Article

Multi-Label Diagnosis of Arrhythmias Based on a Modified Two-Category Cross-Entropy Loss Function

Junjiang Zhu [1,*], Cheng Ma [1], Yihui Zhang [1], Hao Huang [1], Dongdong Kong [2] and Wangjin Ni [1]

[1] College of Mechanical and Electrical Engineering, China Jiliang University, Hangzhou 310018, China; p21010854071@cjlu.edu.cn (C.M.); p22010855036@cjlu.edu.cn (Y.Z.); p22010854109@cjlu.edu.cn (H.H.); 2100110212@cjlu.edu.cn (W.N.)
[2] School of Mechatronic Engineering and Automation, Shanghai University, Shanghai 200444, China; kodon180704@shu.edu.cn
* Correspondence: zhujunjiang@cjlu.edu.cn

Abstract: The 12-lead resting electrocardiogram (ECG) is commonly used in hospitals to assess heart health. The ECG can reflect a variety of cardiac abnormalities, requiring multi-label classification. However, the diagnosis results in previous studies have been imprecise. For example, in some previous studies, some cardiac abnormalities that cannot coexist often appeared in the diagnostic results. In this work, we explore how to realize the effective multi-label diagnosis of ECG signals and prevent the prediction of cardiac arrhythmias that cannot coexist. In this work, a multi-label classification method based on a convolutional neural network (CNN), long short-term memory (LSTM), and an attention mechanism is presented for the multi-label diagnosis of cardiac arrhythmia using resting ECGs. In addition, this work proposes a modified two-category cross-entropy loss function by introducing a regularization term to avoid the existence of arrhythmias that cannot coexist. The effectiveness of the modified cross-entropy loss function is validated using a 12-lead resting ECG database collected by our team. Using traditional and modified cross-entropy loss functions, three deep learning methods are employed to classify six types of ECG signals. Experimental results show the modified cross-entropy loss function greatly reduces the number of non-coexisting label pairs while maintaining prediction accuracy. Deep learning methods are effective in the multi-label diagnosis of ECG signals, and diagnostic efficiency can be improved by using the modified cross-entropy loss function. In addition, the modified cross-entropy loss function helps prevent diagnostic models from outputting two arrhythmias that cannot coexist, further reducing the false positive rate of non-coexisting arrhythmic diseases, thereby demonstrating the potential value of the modified loss function in clinical applications.

Keywords: multi-label diagnosis; deep learning; cross-entropy loss function

1. Introduction

Cardiovascular disease (CVD) is one of the leading causes of death, accounting for over 31% of deaths worldwide [1]. There are many types of cardiovascular diseases, and their impact on human health also varies. Determining the type of CVD plays an important role in follow-up treatment. In the clinic, one of the most commonly used methods to diagnose CVD is the resting electrocardiogram (ECG). Medical personnel place electrodes at fixed positions on the resting patient to acquire and select a high-quality 10 s ECG and make a diagnosis based on the ECG waveform. According to incomplete statistics, there are more than 100 kinds of cardiovascular diseases, and the detection of ECGs depends on the diagnostic experience of medical professionals. Therefore, it is very important to develop ECG-based diagnostic tools.

Most early ECG diagnostic tools were realized by imitating the logical conclusions of the physician. Geddes et al. [1] proposed classifying various premature ventricular contractions

(PVC) using rule-based reasoning. First, the parameters for detection were selected according to the ECG characteristics of PVC, such as the R-R interval and the duration and shape of the QRS complex. Then, certain medical rules were used as criteria for assessing the occurrence of PVC. Kezdi et al. [2] proposed an algorithm for detecting ectopic beats and arrhythmia based on clinical experience. The R-wave was determined by calculating the slope of the QRS complex. Supraventricular tachycardia and ventricular ectopy were detected by calculating the changes in the R-R interval and the width, polarity, and height of the QRS complex. The parameters selected for these methods are clinically interpretable. However, other feature extraction methods (except for R-wave) are not accurate enough because of the strong personalization and nonlinearity of ECG signals, especially in different types of arrhythmias. Since different types of ECG signals have different time-frequency features, large errors can easily occur in the calculation of feature parameters, leading to the failure of this type of method.

Another type of method is pattern recognition. First, certain statistical features are extracted, and then a classifier is created using machine learning (ML) to classify different types of arrhythmias. In many studies, time/morphological statistics [3–7], spectral features [8,9], and higher-order statistical parameters [10–13] have been used to diagnose ventricular arrhythmias in malignant arrhythmias. These mathematical features, in combination with classifiers such as artificial neural networks (ANNs) or support vector machines (SVMs) [14–16], can efficiently filter out rhythms such as ventricular fibrillation and ventricular tachycardia. The two steps (i.e., feature extraction and classification) in pattern recognition help in the diagnosis of cardiac arrhythmias. The accuracy and efficiency of detection are better than simulating the physician's logical conclusions. The disadvantage is that the signal features are artificially determined, or more precisely, the quality of the signal features often depends on artificial experience. Therefore, it is difficult to find effective statistical features because there are too many types of arrhythmias.

In recent years, with the development of deep learning, researchers have begun to use deep learning instead of artificial feature extraction methods [17] to evaluate ECG signals. "Artificial feature extraction methods" refer to the methods used to calculate the features of electrocardiogram signals from different perspectives (such as the time domain, frequency domain, and time-frequency domain) for the classification of arrhythmias. The selection of these features is based on personal subjective experience. Feng et al. [18] employed dynamic time warping (DTW), C-means clustering, and the BP algorithm to optimize the parameters of the probabilistic process neural network (PPNN). The method achieved an F1 score of 0.7615 and an accuracy of 74.16% on the Chinese Cardiovascular Disease Database (CCDD). While PPNN offers advantages such as few-shot learning and computational complexity, the limited size of its parameters hampers its classification performance. Yıldırım et al. [19] proposed a new one-dimensional convolutional neural network model (1D CNN) to classify 17 types of cardiac arrhythmias. Its accuracy and F1 score on the MIT-BIH arrhythmia database were 91.33% and 0.8538, respectively. The model demonstrated efficient and rapid diagnostic capabilities. Luo et al. [20] conducted a study using the same database and proposed a hybrid convolutional recurrent neural network (HCRNet), achieving an accuracy of 99.01%. However, the MIT-BIH data were derived from internal patients, and the ECG signals exhibited highly personalized characteristics. Thus, a model with high accuracy might not necessarily possess a high degree of generalizability across different patients. Yao et al. [21] proposed the ATI-CNN model to address the low performance of a CNN in the detection of variable-length ECG signals. This model integrated a CNN, recurrent cells, and an attention module. On the China Physiological Signal Challenge (CPSC) dataset, ATI-CNN achieved an F1 score of 0.812 and a precision of 0.826. By combining the spatiotemporal features of ECG signals, ATI-CNN improved accuracy while reducing the number of model parameters, thereby lowering training costs. However, this model did not consider the one-to-many relationship between patients and arrhythmia labels. Objectively, deep learning methods learn features from a large number of data to classify ECG signals, which will be the development direction of intelligent ECG diagnosis in the future.

