilistic time series forecasting [41,42].

2.3. Transformers

Transformers [43], relying on self-attention mechanisms, have been widely employed across AI fields for analyzing temporal features of long sequences, such as computer vision, audio processing, and natural language processing [44]. Various transformer-based models have been developed to process highly long sequences and reduce computational cost, such as Informer [45], FEDformer [46], and Quatformer [47]. Ref. [48] introduced Pyraformer, which utilized a pyramidal attention module to analyze long-term and short-term trends within the data and efficiently capture the complex time relationships. Ref. [49] have proposed the Autoformer, which takes the series as a fundamental building block of deep models and uses an auto-correlation mechanism to capture long-range dependencies more efficiently than self-attention methods.

Transformers exhibit high computational costs for long sequences and are susceptible to overfitting due to the absence of structural biases [44]. Local attention mechanisms [50,51] introduce structural biases but compromise the ability to capture global context. A state space augmented transformer (SPADE) [52] addresses this by incorporating S4 model to provide strong structural biases and global information. This work transfers this approach to ECG signal classification, leveraging hand-crafted features.

3. Methodology

3.1. Datasets & Problem Statements

Our work is based on the PTB-XL dataset [53], which comprises 21,837 12-lead ECG recordings at a rate of 100 Hz. These ECGs are obtained from six limb leads (I, II, III, aVF, aVL, aVR) and six precordial leads (V1, V2, V3, V4, V5, V6). All annotations for each sample cover five superclasses comprising normal (NORM), myocardial infarction (MI), ST/T changes (STTC), conduction disturbance (CD), and hypertrophy (HYP).

We employed the publicly available ECG-Image-Kit toolbox (version 1.0) [4] to generate corresponding paper ECGs for the recordings in PTB-XL. These images encompass 12-lead ECG signal segments subjected to various distortions such as rotations, creases, cropping, wrinkles, and high-level noise. Each ECG image consists of 26 classes: 13 lead waveforms and their names. The first 12 lead waveforms represent 2.5 s segments of each lead, while the 13th waveform corresponds to the entire 10 s ECG curve of lead II. Each ECG "lead" and "lead name" are annotated with a rotated bounding box defined by its four vertex coordinates.

This work introduces a three-stage approach to diagnose abnormal cardiac diseases based on paper ECGs, as depicted in Figure 1. The method leverages an ECG-ARCResNet model for signal detection, followed by a comprehensive image processing pipeline to digitize the ECG signals. The digitized ECG samples are then fed into a transformer-based classification model to identify diverse CDVs.

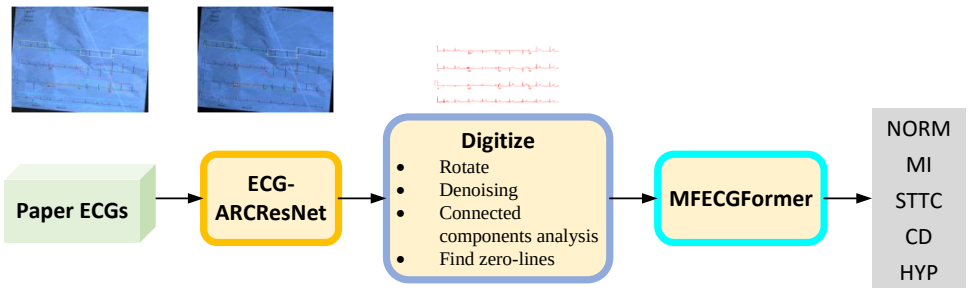


Figure 1. System flowchart for digitizing paper ECGs and classifying five CDVs.

3.2. Waveform Detection

In this work, we employ the oriented R-CNN as the body of the detection model. Given ARC modules' superior ability to learn feature representations of rotated objects, we have adopted ARC-ResNet50 as our backbone, as shown in Figure 2.

Oriented R-CNN, a two-stage detector based on FPN, comprises an oriented RPN and an oriented R-CNN head. The former generates high-quality oriented proposals, while the latter classifies these proposals and refines their spatial locations through bounding box regression. Oriented R-CNN leverages cross-entropy loss to classify proposals and determine object categories. Smooth L1 loss is adopted to regress the bounding boxes, refining their position, size, and orientation to localize the objects better. The positional distribution of the 26 classes in paper ECG records produced by unified ECG devices is fixed. By leveraging this structural information, this paper proposes paperECGLoss to improve detection performance. Next, we describe it in detail.

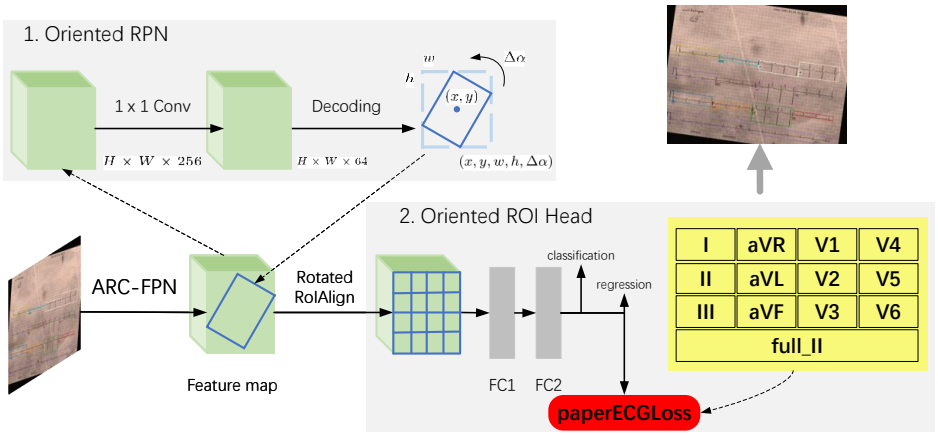


Figure 2. As a two-stage detector, Oriented R-CNN first generates oriented proposals via an oriented RPN, followed by a second stage using an oriented R-CNN head for classification and regression.

As shown in Figure 2, the first 12 ECG leads are arranged symmetrically in a 3×4 grid. The spatial relationships among these classes can be summarized as follows: (1) All lead bounding boxes are parallel along the X-coordinate direction. (2) The bounding boxes of the first 12 leads have a uniform length. (3) The bounding boxes at the two extremities are collinear along the Y-direction. (4) The full_II bounding box has a length four times that of the others.

$$\text{diff-1}_{i,j} = \text{k}_{X:i} - \text{k}_{X:j}, \quad 1 \le i, j \le 13 \tag{2}$$

$$\text{diff-2}_{i,j} = \text{len}_{X:i} - \text{len}_{X:j}, \quad 1 \le i, j \le 12 \tag{3}$$

$$\text{diff-3}_{i} = \text{len}_{X:13} - 4 \times \text{len}_{X:i}, \quad 1 \le i \le 12 \tag{4}$$

$$\text{diff-4}_{i,j,k} = \text{k}_{Y:i,j} - \text{k}_{Y:i,k}, \quad i, j, k \in \{1, 5, 9, 13\} or \{4, 8, 12, 13\} \tag{5}$$

where $\text{k}_{X:i}$ and $\text{len}_{X:i}$ denote the slope, and the length of the bounding box corresponding to the i-th lead in the X-direction, respectively. $\text{k}_{Y:i,j}$ means the slope of the line segment joining the top-left corners of the i-th and j-th lead bounding boxes. The paperECGLoss is defined as follows:

$$paperECGLoss = MSE(\text{diff-1}) + MSE(\text{diff-2}) + MSE(\text{diff-3}) + MSE(\text{diff-4})$$

where $MSE(\text{diff-1})$ means the mean of the squared error for diff-1. Finally, we train the detection model with a combination of smooth L1 loss and paperECGLoss to achieve more precise localization of target bounding boxes.