In ECG signals, some arrhythmias can occur simultaneously, whereas others do not. For example, in ECG signals of a period of sustained atrial fibrillation, PVCs but not premature atrial fibrillation can occur simultaneously. The relationship between the various designations is complex, making multi-label classification of ECG signals challenging [22–24]. Yoo et al. [25] optimized the algorithm from the perspective of multi-label classification of arrhythmia and proposed xECGNet. By incorporating the L2 norm of attention maps of different disease categories into the loss function, xECGNet achieved a multi-label subset accuracy of 84.6% in the classification tasks of eight types of arrhythmias on the CPSC dataset. Yang et al. [26] proposed using a stacking approach to combine the classification results of ResNet and random forest and obtain the final results through voting. Despite the method's accuracy improving to 95%, integrating multiple models increased deployment costs, making it challenging to apply to general medical embedded devices. Nowadays, current methods emphasize learning the relationships between labels from research data (the labels themselves). However, due to the complex relationships between the labels of ECG signals, it is difficult to learn these relationships from only research data. This causes the diagnostic models to output some arrhythmias that cannot coexist, leading to the increased misdiagnosis rate of the multi-label ECG diagnostic algorithm [27].

In this work, we propose a multi-label diagnostic method based on a modified two-category cross-entropy loss function. This method first incorporates LSTM and attention mechanisms to enhance the classification accuracy of the CNN model. Building upon this, to address the issue of certain conclusions being unable to coexist in arrhythmia diagnosis, we add a regularization term to the traditional binary cross-entropy loss function, which disallows the coexistence of certain arrhythmia disease label pairs. The regularization term helps constrain the network's learning direction, enabling it to consider the mutually exclusive relationships between various disease labels. It improves the applicability of the ECG diagnostic algorithm in real-life diagnosis scenarios.

The main innovative points of this article are:

(A) A new multi-label training loss function is proposed by adding a regularization term that does not allow the coexistence of some arrhythmias;

(B) A CNN + LSTM + ATTENTION architecture is presented to improve ECG classification performance;

(C) More than 10,000 ECG recordings of the six most common cardiac arrhythmias are used to test the loss function and classification method, and the performance is compared between patients. Our method improves the accuracy of classifying four types of arrhythmias (normal, sinus tachycardia, atrial flutter, and atrial tachycardia) and reduces the incidence of misdiagnosing atrial flutter and atrial tachycardia as false positives.

This paper is organized as follows. In Section 2, explanations of the CNN + LSTM + ATTENTION architecture and the modified cross-entropy loss function are presented. The new ECG database is described in Section 3. An analysis of the modified cross-entropy loss function and its comparison with other methods are described in Section 4. Further details of the presented method and future research topics are given in Section 5. Section 6 presents the conclusions of this paper.

2. Proposed Method

2.1. Deep Learning Model

In this work, a deep learning model, consisting of a convolutional neural network (CNN) [28], long short-term memory (LSTM) [29], and an attention mechanism [30] is used to classify ECG signals.

2.1.1. Feature Extraction

A CNN is used for feature extraction, as shown in Figure 1. For the convolution operation in the CNN, it is assumed that z_j^i represents the j-th channel output of the i-th

convolutional layer and o_j^l is the input. The input o_j^l and the output z_j^l of the l-th layer can be expressed by Equation (1) and Equation (2), respectively.

$$o_j^l = f(z_j^l), \tag{1}$$

$$z_j^l = \sum_{i \in M_j} x_i^{l-1} * k_{ij}^l + b_j^l, \tag{2}$$

where $f(\cdot)$ is the activation function, M_j is the subset of the feature map of the $(l-1)$-th layer, k_{ij}^l is the convolution kernel matrix, b_j^l is the bias, and '*' is the convolution symbol.
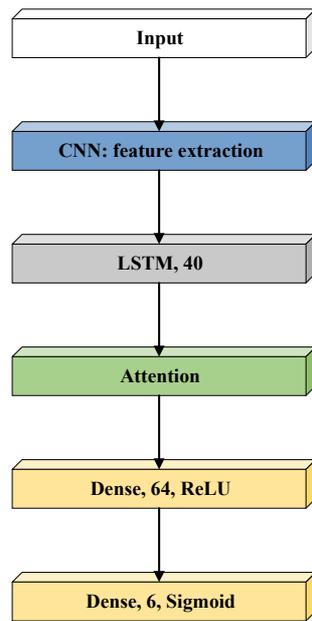

Figure 1. Structure of the deep learning model for multi-label diagnosis of cardiac arrhythmias.

For the pooling operation in the CNN, α stands for the sampling coefficient and represents the maximum pooling ($\cdot$) function. The input o_j^{l+1} and the output z_j^{l+1} of the $(l+1)$-th layer can be expressed by Equation (3) and Equation (4), respectively.

$$o_j^l = f(z_j^l)\epsilon, \tag{3}$$

$$z_j^{l+1} = \alpha_i^{l+1} * Max\mathcal{P}ooling(x_{ij}^l) + b_j^{l+1}, \tag{4}$$

2.1.2. LSTM

The features $Z \in \mathcal{R}^{T \times D}$ obtained by the CNN are input to the following LSTM, where T is the length of the input features and D is the number of input features. The workflow is shown in Figure 2.

The internal state $C_t \in \mathbb{R}^S$ between the units in the LSTM layer is used to determine the relationship between the ECG features extracted by the CNN. S represents the length of the vector output from the LSTM layer. z_t represents the t-th slice in the group of input features ($1 \leq t \leq T$). $h_t \in \mathbb{R}^S$ represents the hidden state of the LSTM layer corresponding to z_t. The final output h_t can be calculated as follows:

$$f_t = \sigma(W_f[h_{t-1}, z_t] + b_f), \tag{5}$$

$$i_t = \sigma(W_i[h_{t-1}, \mathcal{Z}_t] + b_i), \tag{6}$$

$$\tilde{C}_t = tanh(W_{\tilde{C}}[h_{t-1}, z_t] + b_{\tilde{C}}), \tag{7}$$

$$O_t = \sigma(W_0[h_{t-1}, z_t] + b_o), \tag{8}$$

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t, \tag{9}$$

$$h_t = O_t \times tanh(C_t), \tag{10}$$

where f_t, i_t, and O_t represent the update results of the forget gate, input gate, and output gate, respectively. W_f, W_i, $W_{\tilde{c}}$, and W_o represent the weights of the forget gate, input gate, output gate, and LSTM state unit, respectively.
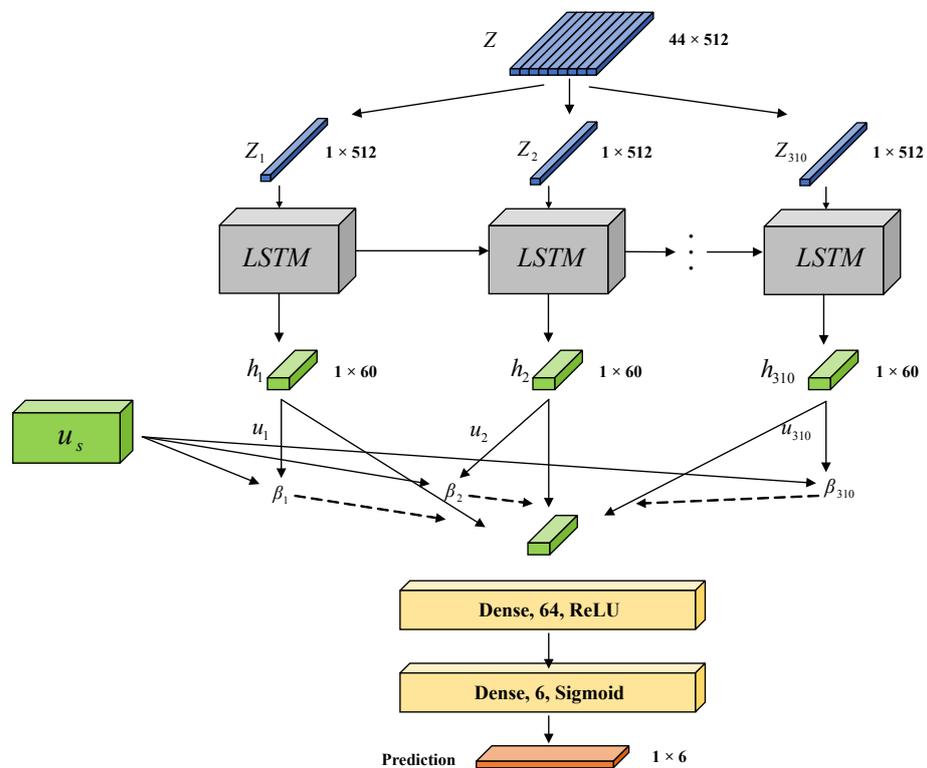

Figure 2. Structure of LTSM and attention mechanism.

2.1.3. Attention Mechanism

The attention mechanism is used to compute the attention distribution in the hidden state h_t ($1 \leq t \leq T$) at each time point. The final output features are then formed by the weighted average of the attention distribution. The computational process is illustrated below:

$$u_i = tan\,h(W_s \cdot h_i + b_s), \tag{11}$$

$$\beta_i = \frac{e^{u_i^T u_s}}{\sum_i e^{u_i^T u_s}}, \tag{12}$$

$$\tilde{z} = \sum_{i=1}^{s} \beta_i h_i, \tag{13}$$

where $\tilde{Z} \in \mathbb{R}^S$ represents the results after the weighted average, β_i represents the weighting factor in the hidden state h_i ($1 \leq t \leq T$) , b_s and W_s are both trainable weights, W_s

represents the query vector, and u_i ($1 \leq t \leq T$) represents the intermediate weighting factor in the calculation.

2.1.4. Fully Connected Layer

Finally, the features $\tilde{Z}$ obtained by the attention mechanism are input to the fully connected layer to perform the final classification. The final prediction vector z is obtained as follows:

$$z^1 = f(W^1 \cdot \tilde{z} + b^1), \tag{14}$$

$$z = Sigmiod(W^2 \cdot z^1 + b^2), \tag{15}$$

where z^1 and z^2 each represent a weighing matrix in the fully connected layer. b^1 and b^2 represent the bias matrices. z^1 represents the output of the first fully connected layer, and *Sigmiod* represents the activation function.

2.2. The Modified Cross-Entropy Loss Function

In multi-label classification, a two-category cross-entropy loss function is usually used to calculate the loss between the labels and the predicted outcomes. In this work, two types of cross-entropy loss functions are studied, given by Equations (16) and (17).

$$loss\_1 = \frac{1}{n} \sum_{k=1}^{N} [y_k ln(a_k) + (1 - y_k) ln(1 - a_k)], \tag{16}$$

$$loss\_2 = \frac{1}{n} \sum_{k=1}^{N} [y_k ln(a_k) + (1 - y_k) ln(1 - a_k)] + \sum_{\substack{l=1 \\ a_i...l \in A_i}}^{M} In\left(1 - \sin\left(\frac{\pi \cdot a_i \cdot a_j}{2}\right)\right), \tag{17}$$

where N is the number of arrhythmia disease types, y_k is the k-th element in the real ECG label vector, and a_k is the k-th element in the predicted ECG label vector. M is the number of combinations belonging to the coexistence of arrhythmia diseases with strong negative correlations, a_l is the l-th combination of arrhythmia diseases that cannot coexist, and a_i and a_j are the i-th element and j-th element, respectively, in the predicted ECG label vector in a_j.

According to Equation (16), $loss\_1$ is the traditional cross-entropy loss function used for multi-label classification and is widely used in deep learning. However, the traditional cross-entropy loss function does not consider the correlations between different labels [31–33]. This results in cardiac arrhythmias, which almost never occur simultaneously, in the predicted results.

According to Equation (17), $loss\_2$ is the modified cross-entropy loss function, obtained by introducing a regularization term. The regularization term increases the penalty of the co-occurrence of cardiac arrhythmias that cannot coexist, which is expected to improve the prediction performance of deep learning models.

Specifically, when the model predicts the presence of non-coexisting arrhythmia disease label pairs in the results due to the logarithmic function's derivative property, it rapidly increases the value of the regularization term, allowing the model to continue training. Conversely, this regularization term tends toward 0, resulting in the degeneration of the loss function into a binary cross-entropy loss function, which does not affect the prediction of other coexisting disease labels.

3. ECG Database

This work is based on the 12-lead ECG data collected by SID MEDICAL TECHNOLOGY CO., LTD from many hospitals in Shanghai. The device used to acquire the ECG signals was the Inno-12 ECG acquisition workstation, as shown in Figure 3. The ECG signals collected were 10 seconds long, and the sampling frequency was 500 Hz. The ECG signals were first magnified 400 times using electrode tabs and then discretized, ensuring the accuracy of acquisition. Considering the power frequency interference, a trap filter was developed in the hardware circuit. Each ECG sample was processed with a Butterworth bandpass filter (0.5~100 Hz) to remove high- and low-frequency noise.

A total of 39,069 data were collected, including six types of ECGs: normal ECG, sinus tachycardia, sinus bradycardia, atrial flutter, atrial tachycardia, and premature ventricular contraction (PVC), as shown in Figure 4. In a 10-second ECG signal, atrial flutter and atrial tachycardia cannot coexist simultaneously. In this work, they are considered non-coexisting arrhythmia disease label pairs. All ECG data were labeled by two professional cardiologists. If the two cardiologists disagreed, the label was determined by a third chief cardiologist. Then, these ECG data were divided into a training dataset (23,322), a validation dataset (2591), and a test dataset (13,156). The distribution of arrhythmias in the different datasets is shown in Table 1.

Table 1. Distribution of cardiac arrhythmias in different datasets.