3.3. Digitization

Segmenting and digitizing the lead curves from paper ECG images is essential to identify abnormal samples automatically. The specific procedures are as follows:

(1) Zero-line Alignment: The center points of the bounding boxes for lead names on the same row should align on a straight line. By calculating the slope of the lines connecting these center points, the rotation angle of the image can be determined, enabling the correction of rotated images.

(2) Noise Reduction and Segmentation: an unsupervised denoising model, Noise2Void [54], is trained to denoise the original image and enhance the foreground, making it easier to segment the ECG signal curves. Based on the output coordinates from the detection model, each lead curve region is segmented from the denoised image. For each sub-

image containing a lead, binarization is performed, followed by clustering of pixels to separate the background, grid, and signal curve. Finally, connected component analysis is employed to isolate the lead curve.

(3) Zero-line Detection: A "multi-level sliding window" technique is employed to approximate the zero-line of each lead curve. Specifically, a window with a decreasing width is slid along the vertical axis iteratively. At each iteration, the window containing the most data points is considered the region where the zero-line is. The average vertical coordinate of the data points within the final 4-pixel-wide window is taken as the vertical coordinate of the zero-line.

(4) Horizontal Coordinate Determination: Based on the range of each lead curve, the horizontal coordinate range of the 10 s curve is determined. Given the image resolution *res* and the horizontal time resolution of 25 mm/s, the number of data points contained in a 10 s segment can be calculated as:

$$N = \frac{10[\text{s}] \times 25[\text{mm/s}]}{25.4[\text{mm/inch}]} \times res[/\text{inch}]. \tag{6}$$

By combining the position of each curve, the horizontal coordinate range of each 2.5 s signal segment is separated.

(5) Vertical Coordinate Determination: The signal intensity y corresponding to each data point can be determined using the following formula:

$$y = \frac{pixel\_y}{res[/\text{inch}]} \times \frac{25.4[\text{mm/inch}]}{10[\text{mm/mv}]}. \tag{7}$$

Here, *pixel_y* represents the vertical coordinate of the data point relative to the zero-line, and the spatial resolution is 10 mm/mv.

The digitization process may result in the loss of a few pixels in the curves, which are then interpolated linearly. Due to varying image resolutions, the digitized signals have different lengths. All digitized signals are resampled to 250 or 1000 samples using interpolation or averaging for consistency.

3.4. Classification

Based on the S4-augmented transformer (SPADE), this work trains a multi-feature ECGFormer for classifying abnormal ECG samples, as illustrated in Figure 3. SPADE can capture both global and local dependencies through a hierarchical transformer-based architecture. At the bottom layer, a S4 model captures coarse global information by inducing a strong structural bias. The subsequent conventional transformer layers extract more intricate local dependencies. The implementation details of this model are described below.

Let $\boldsymbol{x} \in \mathbb{R}^{L \times D}$ be an input sample where the length is L, and the embedding size is D. The preliminary features across dimensions at each time step are processed through a one-dimensional convolution operation. It expands the feature dimension, enabling the model to learn richer representations. The SPADE module comprises three multi-head self-attention (MSA) encoders. The first layer is a S4-augmented self-attention encoder. The computational process is as follows:

$$\begin{aligned} \boldsymbol{x}_{S4} &= \text{LN}\ (\text{S4}(\boldsymbol{x}_i)), \\ \boldsymbol{x}_{MSA} &= \text{MSA}\ (\boldsymbol{x}_i), \\ \bar{\boldsymbol{x}}_i &= \text{LN}\ (\mathbf{W}\ [\boldsymbol{x}_{S4},\ \boldsymbol{x}_{MSA}] + \boldsymbol{x}_i), \\ \boldsymbol{x}_{i+1} &= \text{LN}\ (\text{FFN}(\bar{\boldsymbol{x}}_i) + \bar{\boldsymbol{x}}_i), \end{aligned} \tag{8}$$

where x_i denotes the input of the i-th layer. The subsequent two layers employ standard MSA networks [44]. Subsequently, the SPADE output feature vector y is compressed via max pooling and serves as the classifier's input.

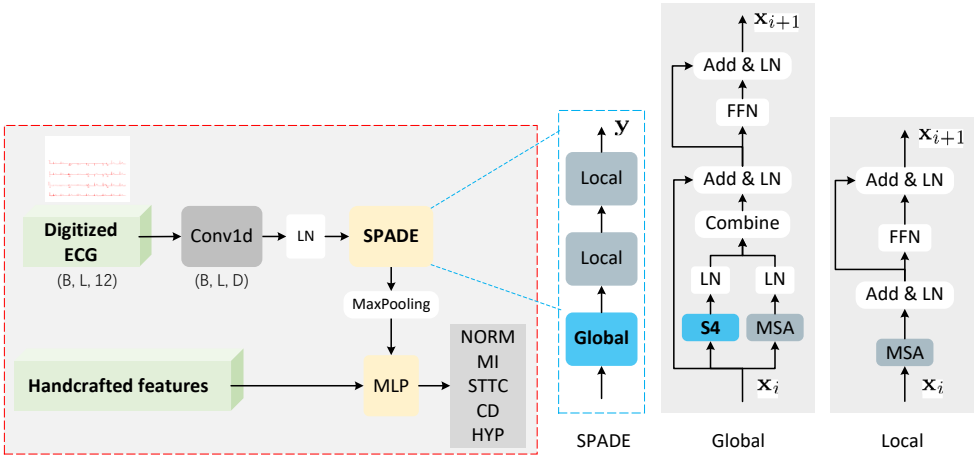


Figure 3. The overall architecture of MFECGFormer for ECG signals classification. This model's backbone, SPADE, consists of one S4-augmented attention network and two standard multi-head self-attention networks.

In this work, we design a set of time-frequency domain features from the lead II of ECG signals. These features, which encompass waveform statistics and two crucial metadata attributes (age and sex), are integrated into the input feature vector. The main handcrafted features (HCFeats) are listed in Appendix A. Finally, these features are concatenated with the output of SPADE and fed into a two-layer perceptron for ECG classification.

$$\textbf{Outputs} = \text{MLP}([\text{MaxPooling}(\boldsymbol{y}); \textbf{HCFeats}]). \tag{9}$$

4. Experiments

4.1. Experiment Settings

All models were carried out using the PyTorch Library. All experiments took place on a Linux server with an Intel(R) Xeon(R) CPU E5-2680, 256 GB of RAM, and an NVIDIA GeForce RTX 3090 GPU. Using hierarchical sampling, we stratified the real dataset into ten folds to ensure class balance. The model was trained on folds 1 through 8, with fold 9 used for validation and fold 10 reserved for testing. The network's hyperparameter settings are elaborated in Appendix B. The training process utilized the following parameters: a batch size of 32, a learning rate of 0.001, and the AdamW optimizer.