	Total	Normal	Sinus Tachycardia	Sinus Bradycardia	Atrial Flutter	Atrial Tachycardia	PVC
Training	23,322	9653	4462	4650	4557	2568	4194
Validation	2591	1057	524	555	455	292	458
Test	13,156	3728	2431	4816	2181	250	1692

Note: atrial flutter and atrial tachycardia are non-coexisting arrhythmia disease label pairs.
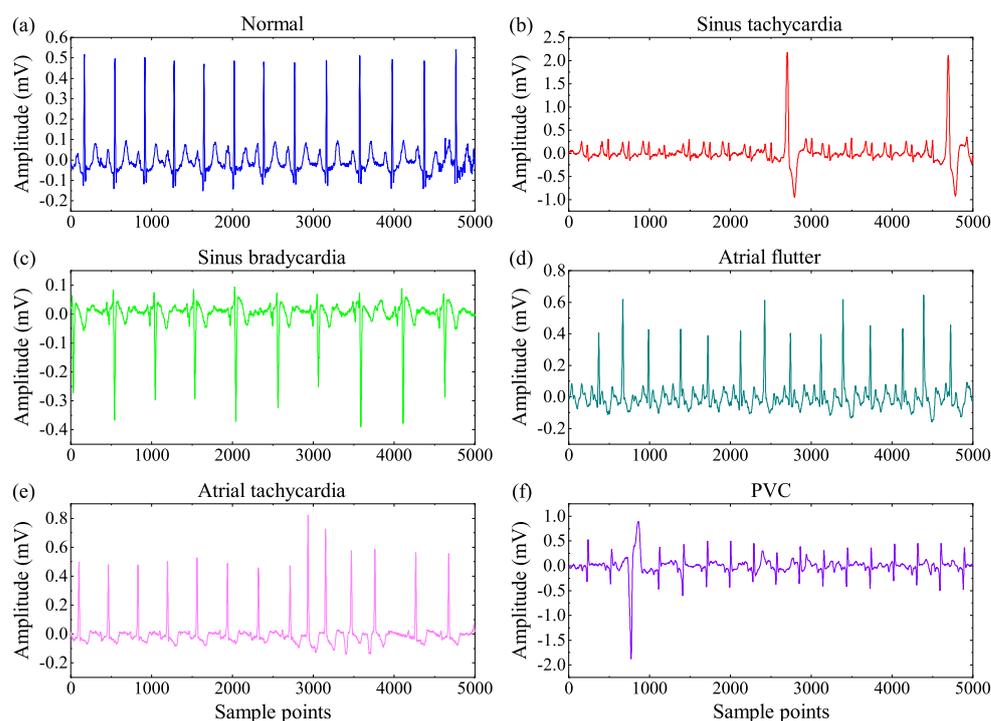

Figure 3. Inno-12 ECG acquisition workstation.

Figure 4. The lead-II ECG signals of different arrhythmias. mV represents voltage. (**a**) Normal; (**b**) Sinus tachycardia; (**c**) Sinus bradycardia; (**d**) Atrial flutter; (**e**) Atrial tachycardia; (**f**) PVC.

4. Experimental Setup and Analysis

4.1. Experimental Setup

In terms of hardware, all experiments were carried out on a Dell T5820 workstation with an Intel Core i9-10900X CPU, 64 GB of RAM, and two graphics cards (NVIDIA RTX 3060 12GB) sourced from Dell in Shanghai, China. In terms of software, all deep learning models were constructed using Numpy 1.19.5, TensorFlow 1.13.1, and Keras 2.2.4, which were installed on Ubuntu 20.04.

4.2. Parameter Setting

4.2.1. The Deep Learning Model

Three CNN models (i.e., 1D VGG16 [34], 1D ResNet34 [35], and 1D ResNet50 [35]) were used to compare whether our proposed method leads to performance improvements in the CNN models, as shown in Figure 5. In our method, only one CNN model is used for feature extraction. The character '/2' in each sub-image means that the stride size in the corresponding network layer is 2. The VGG16 model comprises 16 convolutional layers and adopts the traditional stacked convolution layer approach. Its model structure is relatively deep but simple. The ResNet34 model has 34 convolutional layers and adds residual structures, in contrast to VGG16. It resolves the issue of gradient vanishing during model training by incorporating skip connections that directly add the input to the output. The ResNet50 model, on the other hand, has 50 convolutional layers and utilizes bottleneck structures to reduce computational complexity and improve model efficiency.

Three deep learning models (i.e., VGG16 + LSTM + ATTENTION, ResNet34 + LSTM + ATTENTION, and ResNet34 + LSTM + ATTENTION) were used to verify whether our proposed method leads to performance improvements in the CNN models. The corresponding parameter settings and network structures are shown in Table 2. The input size of all three deep learning models was 5000×12. The output sizes of the three deep learning models were 44×512, 22×512, and 22×2048, respectively.
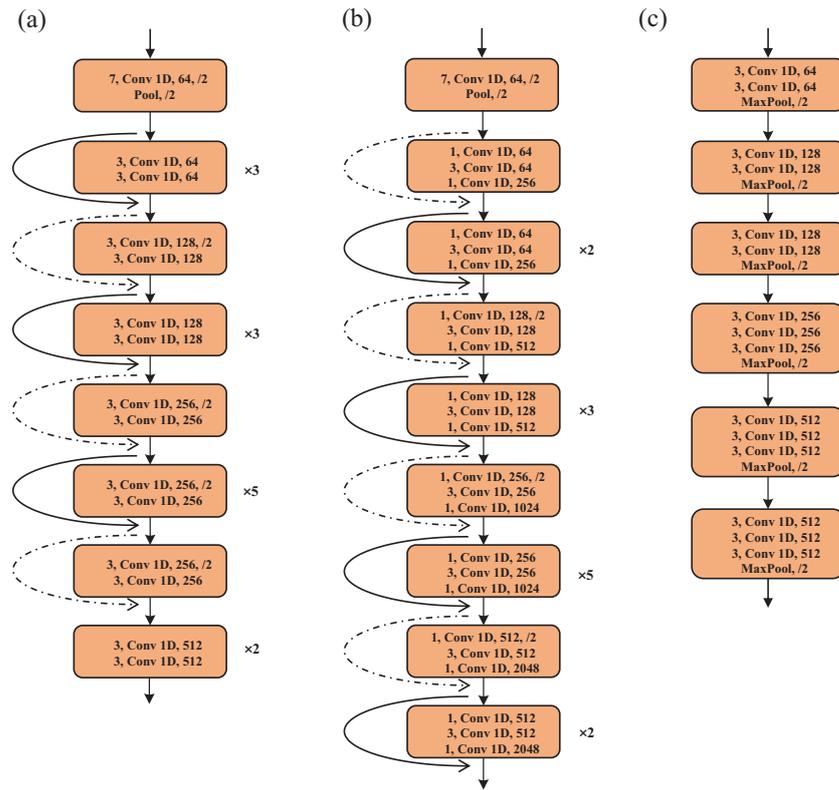
Figure 5. The three CNN models: (**a**) 1D ResNet34; (**b**) 1D ResNet50; (**c**) 1D VGG16.

Table 2. Parameter settings and network structures of the three deep learning models.