4.2. Results and Discussion

This work employs the average precision (AP) and signal-to-noise ratio (SNR) to evaluate the performance of paper ECG digitization methods.

$$\text{SNR} = 10 \log_{10} \left(\frac{\sum_{i=1}^{N} s[i]^2}{\sum_{i=1}^{N} (x[i] - s[i])^2} \right),$$

where $s[i]$ and $x[i]$ signify the original and digitized ECG signals, respectively.

AUROC, Precision, Recall, F1-score, and accuracy are used to compare the performance of different classification models.

$$\begin{aligned} \text{Precision} &= \frac{\text{TP}}{\text{TP+FP}}, \\ \text{Recall} &= \frac{\text{TP}}{\text{TP+FN}}, \\ \text{F1-score} &= 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}, \\ \text{Accuracy} &= \frac{\text{TP + TN}}{\text{TP + TN+FP+FN}}, \end{aligned} \tag{10}$$

where FP, TN, TP, and FN represent the number of false positives, true negatives, true positives, and false negatives, respectively.

Digitization

Accurately detecting each lead's location and name serves two primary purposes. Firstly, the image skew angles can be corrected using the bounding box coordinates of the lead names. Secondly, the lead curves can be separated from their respective image regions, reducing the impact of image noise from other areas.

This work built the detection model with hyperparameters from Appendix B Table A2 and trained for 20 epochs. We report detailed experimental results, including category-wise average precision (AP) and the mean average precision (mAP), for comparison with existing state-of-the-art oriented object detectors. The AP and mAP for all lead names and curves are presented in Table 1 and Table 2, respectively.

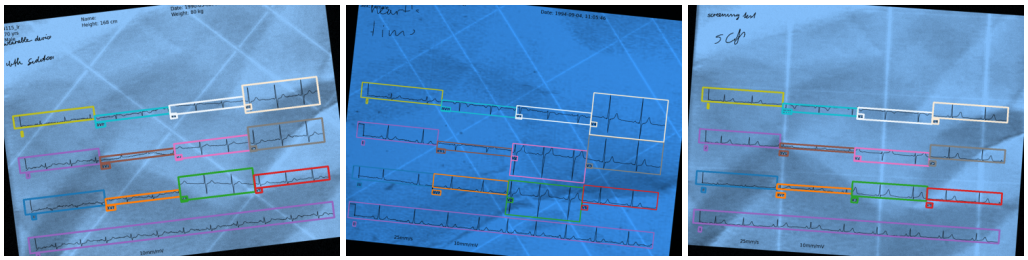
Table 1. Comparison of the average precision of each lead name and the mean average precision (mAP) with state-of-the-art methods.

Methods	I	II	III	aVR	aVL	aVF	V1	V2	V3	V4	V5	V6	full_II	mAP
Oriented R-CNN	58.16	77.39	89.62	97.11	98.04	95.80	95.46	95.78	95.43	94.56	94.62	95.71	77.55	89.63
ARC R-CNN	57.44	79.92	90.20	97.92	98.20	96.35	96.60	95.89	96.33	94.51	94.92	95.22	78.59	90.16
Ours	60.33	76.98	92.49	97.95	98.71	97.15	96.32	97.02	97.12	97.03	96.93	96.75	89.30	**91.85**

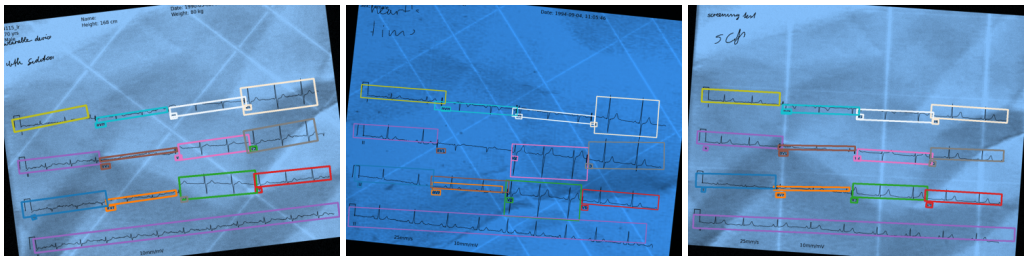
Table 2. Comparison of the average precision of each lead curve and the mean average precision (mAP) with state-of-the-art methods. The evaluation metrics are mAP and SNR.

Methods	I	II	III	aVR	aVL	aVF	V1	V2	V3	V4	V5	V6	full_II	mAP	SNR
Oriented R-CNN	91.89	92.67	92.71	93.41	93.71	93.02	93.63	94.42	94.06	92.87	92.14	92.16	89.75	92.80	7.20
ARC R-CNN	91.61	92.67	92.99	92.93	94.36	93.20	93.94	94.76	94.31	93.04	92.36	92.39	90.27	92.99	7.38
Ours	92.52	93.18	93.26	93.36	94.72	93.27	94.11	94.75	94.53	93.45	92.58	93.01	91.28	**93.39**	**8.22**

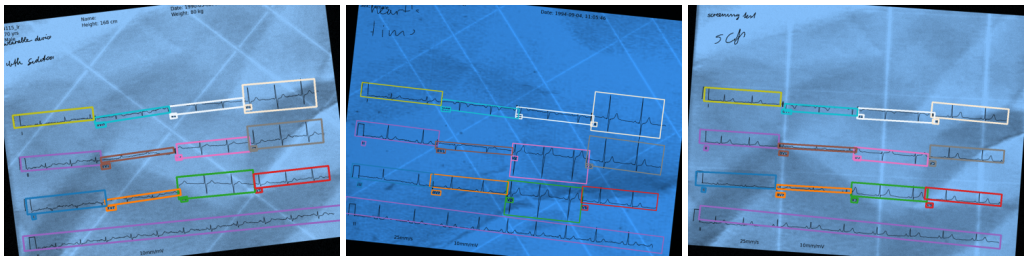
The ARC R-CNN model is used as the baseline model. The proposed model performs better in detecting most categories than other oriented R-CNN-based rotation object detection models. However, we observed a lower accuracy in predicting the names of lead I and II. This issue can be attributed to these two classes' relatively small pixel occupancy, making them difficult to distinguish from image noise in many samples. Visualization results of detection and digitization for three ECG images from the test set are shown in Figure 4.



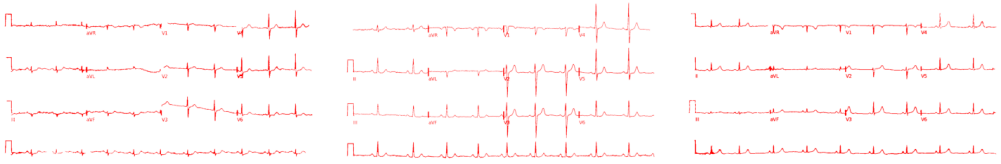
(1) Ground truth



(2) Baseline



(3) Ours



(4) Digitized ECGs

Figure 4. The visualization results of detection and digitization for three ECG images with different noise levels and rotation angles are presented. The rows show ground truth (**1**), baseline predictions (**2**), our detection model's results (**3**), and digitized ECGs (**4**). Curves of all leads with their names are marked with distinct color boxes.

The first row corresponds to the ground truth annotations. Curves of all leads with their names are marked with distinct color boxes. The second row displays the bounding box predictions generated by the baseline model (i.e., oriented R-CNN with ARC modules). The results obtained by our detection model are shown in the third row. The fourth row illustrates the digitized ECG signals. It can be observed that, compared with the baseline, our method can more accurately detect all targets, especially the curves of each ECG lead. Taking the 10 s lead II waveform as an example, the bounding boxes identified by our model are closer to the ground truth. For the second image, the baseline misses the names of leads I and II, while our model identifies the bounding box for the name of lead II but also misses the name of lead I. Upon examination of the digitized ECG signals, minor data distortions can be observed, namely, partial discontinuities within the signal waveforms. This issue reduces the accuracy of the reconstructed digital signal, mainly due to the high-level noise in the image, causing some curves to be misclassified as noise. Overall, the digitized ECGs can recover the waveform characteristics of the original signal.

To further assess the similarity between the digitized ECG and the original signal, we extracted the 10 s lead II and identified the P waves, QRS complexes, T waves, and R peaks [55]. As shown in Figure 5, we examined the Pearson correlation coefficients [56] of seven indices, including QRS duration, PR interval, RR interval, QT interval, P wave amplitude, R peak amplitude, and T wave amplitude.

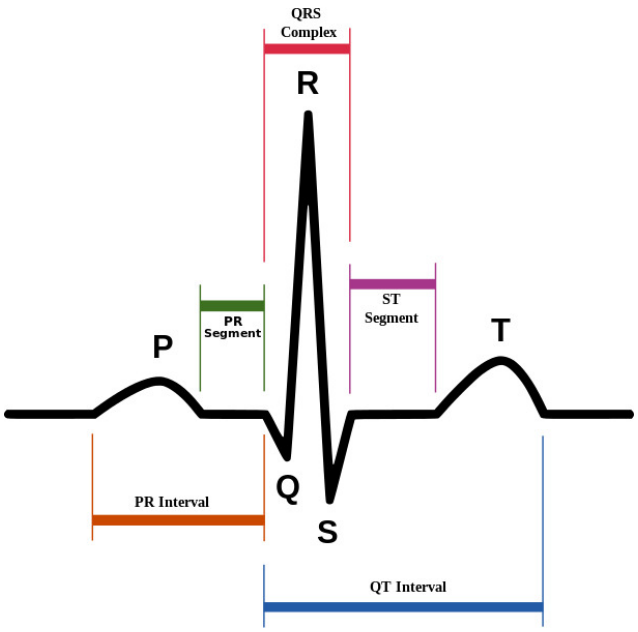


Figure 5. A graphical depiction of fundamental ECG waves and intervals.

Table 3 presents the experimental results. It indicates that all correlation coefficients for intervals and peak amplitudes fall between 0.901 and 0.981, suggesting the effectiveness of the digitalization method. Notably, the QRS duration and RR interval, which are crucial for clinical diagnosis of cardiovascular diseases, exhibited the highest correlation coefficients, suggesting that the proposed digitization method is reliable for assisting clinical diagnosis. The slightly lower correlation coefficients for P wave and T wave amplitudes might be attributed to slight data loss during digitization.

Table 3. Correlation coefficients between the original and digitized ECGs for seven ECG characteristics.

Index	Correlation Coefficient	*p*-Value	95% Confidence Interval	
			LCB	UCB
QRS (ms)	0.979	<0.001	0.947	0.989
PR (ms)	0.933	<0.001	0.909	0.959
RR (ms)	0.981	<0.001	0.978	0.985
QT (ms)	0.936	<0.001	0.927	0.944
P (mV)	0.906	<0.002	0.863	0.949
R (mV)	0.959	<0.001	0.939	0.972
T (mV)	0.901	<0.002	0.866	0.938

Classification

We trained a classification model, MFECGFormer, for abnormal ECG signal identification, configured with parameters detailed in Appendix B Table A3. Trained on the real data and the synthetic data (Appendix C) for 50 epochs, our model outperforms other counterparts, including Trans [17], SSM_ECG [18], and MVMS [19], on the real test set. Trans employs four self-attention encoders, each composed of three multi-head attention layers and a ResNet18 module to model ECG features. MVMS adopts a multi-view approach to capture multi-scale features from each lead and utilizes knowledge distillation to reduce model parameters. It outperforms other models on the 5-way classification

task of the PTB-XL dataset. SSM_ECG is a novel deep structural state space model for ECG classification.

In this paper, we conducted experiments using original and digitized ECG data as ground truth. We trained and evaluated our model on an identical test set. As shown in Table 4, the proposed model consistently outperforms other methods. However, the classification accuracy of the digitized data is relatively lower than that of the original data, suggesting that information loss and potential biases introduced during digitization may hinder performance.

Table 4. Predictive performance comparison of baseline methods and our classification model on the test set.

Methods		AUROC	Precision	Recall	F1-Score	Accuracy
Trans [17]	original	0.921	0.730	0.756	0.741	0.914
	digitized	0.909	0.713	0.725	0.717	0.904
SSM_ECG [18]	original	0.920	0.737	0.770	0.752	0.915
	digitized	0.919	0.732	0.758	0.744	0.912
MVMS [19]	original	0.949	0.813	0.845	0.827	0.943
	digitized	0.922	0.752	0.762	0.757	0.918
Ours	original	0.951	0.831	0.846	0.838	0.947
	digitized	**0.927**	**0.775**	**0.788**	**0.781**	**0.926**

Furthermore, the confusion matrices of different models are illustrated in Figure 6, in which each row shows the predicted distribution for the corresponding category. These results indicate that all models tend to misclassify samples as NORM. This observation aligns with clinical findings where certain ECGs are labeled as both NORM and other abnormal patterns. Moreover, the relatively low recall for the STTC category is mainly attributed to its small proportion in the test set (8.9%). However, although the proportion of the MI category (15.8%) in the test set is comparable to other abnormal categories, its recall is higher, potentially indicating that the proposed model is more effective in learning the characteristic ECG morphologies associated with this category.

The proposed model achieves higher classification accuracy, but this comes at the expense of increased computational cost. As shown in Table 5, the proposed model, based on the transformer framework, exhibits higher parameter counts and computational complexity compared with MVMS and SSM_ECG. This may limit its applicability in real-time scenarios. Notably, while SSM_ECG demonstrates slightly lower classification performance, its utilization of the Hippo theory and Fast Fourier Transform for convolutional computations results in significantly higher computational efficiency, rendering it more suitable for deployment on the embedded devices.