	Layer Name	Kernel	Parameter	Output Size	Connected from
	1D VGG16 [34]	×	3,713,648	44 × 512	INPUT
	LSTM	60	137,520	22 × 60	1D VGG16
VGG16 + LSTM + ATTENTION	Attention	×	3720	1 × 60	LSTM
	Dense	64	3904	1 × 64	Attention
	Dense	6	390	1 × 6	Dense
	Total Parameter: 3,859,182				
	Layer Name	Kernel	Parameter	Output Size	Connected from
	1D ResNet34 [35]	×	7,598,916	22 × 512	INPUT
	LSTM	60	137,520	22 × 60	1D ResNet34
ResNet34 + LSTM + ATTENTION	Attention	×	3720	1 × 60	LSTM
	Dense	64	3904	1 × 64	Attention
	Dense	6	390	1 × 6	Dense
	Total Parameter: 8,320,266				
	Layer Name	Kernel	Parameter	Output Size	Connected from
	1D ResNet50 [35]	×	21,945,280	22 × 2048	INPUT
	LSTM	60	137,520	22 × 60	1D ResNet34
ResNet50 + LSTM + ATTENTION	Attention	×	3720	1 × 60	LSTM
	Dense	64	3904	1 × 64	Attention
	Dense	6	390	1 × 6	Dense
	Total Parameter: 22,090,814				

In the LSTM layer, an intermediate output with a size of 1×60 was generated at each iteration. The activation function 'sigmoid' was used for the forget gate, the input gate,

and the output gate. The activation function 'tanh' was used for updating the state C_t. The initialization method used for the matrix weight was 'glorot uniform'.

In the attention layer, the sizes of the matrix weight W_s, bias b_s, and query vector u_s were 60×60, 60×1, and 60×1, respectively. The initialization method used was 'glorot uniform'. The first dense layer consisted of 64 neurons and used the activation function 'ReLU'. The second dense layer consisted of six neurons (corresponding to the different diseases) and used the activation function 'Sigmoid'. The initialization method used in the two fully connected layers was 'glorot uniform'.

4.2.2. The Modified Cross-Entropy Loss Function

In the database created in this work, atrial flutter and atrial tachycardia have a high negative correlation. It was found that the correlation (Poisson correlation degree) between atrial flutter and atrial tachycardia was -0.98 according to the correlation analysis of arrhythmia diseases based on the 200,000 ECG conclusions obtained from Shanghai Zhongshan Hospital. Thus, the loss function used in this work can be expressed using Equation (18).

$$loss = \frac{1}{n} \sum_{k=1}^{6} [y_k ln(a_k) + (1 - y_k) ln(1 - a_k)] + ln\left(1 - sin\left(\frac{\pi \cdot a_4 \cdot a_5}{2}\right)\right), \qquad (18)$$

where a_4 and a_5 represent the existence probabilities of atrial flutter and atrial tachycardia, respectively, obtained from the predicted ECG label vector.

The influence of the presence of both atrial flutter and atrial tachycardia in the predicted outcomes on the regularization term is shown in Figure 6a. The regularization term tended toward 0 when either only one or neither (i.e., atrial flutter and atrial tachycardia) appeared in the predicted outcomes. The regularization term increased rapidly when the probability of atrial tachycardia and atrial flutter simultaneously exceeded 0.5. Figure 6b shows the influence of the regularization term on the model's weight matrix concerning the partial derivative values of the loss function. When two labels with a negative correlation were present simultaneously, the corresponding $\partial Loss / \partial W$ value increased, thereby enhancing the speed of weight updates in the backward propagation process of the model. This enabled the model to promptly recognize negative correlations between the labels and adjust the weights accordingly. Conversely, when the model experienced a decrease in the speed of the weight updates, it tended to achieve stability.
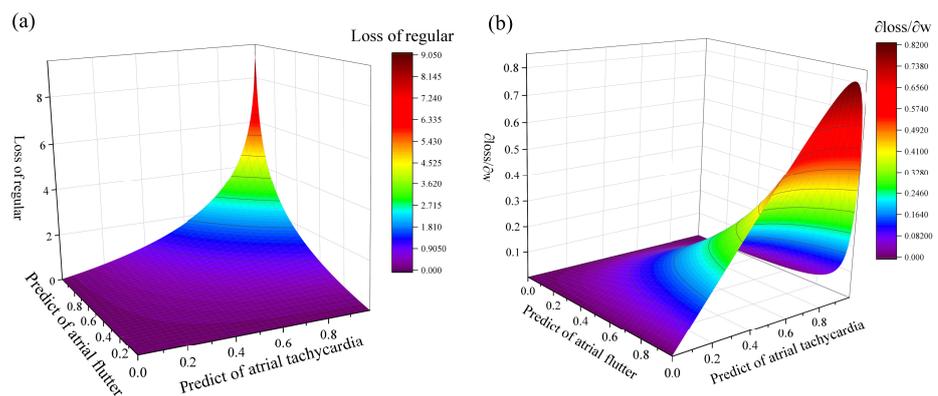

Figure 6. Changes in the regularization term: (**a**) The impact of negatively correlated labels on the regularization term. (**b**) The impact of the regularization term on the partial derivative values of the loss function with respect to the model's weight matrix.

4.3. Evaluation Indicators

In this section, six evaluation indicators are examined to assess the performance of the presented models. The six evaluation indicators are (1) Error Num, (2) Hamming Loss, (3)

Subset Accuracy, (4) Jaccard Index, (5) Precision, (6) Recall, and (7) F1 score, and they are expressed as follows.

$$Error\ Num = \sum_{x \in X} h_m(x) = 1\ and\ h_n(x) = 1, \tag{19}$$

$$Hamming\ Loss\ (h) = \frac{1}{|X|} \sum_{x \in X} \frac{1}{l} \sum_{j=1} [(L_j \in h_{(x)}) \otimes (L_j \in y)], \tag{20}$$

$$Subset\ Accuracy\ (h) = \frac{1}{|X|} \sum_{x \in x} [h(x) = y], \tag{21}$$

$$Jaccard\ Index\ (h) = \frac{1}{|X|} \sum_{x \in X} \frac{h(x)\ (y)}{h(x) \cup y}, \tag{22}$$

$$Precison = \frac{\sum_{j=1}^{j=l} TP_j}{\sum_{j=1}^{j=l} TP_j + FP_j}, \tag{23}$$

$$Recall = \frac{\sum_{j=1}^{j=l} TP_j}{\sum_{j=1}^{j=l} TP_j + FN_j}, \tag{24}$$

$$F1\ score = \frac{2 * precision * recall}{precsion + recall}, \tag{25}$$

where m and n refer to the positions (i.e., arrhythmia diseases) in the predicted label vector. The two arrhythmia diseases (m and n) cannot be present simultaneously in the ECG diagnostic results. $\otimes$ represents the logical symbol AND, y represents the label corresponding to instance x, represents the classification results of the multi-label model for x, L_j represents the j-th label in the label vector for instance x, and X represents the set of all instances x. TP_i (true positive) represents the number of positive samples correctly predicted by the multi-label model for the j-th label, FP_j (false positive) represents the number of positive samples incorrectly predicted by the multi-label model for the j-th label, and FN_j (false negative) represents the number of negative samples incorrectly predicted by the multi-label model for the j-th label.

'Error Num' is defined as the number of label pairs output by the model that cannot exist simultaneously, and it is used to study the effects of the modified loss function. The smaller the 'Error Num', the better the predictive performance of the model.