Table 5. For each model, the number of parameters, FLOPs, and inference time are reported.

Methods	Params (10^6)	FLOPs (10^6)	Inference (ms)
Trans	13.64	103.93	11.91
SSM_ECG	2.15	**4.61**	**6.36**
MVMS	**0.39**	82.27	7.04
Ours	6.72	69.33	8.12

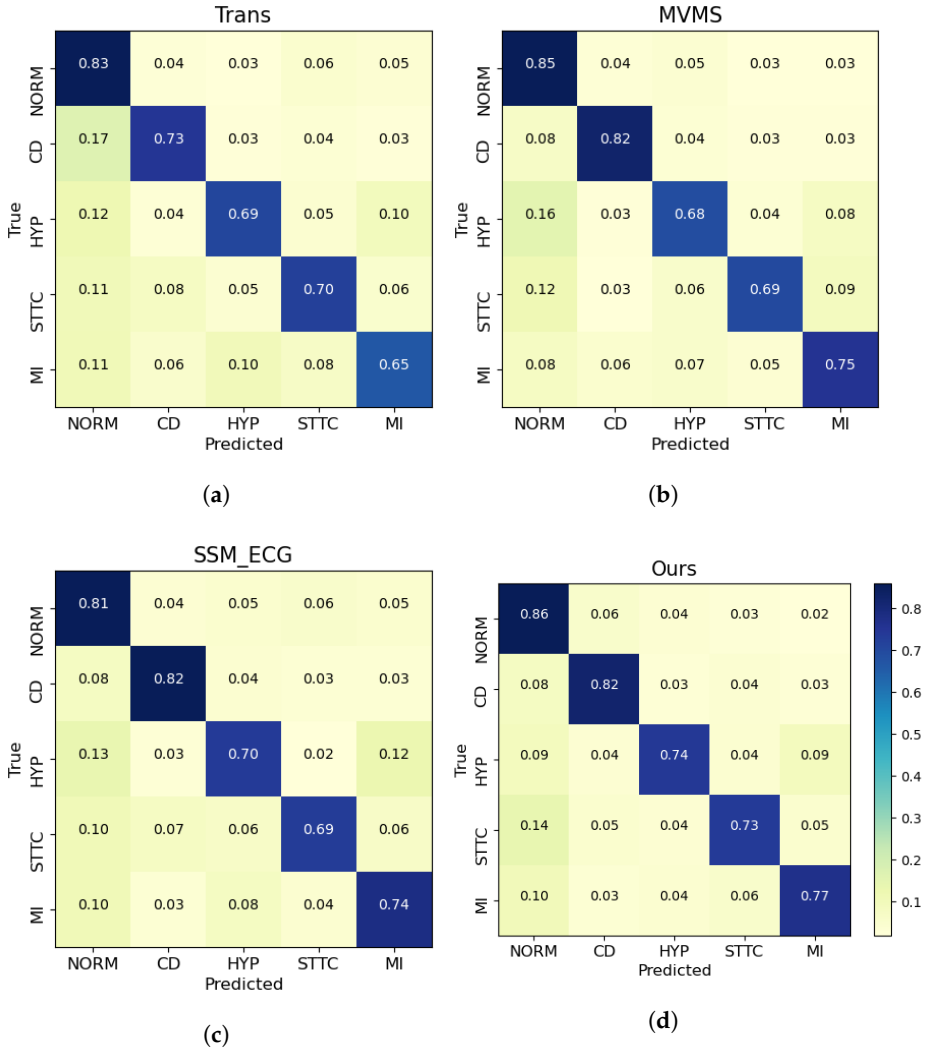


Figure 6. The figure presents the confusion matrices for four classification models (Trans, MVMS, SSM_ECG, and Ours) evaluated on the same test set. The x-axis and y-axis denote the predicted and actual categories, respectively. Each row shows the predicted distribution for the corresponding category.

5. Limitation and Future Work

While the proposed models demonstrate promising results, there are several limitations. First, the model exhibits lower accuracy in detecting the names of lead I and II in paper ECGs. Second, a small amount of information loss occurs during the digitization process, which can affect the accuracy of the reconstructed ECG signals and the classification performance of abnormal samples. These issues are primarily due to high-level noise in paper ECGs. Future work should investigate more robust denoising techniques or curve segmentation models. Third, our model's performance depends on a specific lead layout on paper ECGs. Future research should explore methods to adapt to various layouts. Additionally, the computational efficiency of the classification model needs to be improved.

6. Conclusions

This work proposes a systematic approach for automatically identifying abnormal samples in paper ECGs characterized by rotations, noise, and creases. Our method introduces paperECGLoss, which utilizes the symmetric distribution of 12 leads in paper ECGs to enhance detection model's performance. The reliability of the digitized ECGs is evaluated using multiple metrics. Then, we build an S4-augmented transformer for classification, combining handcrafted features. Using the PTB-XL dataset, we comprehensively

evaluate the proposed models. The results demonstrate that the model effectively improves the detection accuracy of ECG waveforms and achieves competitive results on the five-class classification task compared with other counterparts. Our proposed method offers a promising solution for the automated analysis of paper ECGs. This model significantly improves diagnostic efficiency by automating the identification of critical cases. It enhances diagnostic accuracy by detecting subtle signal variations and reduces the risk of human error. Furthermore, it facilitates the digitization of paper ECGs for large-scale analysis.

Author Contributions: Conceptualization, X.W. and J.Y.; methodology, X.W.; software, X.W.; formal analysis, X.W.; writing—original draft preparation, X.W.; writing—review and editing, X.W. and J.Y.; funding acquisition, J.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by 19ZR1476300.

Data Availability Statement: The data used in this paper are all from public datasets. Data are contained within the article.

Acknowledgments: We thank Xiaolin Huang for the suggestions to improve the methodology and helping to refine the English writing.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Appendix A. Data Extraction

Table A1. Handcrafted Features.

Name	Description
SDNN	Standard deviation of NN (normal-to-normal heartbeat) intervals.
RMSSD	Root mean square of successive differences between normal heartbeats.
NN50	The number of successive NN intervals with a difference exceeding 50 ms.
pNN50	The ratio of NN50 to the total number of NN intervals.
DSD	Standard deviation of successive differences between adjacent NN intervals.
Mean NN	Mean of NN intervals.
CVNN	Coefficient of variation of NN intervals, calculated as the ratio of SDNN to mean NN.
CVSD	Coefficient of variation of successive differences, calculated as the ratio of RMSSD to mean NN.
HR_min	The minimum heart rate value
HR_max	The maximum heart rate value
HTI	Heart rate turbulence index
TINN	Application of triangular interpolation to the histogram of NN intervals
PSEmedian	The median of the Shannon entropy values of the P wave segments.
PAEmedian	The median of the approximate entropy values of the P wave segments.
PPEmedian	The median of the permutation entropy values of the P wave segments.
RAE	The approximate entropy of the R wave segments.
RSE	The Shannon entropy of the R wave segments.
RPE	The permutation entropy of the R wave segments.
SWTL2E	The entropy value of the signal after being decomposed to level 2 using SWT.
fftmean	$Z_1 = \frac{1}{N}\sum_{k=1}^{N} F(k)$
fftvar	$Z_2 = \frac{1}{N-1}\sum_{k=1}^{N}(F(k) - Z_1)^2$

Table A1. *Cont.*

Name	Description
fftentropy	$Z_3 = -1 \times \sum_{k=1}^{N} (\frac{F(k)}{Z_1 N} log_2 \frac{F(k)}{Z_1 N})$
fftenergy	$Z_4 = \frac{1}{N} \sum_{k=1}^{N} (F(k))^2$
fftskew	$Z_5 = \frac{1}{N} \sum_{k=1}^{N} (\frac{F(k)-Z_1}{\sqrt{Z_2}})^3$
fftkurt	$Z_6 = \frac{1}{N} \sum_{k=1}^{N} (\frac{F(k)-Z_1}{\sqrt{Z_2}})^4$
age	
sex	

Appendix B. Hyperparameters

Table A2. Hyperparameters for ECG-ARCResNet.

Hyperparameter	Value
Backbone	ResNet-50-FPN
Rotation angles' number	512
FPN: pyramid levels	4
RPN: anchor scales RPN: anchor ratios RPN: IoU threshold RPN: NMS threshold	[16, 32, 64] [0.5, 1.0, 2.0] 0.3, 0.7 0.4
ROI: fc	1024

Table A3. Hyperparameters for MFECGFormer.

Hyperparameter	Value
Embedding size	512
S4: state_num Encoder layers d_model header_num	64 3 256 4
MLP: fc Activation	1024 GELU

Appendix C. Data Augmentation

Given the scarcity of balanced datasets for CDVs, this work employs a diffusion model with structured state-space architecture (SSSD-ECG) to augment the training set. This method, initially proposed by [39], has presented a conditional generative model for 71 ECG classes. The classification model trained on the generated ECG data performs better than other generative models, including a conditional generative adversarial network (CGAN), despite a slight performance gap relative to the model trained on real data.

In this work, we reproduce this model and introduce slight modifications to generate samples for four abnormal classes. Figure A1 illustrates the framework, which comprises three stacked SSSD layers, each containing two S4 modules. The S4 module better captures long-term dependencies in time series, which has been proven effective in various time series prediction tasks [42]. We add more skip connections (red lines) in each SSSD layer to make the model robust.

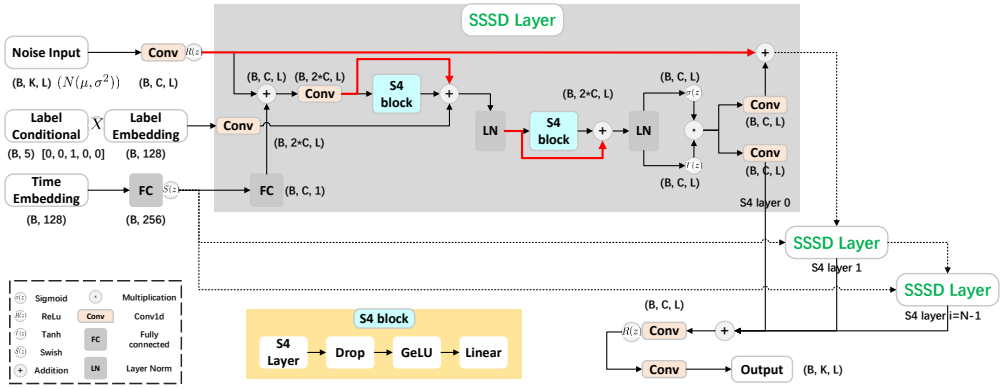


Figure A1. The architecture of conditional SSSD-ECG.

To handle the imbalanced class distribution in the original dataset (Table A4), we trained the SSSD-ECG model to generate synthetic ECG samples, following the hyperparameters in Table A5.

Table A4. Statistics of the original and augmented datasets.

Category	NORM	MI	STTC	CD	HYP
Original	46.3%	15.8%	8.9%	15.3%	13.7%
Augmented	32.2%	16.9%	16.9%	16.9%	16.9%

Table A5. Hyperparameters for SSSD-ECG.

Hyperparameter	Value
S4 layers	36
S4: state_num	64
Residual channels	256
Diffusion embedding fc1	128
Diffusion embedding fc2	256
Diffusion embedding fc3	256
Schedule	Linear
Diffusion steps T	200
B_0	0.0001
B_1	0.02
Loss function	MSE

Figure A2 illustrates the generated ECG samples for each class. The generated samples generally replicate the characteristics of real ECG data, including trends and QRS complex counts. To mitigate the potential negative impact of oversampling, we balanced the class distribution by expanding the minority classes to approximately half the size of the *Norm* class.

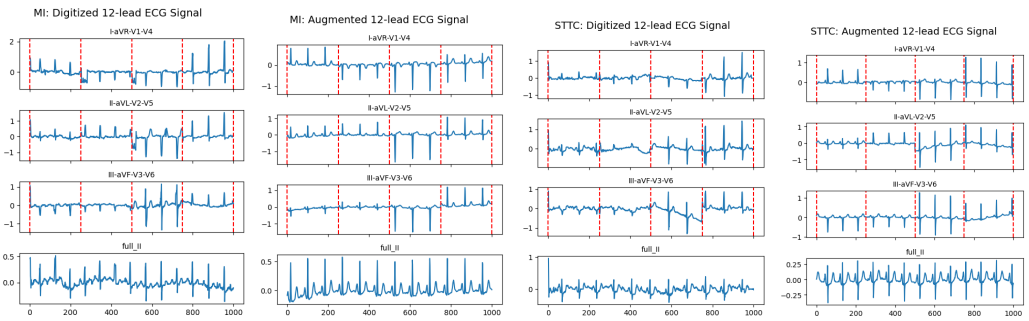


Figure A2. *Cont.*

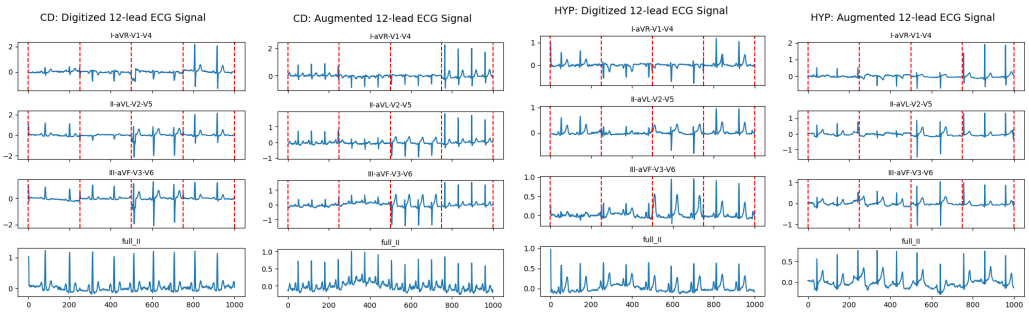


Figure A2. Comparison of original and generated data for four abnormal classes. Each image comprises 12 lead curves of 2.5 s duration (separated by red dashed lines) and a 10 s waveform of Lead II.