'Hamming Loss' is used to measure the fitting ability of the multi-label model. 'Subset Accuracy' is defined as the ratio between the number of correctly predicted samples and the total number of samples, and it is used to evaluate the predictive ability of the multi-label model. The 'Jaccard Index' is used to calculate the similarity between the label and the prediction score.

'Precision' is defined as the ratio between the number of correctly predicted positive samples and the total number of positive samples, and it is used to evaluate the accuracy of the multi-label model. 'Recall' is defined as the ratio between the number of correctly predicted positive samples and the number of samples predicted as positive, and it is used to measure the recall rate of the multi-label model. The 'F1 score' is defined as the weighted average of 'Recall' and 'Precision'. In general, a larger 'F1 score' indicates that the model has better predictive performance.

4.4. Performance After Adding LSTM + ATTENTION

To compare the effectiveness of LSTM + ATTENTION, this study selected the classic CNN models VGG16 [34], ResNet34 [35], and ResNet50 [35]. In terms of the Error Num metric shown in Table 3, all three models exhibited a decrease after applying the LSTM + ATTENTION structure. The F1 scores of VGG16, ResNet34, and ResNet50 reached up to

94.74%, 93.64%, and 93.52%, respectively. It is proved that a CNN with a suitable structure is effective in multi-label ECG classification. After adding LSTM + ATTENTION, the F1 scores of the three methods were 95.21%, 93.98%, and 94.16%, respectively. This shows that the prediction performance of a CNN can be improved or ensured by adding LSTM + ATTENTION.

Table 3. Comparison of the performance of the three CNN models after adding LSTM + ATTENTION.

Model	VGG16 [34]		ResNet34 [35]		ResNet50 [35]	
Add LSTM + ATTENTION	No	Yes	No	Yes	No	Yes
Error Num (Num)	18	12	41	30	5	4
Hamming Loss	0.0200	0.0187	0.0250	0.0240	0.0252	0.0230
Subset Accuracy	0.9265	0.9392	0.9188	0.9245	0.9213	0.9269
Jaccard Index	0.9441	0.9521	0.9343	0.9397	0.9383	0.9417
Precision	0.9564	0.9561	0.9424	0.9435	0.9452	0.9495
Recall	0.9409	0.9491	0.9327	0.9380	0.9338	0.9363
F1 score	0.9474	0.9521	0.9364	0.9398	0.9352	0.9416

4.5. Performance of the Traditional Cross-Entropy Loss Function

In this section, the traditional cross-entropy loss function (see Equation (16)) is used to train the presented deep learning methods, as shown in Table 2.

VGG16 [34], ResNet34 [35], and ResNet50 [35] are the most commonly used CNN models for ECG classification. In this work, VGG16, ResNet34, ResNet50, and their combinations with LSTM + ATTENTION were used to verify the performance of the improved loss function. 'Adam' was chosen as the optimizer, the initial learning rate was set to 0.001, and the number of training epochs was set to 100. Regarding the hyperparameter settings for each CNN model, we established them based on parameters published in the literature [36,37] and determined the optimal model training configuration using the GridsearchCV algorithm[38]. To evaluate the performance of the multi-label model on the validation dataset, an early stop mechanism was introduced into the training process to prevent overfitting. The training of the model was stopped if the loss of the multi-label model on the validation dataset did not decrease in 10 consecutive training sessions.

The training dataset (23,322) and the validation dataset (2591), as described in Section 3, were both used to train the three deep learning models. The test dataset (13,156) was used to test the effectiveness of the trained models.

The training process using the traditional cross-entropy loss function is shown in Figure 7a. The corresponding experimental results are shown in Table 4.
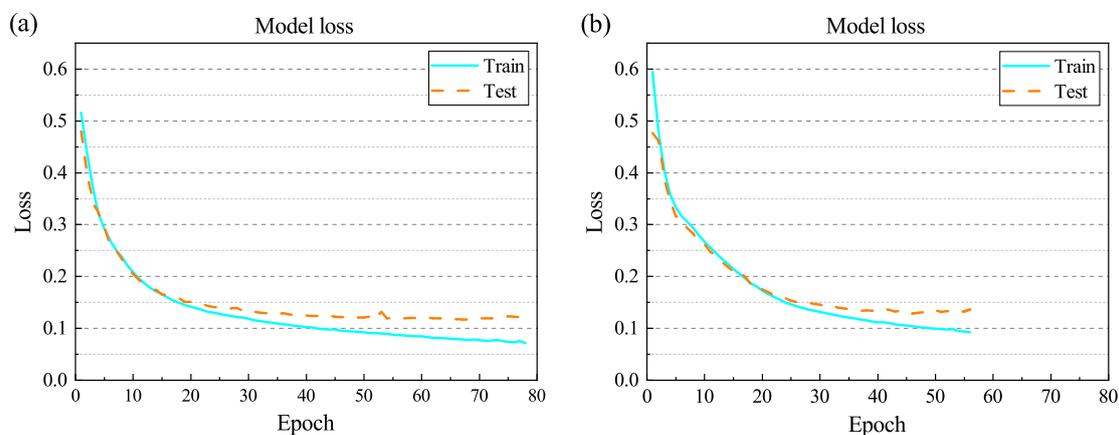

Figure 7. Comparison of the traditional and modified loss functions: (**a**) training process using the traditional loss function; (**b**) training process using the modified loss function.

Table 4. Comparison between the modified and traditional loss functions.

Model	VGG16 + LSTM + ATTENTION		ResNet34 + LSTM + ATTENTION		ResNet50 + LSTM + ATTENTION	
Loss function	Traditional	Modified	Traditional	Modified	Traditional	Modified
Error Num (Num)	12	0	30	0	4	0
Hamming Loss	0.0187	0.0197	0.0240	0.0234	0.0230	0.0208
Subset Accuracy	0.9392	0.9287	0.9245	0.9179	0.9269	0.9242
Jaccard Index	0.9521	0.9419	0.9397	0.9313	0.9417	0.9392
Precision	0.9561	0.9595	0.9435	0.9499	0.9495	0.9549
Recall	0.9491	0.9360	0.9380	0.9283	0.9363	0.9340
F1 score	0.9521	0.9474	0.9398	0.9382	0.9416	0.9440

4.6. Effectiveness of the Modified Cross-Entropy Loss Function

In this section, the modified cross-entropy loss function (see Equation (18)) is used for training the presented deep learning methods, as shown in Table 2. The other parameter settings are the same as those used in Section 4.2.

The training process using the modified cross-entropy loss function is shown in Figure 7b. The corresponding experimental results are given in Table 4.

The early stop mechanism stopped the training of the model when it entered a stable phase. In Figure 7, it can be seen that (1) the model loss did not decrease after 68 training epochs when using the traditional loss function, and (2) the model loss did not decrease after 52 training epochs when using the modified loss function. It can be concluded that the training epochs were shorter when using the modified loss function.