References

1. Li, H.; Lin, Z.; An, Z.; Zuo, S.; Zhu, W.; Zhang, Z.; Mu, Y.; Cao, L.; Garcia, J.D.P. Automatic electrocardiogram detection and classification using bidirectional long short-term memory network improved by Bayesian optimization. *Biomed. Signal Process. Control.* **2022**, *73*, 103424. [CrossRef]
2. Murat, F.; Yildirim, O.; Talo, M.; Baloglu, U.B.; Demir, Y.; Acharya, U.R. Application of deep learning techniques for heartbeats detection using ECG signals-analysis and review. *Comput. Biol. Med.* **2020**, *120*, 103726. [CrossRef]
3. Tison, G.H.; Zhang, J.; Delling, F.N.; Deo, R.C. Automated and interpretable patient ECG profiles for disease detection, tracking, and discovery. *Circ. Cardiovasc. Qual. Outcomes* **2019**, *12*, e005289. [CrossRef] [PubMed]
4. Shivashankara, K.K.; Shervedani, A.M.; Clifford, G.D.; Reyna, M.A.; Sameni, R.; et al. ECG-Image-Kit: A synthetic image generation toolbox to facilitate deep learning-based electrocardiogram digitization. *Physiol. Meas.* **2024**, *45*, 055019. [CrossRef]
5. Öztürk, S.; Şahin, S.A.; Aksoy, A.N.; Ari, B.; Akinbi, A. A novel approach for cardiotocography paper digitization and classification for abnormality detection. *IEEE Access* **2023**, *11*, 42521–42533. [CrossRef]
6. Sun, X.; Li, Q.; Wang, K.; He, R.; Zhang, H. A Novel Method for ECG Paper Records Digitization. In Proceedings of the 2019 Computing in Cardiology (CinC), Singapore, 8–11 September 2019; pp. 1–4. [CrossRef]
7. Wu, H.; Patel, K.H.K.; Li, X.; Zhang, B.; Galazis, C.; Bajaj, N.; Sau, A.; Shi, X.; Sun, L.; Tao, Y.; et al. A fully-automated paper ECG digitisation algorithm using deep learning. *Sci. Rep.* **2022**, *12*, 20963. [CrossRef]
8. Randazzo, V.; Puleo, E.; Paviglianiti, A.; Vallan, A.; Pasero, E. Development and Validation of an Algorithm for the Digitization of ECG Paper Images. *Sensors* **2022**, *22*, 7138. [CrossRef] [PubMed]
9. Ravichandran, L.; Harless, C.; Shah, A.J.; Wick, C.A.; Mcclellan, J.H.; Tridandapani, S. Novel tool for complete digitization of paper electrocardiography data. *IEEE J. Transl. Eng. Health Med.* **2013**, *1*, 1800107–1800107. [CrossRef]
10. Baydoun, M.; Safatly, L.; Abou Hassan, O.K.; Ghaziri, H.; El Hajj, A.; Isma'eel, H. High precision digitization of paper-based ECG records: A step toward machine learning. *IEEE J. Transl. Eng. Health Med.* **2019**, *7*, 1–8. [CrossRef] [PubMed]
11. Mishra, S.; Khatwani, G.; Patil, R.; Sapariya, D.; Shah, V.; Parmar, D.; Dinesh, S.; Daphal, P.; Mehendale, N. ECG paper record digitization and diagnosis using deep learning. *J. Med Biol. Eng.* **2021**, *41*, 422–432. [CrossRef] [PubMed]
12. Yu, X.; Huang, Y.; Wu, J.; Wang, J.; Cai, W. From Paper to Digital: ECG Processing with U-Net Digitization and ResNet Classification. In Proceedings of the 51st Computing in Cardiology Conference, Karlsruhe, Germany, 8–11 September 2024.
13. Li, Y.; Qu, Q.; Wang, M.; Yu, L.; Wang, J.; Shen, L.; He, K. Deep learning for digitizing highly noisy paper-based ECG records. *Comput. Biol. Med.* **2020**, *127*, 104077. [CrossRef]
14. Petmezas, G.; Papageorgiou, V.E.; Vassilikos, V.; Pagourelias, E.; Tsaklidis, G.; Katsaggelos, A.K.; Maglaveras, N. Recent advancements and applications of deep learning in heart failure: A systematic review. *Comput. Biol. Med.* **2024**, *176*, 108557. [CrossRef] [PubMed]
15. Ebrahimi, Z.; Loni, M.; Daneshtalab, M.; Gharehbaghi, A. A review on deep learning methods for ECG arrhythmia classification. *Expert Systems with Applications: X* **2020**, *7*, 100033. [CrossRef]
16. Pessoa, D.; Petmezas, G.; Papageorgiou, V.E.; Rocha, B.M.; Stefanopoulos, L.; Kilintzis, V.; Maglaveras, N.; Frerichs, I.; de Carvalho, P.; Paiva, R.P. Pediatric Respiratory Sound Classification Using a Dual Input Deep Learning Architecture. In Proceedings of the 2023 IEEE Biomedical Circuits and Systems Conference (BioCAS), Toronto, ON, Canada, 19–21 October 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1–5.
17. Natarajan, A.; Chang, Y.; Mariani, S.; Rahman, A.; Boverman, G.; Vij, S.; Rubin, J. A wide and deep transformer neural network for 12-lead ECG classification. In Proceedings of the 2020 Computing in Cardiology, Rimini, Italy, 13–16 September 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1–4.

18. Mehari, T.; Strodthoff, N. Advancing the state-of-the-art for ECG analysis through structured state space models. *arXiv* **2022**, arXiv:2211.07579.
19. Yang, S.; Lian, C.; Zeng, Z.; Xu, B.; Zang, J.; Zhang, Z. A multi-view multi-scale neural network for multi-label ECG classification. *IEEE Trans. Emerg. Top. Comput. Intell.* **2023**, *7*, 648–660. [CrossRef]
20. Chen, S.W.; Wang, S.L.; Qi, X.Z.; Samuri, S.M.; Yang, C. Review of ECG detection and classification based on deep learning: Coherent taxonomy, motivation, open challenges and recommendations. *Biomed. Signal Process. Control.* **2022**, *74*, 103493. [CrossRef]
21. Yang, X.; Yang, J.; Yan, J.; Zhang, Y.; Zhang, T.; Guo, Z.; Sun, X.; Fu, K. Scrdet: Towards more robust detection for small, cluttered and rotated objects. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 8232–8241.
22. Yang, S.; Luo, P.; Loy, C.C.; Tang, X. Wider face: A face detection benchmark. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 5525–5533.
23. Xia, G.S.; Bai, X.; Ding, J.; Zhu, Z.; Belongie, S.; Luo, J.; Datcu, M.; Pelillo, M.; Zhang, L. DOTA: A large-scale dataset for object detection in aerial images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 3974–3983.
24. Xu, Y.; Fu, M.; Wang, Q.; Wang, Y.; Chen, K.; Xia, G.S.; Bai, X. Gliding vertex on the horizontal bounding box for multi-oriented object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *43*, 1452–1459. [CrossRef] [PubMed]
25. Li, W.; Chen, Y.; Hu, K.; Zhu, J. Oriented reppoints for aerial object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 1829–1838.
26. Yang, X.; Zhang, G.; Yang, X.; Zhou, Y.; Wang, W.; Tang, J.; He, T.; Yan, J. Detecting rotated objects as gaussian distributions and its 3-d generalization. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 4335–4354. [CrossRef]
27. Qian, W.; Yang, X.; Peng, S.; Yan, J.; Guo, Y. Learning modulated loss for rotated object detection. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtually, 2–9 February 2021; Volume 35, pp. 2458–2466.
28. Yang, X.; Yang, X.; Yang, J.; Ming, Q.; Wang, W.; Tian, Q.; Yan, J. Learning high-precision bounding box for rotated object detection via kullback-leibler divergence. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 18381–18394.
29. Yang, X.; Zhou, Y.; Zhang, G.; Yang, J.; Wang, W.; Yan, J.; Zhang, X.; Tian, Q. The KFIoU loss for rotated object detection. *arXiv* **2022**, arXiv:2201.12558.
30. Yang, X.; Yan, J.; Feng, Z.; He, T. R3det: Refined single-stage detector with feature refinement for rotating object. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtually, 2–9 February 2021; Volume 35, pp. 3163–3171.
31. Yang, X.; Yan, J.; Liao, W.; Yang, X.; Tang, J.; He, T. Scrdet++: Detecting small, cluttered and rotated objects via instance-level feature denoising and rotation loss smoothing. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 2384–2399. [CrossRef]
32. Hou, L.; Lu, K.; Xue, J.; Li, Y. Shape-adaptive selection and measurement for oriented object detection. In Proceedings of the AAAI Conference on Artificial Intelligence, Philadelphia, PA, USA, 27 February–2 March 2022; Volume 36, pp. 923–932.
33. Cheng, G.; Wang, J.; Li, K.; Xie, X.; Lang, C.; Yao, Y.; Han, J. Anchor-free oriented proposal generator for object detection. *IEEE Trans. Geosci. Remote. Sens.* **2022**, *60*, 5625411. [CrossRef]
34. Pu, Y.; Wang, Y.; Xia, Z.; Han, Y.; Wang, Y.; Gan, W.; Wang, Z.; Song, S.; Huang, G. Adaptive rotated convolution for rotated object detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 1–6 October 2023; pp. 6589–6600.
35. Alzubaidi, L.; Zhang, J.; Humaidi, A.J.; Al-Dujaili, A.; Duan, Y.; Al-Shamma, O.; Santamaría, J.; Fadhel, M.A.; Al-Amidie, M.; Farhan, L. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *J. Big Data* **2021**, *8*, 44. [CrossRef] [PubMed]
36. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
37. Gu, A.; Goel, K.; Ré, C. Efficiently modeling long sequences with structured state spaces. *arXiv* **2021**, arXiv:2111.00396.
38. Gu, A.; Dao, T.; Ermon, S.; Rudra, A.; Ré, C. Hippo: Recurrent memory with optimal polynomial projections. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 1474–1487.
39. Alcaraz, J.M.L.; Strodthoff, N. Diffusion-based conditional ECG generation with structured state space models. *Comput. Biol. Med.* **2023**, *163*, 107115. [CrossRef] [PubMed]
40. Zhu, L.; Liao, B.; Zhang, Q.; Wang, X.; Liu, W.; Wang, X. Vision mamba: Efficient visual representation learning with bidirectional state space model. *arXiv* **2024**, arXiv:2401.09417.
41. Du, H.; Du, S.; Li, W. Probabilistic time series forecasting with deep non-linear state space models. *CAAI Trans. Intell. Technol.* **2023**, *8*, 3–13. [CrossRef]
42. Alcaraz, J.M.L.; Strodthoff, N. Diffusion-based time series imputation and forecasting with structured state space models. *arXiv* **2022**, arXiv:2208.09399.

43. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, pp. 5998–6008.
44. Lin, T.; Wang, Y.; Liu, X.; Qiu, X. A survey of transformers. *AI Open* **2022**, *3*, 111–132. [CrossRef]
45. Zhou, H.; Zhang, S.; Peng, J.; Zhang, S.; Li, J.; Xiong, H.; Zhang, W. Informer: Beyond efficient transformer for long sequence time-series forecasting. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtually, 2–9 February 2021; Volume 35, pp. 11106–11115.
46. Zhou, T.; Ma, Z.; Wen, Q.; Wang, X.; Sun, L.; Jin, R. Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting. In Proceedings of the International Conference on Machine Learning. PMLR, Baltimore, MD, USA, 17–23 July 2022; pp. 27268–27286.
47. Chen, W.; Wang, W.; Peng, B.; Wen, Q.; Zhou, T.; Sun, L. Learning to rotate: Quaternion transformer for complicated periodical time series forecasting. In Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, 14–18 August 2022; pp. 146–156.
48. Liu, S.; Yu, H.; Liao, C.; Li, J.; Lin, W.; Liu, A.X.; Dustdar, S. Pyraformer: Low-complexity pyramidal attention for long-range time series modeling and forecasting. In Proceedings of the International Conference on Learning Representations, Virtual, 3–7 May 2021.
49. Wu, H.; Xu, J.; Wang, J.; Long, M. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 22419–22430.
50. Beltagy, I.; Peters, M.E.; Cohan, A. Longformer: The long-document transformer. *arXiv Prepr.* **2020**, arXiv:2004.05150.
51. Kitaev, N.; Kaiser, Ł.; Levskaya, A. Reformer: The efficient transformer. *arXiv Prepr.* **2020**, arXiv:2001.04451.
52. Zuo, S.; Liu, X.; Jiao, J.; Charles, D.; Manavoglu, E.; Zhao, T.; Gao, J. Efficient long sequence modeling via state space augmented transformer. *arXiv* **2022**, arXiv:2212.08136.
53. Wagner, P.; Strodthoff, N.; Bousseljot, R.D.; Kreiseler, D.; Lunze, F.I.; Samek, W.; Schaeffter, T. PTB-XL, a large publicly available electrocardiography dataset. *Sci. Data* **2020**, *7*, 154. [CrossRef] [PubMed]
54. Krull, A.; Buchholz, T.O.; Jug, F. Noise2void-learning denoising from single noisy images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 2129–2137.
55. Papageorgiou, V.E.; Zegkos, T.; Efthimiadis, G.; Tsaklidis, G. Analysis of digitalized ECG signals based on artificial intelligence and spectral analysis methods specialized in ARVC. *Int. J. Numer. Methods Biomed. Eng.* **2022**, *38*, e3644. [CrossRef] [PubMed]
56. Georgakis, A.; Papageorgiou, V.E.; Gatziolis, D.; Stamatellos, G. Temporal-Like Bivariate Fay-Herriot Model: Leveraging Past Responses and Advanced Preprocessing for Enhanced Small Area Estimation of Growing Stock Volume. *Oper. Res. Forum* **2024**, *5*, 9. [CrossRef]