In addition, it can be seen in Table 4 that (1) for both loss functions, VGG16 outperformed the ResNet34 and ResNet50 models across all evaluation metrics; (2) 'Error Num' was significantly reduced when using the modified loss function; (3) 'Precision' slightly increased when using the modified loss function; and (4) 'Subset Accuracy', 'Jaccard Index', 'Recall', and 'F1 score' decreased slightly when using the modified loss function. It can be concluded that the modified loss function can significantly reduce the number of coexisting strongly negatively correlated labels while guaranteeing model performance. Therefore, it can be concluded that the modified loss function can effectively prevent the occurrence of strongly negatively correlated arrhythmias in the multi-label diagnosis of arrhythmias.

Table 5 compares the accuracy of classifying different arrhythmias using the two different loss functions. In the table, it can be seen that there was a slight improvement in accuracy when diagnosing normal ECG, sinus tachycardia, atrial flutter, and atrial tachycardia with the improved loss function. However, it should be noted that the model's accuracy in classifying PVCs decreased by more than 1%. Furthermore, with regard to the overall improvement in precision evident in Table 4 and Figure 8, we conclude that using the modified two-category cross-entropy loss function significantly reduces the number of misdiagnoses of atrial tachycardia.

Table 5. Comparison of the accuracy of 6 types of cardiac arrhythmias between the modified and the traditional loss functions.

| Model	VGG16 + LSTM + ATTENTION		ResNet34 + LSTM + ATTENTION		ResNet50 + LSTM + ATTENTION	
Loss function	Traditional	Modified	Traditional	Modified	Traditional	Modified
Normal	0.9670	0.9707	0.9652	0.9676	0.9615	0.9678
Sinus tachycardia	0.9823	0.9826	0.9854	0.9879	0.9822	0.9825
Sinus bradycardia	0.9856	0.9853	0.9870	0.9649	0.9861	0.9752
Atrial flutter	0.9744	0.9786	0.9569	0.9769	0.9742	0.9770
Atrial tachycardia	0.9776	0.9870	0.9749	0.9806	0.9764	0.9869
PVC	0.9806	0.9785	0.9830	0.9405	0.9786	0.9481
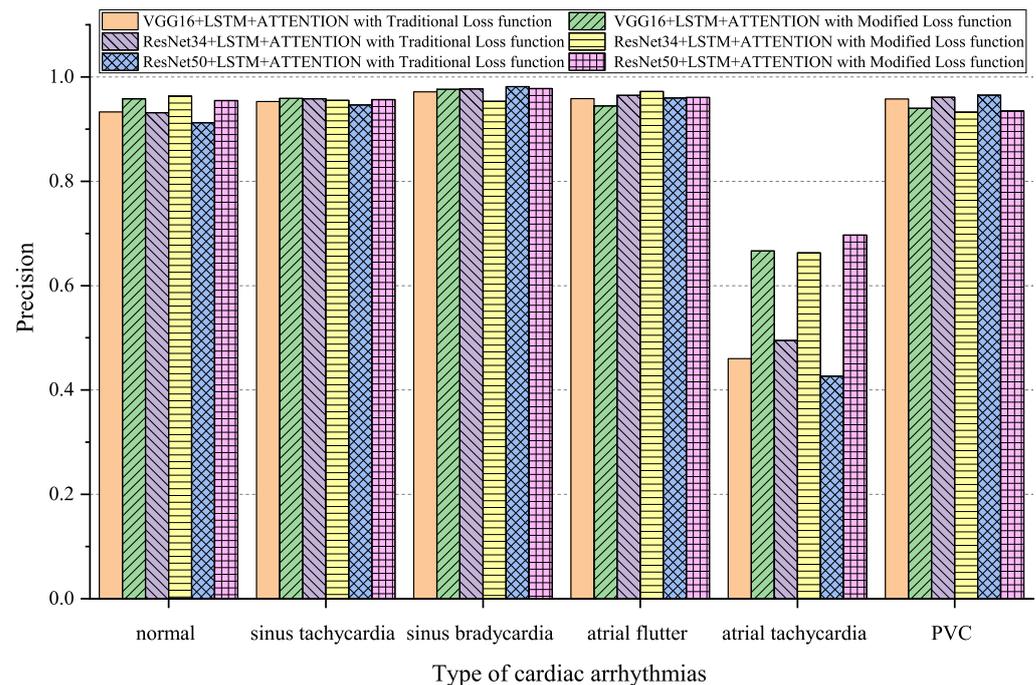
Figure 8. Comparison of the precision of 6 types of cardiac arrhythmias between the modified and traditional loss functions.

5. Discussion

This article proposes a multi-label diagnosis method for cardiac arrhythmias based on a modified two-category cross-entropy loss function. In order to validate the performance of LSTM + ATTENTION, the classic neural networks VGG16, ResNet34, and ResNet50 are used for evaluation. The results show that the prediction performance of the CNN can be improved or ensured by adding LSTM + ATTENTION.

Many types of diseases can be identified from ECG signals, and some of these diseases cannot exist simultaneously. We compare the traditional loss function to our improved loss function across different CNN models. The results indicate that using the traditional loss function still produces non-coexisting labels. However, when using the proposed modified loss function in this paper with the addition of a regularization term, the model's weight update rate between negatively correlated labels is strengthened, forcing the CNN model to learn the connections between non-coexisting labels and preventing the appearance of non-coexisting label pairs in the diagnostic results. In addition, the improved loss function shortens the required training period of the model, demonstrating the effectiveness of our approach in reducing model training costs and enhancing the feasibility of clinical applications.

To validate the classification performance of the modified loss function in diagnosing cardiac arrhythmias using neural network models, we compare the accuracy of the two different loss functions in six types of ECG arrhythmias. The results indicate that our method can improve the precision of the model for negatively correlated atrial tachycardia and atrial flutter labels. This means that it can reduce the risk of false positives in medical diagnosis, demonstrating the potential value of the improved loss function in clinical applications. However, our method shows decreased accuracy in the identification of PVCs and sinus bradycardia. Currently, our research focuses on six common types of cardiac arrhythmias. In the future, we will expand our scope to include a broader range of cardiac arrhythmia datasets.

6. Conclusions

This work applies a CNN + LSTM + ATTENTION model to multi-label ECG classification. To prevent the occurrence of label pairs that cannot exist simultaneously, in

the presented method, a modified cross-entropy loss function is proposed. The modified loss function introduces a regularization term to increase the penalty for the coexistence of arrhythmias exhibiting a strong negative correlation. Experimental results show that the modified loss function helps prevent the occurrence of strongly negatively correlated arrhythmias, sacrificing prediction accuracy by only a small margin. This work provides theoretical evidence for multi-label ECG classification in clinical diagnosis.

References

1. Geddes, J.S.; Warner, H.R. A PVC detection program. *Comput. Biomed. Res.* **1971**, *4*, 493–508. [CrossRef] [PubMed]
2. Kezdi, P.; Naylor, W.S.; Rambousek, R.; Stanley, E. A practical heart rate and ectopic beat detector. *J. Electrocardiol.* **1968**, *1*, 213–219. [CrossRef]
3. Arafat, M.A.; Chowdhury, A.W.; Hasan, M.K. A simple time domain algorithm for the detection of ventricular fibrillation in electrocardiogram. *Signal Image Video Process.* **2011**, *5*, 1–10. [CrossRef]
4. Thakor, N.V.; Zhu, Y.S.; Pan, K.Y. Ventricular tachycardia and fibrillation detection by a sequential hypothesis testing algorithm. *IEEE Trans. Biomed. Eng.* **1990**, *37*, 837–843. [CrossRef] [PubMed]
5. Anas, E.M.A.; Lee, S.Y.; Hasan, M.K. Sequential algorithm for life threatening cardiac pathologies detection based on mean signal strength and EMD functions. *Biomed. Eng. Online* **2010**, *9*, 1–22. [CrossRef] [PubMed]
6. Jekova, I.; Krasteva, V. Real time detection of ventricular fibrillation and tachycardia. *Physiol. Meas.* **2004**, *25*, 1167. [CrossRef] [PubMed]
7. Barro, S.; Ruiz, R.; Cabello, D.; Mira, J. Algorithmic sequential decision-making in the frequency domain for life threatening ventricular arrhythmias and imitative artefacts: A diagnostic system. *J. Biomed. Eng.* **1989**, *11*, 320–328. [CrossRef]
8. Kuo, S. Computer detection of ventricular fibrillation. In *Computers in Cardiology*; IEEE Comupter Society: Washington, DC, USA, 1978; pp. 347–349.
9. Zhang, X.S.; Zhu, Y.S.; Thakor, N.V.; Wang, Z.Z. Detecting ventricular tachycardia and fibrillation by complexity measure. *IEEE Trans. Biomed. Eng.* **1999**, *46*, 548–555. [CrossRef]
10. Amann, A.; Tratnig, R.; Unterkofler, K. Detecting ventricular fibrillation by time-delay methods. *IEEE Trans. Biomed. Eng.* **2006**, *54*, 174–177. [CrossRef]
11. Amann, A.; Tratnig, R.; Unterkofler, K. A new ventricular fibrillation detection algorithm for automated external defibrillators. In *Computers in Cardiology*; IEEE: New York, NY, USA, 2005; pp. 559–562.
12. Li, H.; Han, W.; Hu, C.; Meng, M.Q.H. Detecting ventricular fibrillation by fast algorithm of dynamic sample entropy. In Proceedings of the 2009 IEEE International Conference on Robotics and Biomimetics (ROBIO), Guilin, China, 19–23 December 2009; pp. 1105–1110.
13. Martis, R.J.; Acharya, U.R.; Mandana, K.M.; Ray, A.K.; Chakraborty, C. Cardiac decision making using higher order spectra. *Biomed. Signal Process. Control* **2013**, *8*, 193–203. [CrossRef]
14. Roza, V.C.C.; de Almeida, A.M.; Postolache, O.A. Design of an artificial neural network and feature extraction to identify arrhythmias from ECG. In Proceedings of the 2017 IEEE International Symposium on Medical Measurements and Applications (MeMeA), Rochester, MN, USA, 7–10 May 2017; pp. 391–396.
15. Yang, W.; Si, Y.; Wang, D.; Guo, B. Automatic recognition of arrhythmia based on principal component analysis network and linear support vector machine. *Comput. Biol. Med.* **2018**, *101*, 22–32. [CrossRef]
16. Houssein, E.H.; Ibrahim, I.E.; Neggaz, N.; Hassaballah, M.; Wazery, Y.M. An efficient ECG arrhythmia classification method based on Manta ray foraging optimization. *Expert Syst. Appl.* **2021**, *181*, 115131. [CrossRef]
17. Singh, A.K.; Krishnan, S. *ECG Signal Feature Extraction Trends in Methods and Applications*; BioMed Central: London, UK, 2023; Volume 22, pp. 1–36.
18. Feng, N.; Xu, S.; Liang, Y.; Liu, K. A probabilistic process neural network and its application in ECG classification. *IEEE Access* **2019**, *7*, 50431–50439. [CrossRef]
19. Yıldırım, Ö.; Pławiak, P.; Tan, R.S.; Acharya, U.R. Arrhythmia detection using deep convolutional neural network with long duration ECG signals. *Comput. Biol. Med.* **2018**, *102*, 411–420. [CrossRef] [PubMed]
20. Luo, X.; Yang, L.; Cai, H.; Tang, R.; Chen, Y.; Li, W. Multi-classification of arrhythmias using a HCRNet on imbalanced ECG datasets. *Comput. Methods Programs Biomed.* **2021**, *208*, 106258. [CrossRef] [PubMed]

21. Yao, Q.; Wang, R.; Fan, X.; Liu, J.; Li, Y. Multi-class arrhythmia detection from 12-lead varied-length ECG using attention-based time-incremental convolutional neural network. *Inf. Fusion* **2020**, *53*, 174–182. [CrossRef]
22. Li, Y.; Zhang, Z.; Zhou, F.; Xing, Y.; Li, J.; Liu, C. Multi-Label Classification of Arrhythmia for Long-Term Electrocardiogram Signals with Feature Learning. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–11. [CrossRef]
23. Ran, S.; Li, X.; Zhao, B.; Jiang, Y.; Yang, X.; Cheng, C. Label correlation embedding guided network for multi-label ECG arrhythmia diagnosis. *Knowl.-Based Syst.* **2023**, *270*, 110545. [CrossRef]
24. Yang, J.; Li, J.; Lan, K.; Wei, A.; Wang, H.; Huang, S.; Fong, S. Multi-Label Attribute Selection of Arrhythmia for Electrocardiogram Signals with Fusion Learning. *Bioengineering* **2022**, *9*, 268. [CrossRef]
25. Yoo, J.; Jun, T.J.; Kim, Y.H. xECGNet: Fine-tuning attention map within convolutional neural network to improve detection and explainability of concurrent cardiac arrhythmias. *Comput. Methods Programs Biomed.* **2021**, *208*, 106281. [CrossRef]
26. Yang, X.; Zhang, X.; Yang, M.; Zhang, L. 12-Lead ECG arrhythmia classification using cascaded convolutional neural network and expert feature. *J. Electrocardiol.* **2021**, *67*, 56–62. [CrossRef] [PubMed]
27. Ge, Z.; Jiang, X.; Tong, Z.; Feng, P.; Zhou, B.; Xu, M.; Wang, Z.; Pang, Y. Multi-label correlation guided feature fusion network for abnormal ECG diagnosis. *Knowl.-Based Syst.* **2021**, *233*, 107508. [CrossRef]
28. LeCun, Y.; Boser, B.; Denker, J.S.; Henderson, D.; Howard, R.E.; Hubbard, W.; Jackel, L.D. Backpropagation applied to handwritten zip code recognition. *Neural Comput.* **1989**, *1*, 541–551. [CrossRef]
29. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]
30. Yang, Z.; Yang, D.; Dyer, C.; He, X.; Smola, A.; Hovy, E. Hierarchical attention networks for document classification. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2016; pp. 1480–1489.
31. He, T.; Zhang, L.; Guo, J.; Yi, Z. Multilabel classification by exploiting data-driven pair-wise label dependence. *Int. J. Intell. Syst.* **2020**, *35*, 1375–1396. [CrossRef]
32. Gao, W.; Zhou, Z.H. On the consistency of multi-label learning. *Artif. Intell.* **2013**, *199–200*, 22–44. [CrossRef]
33. Kobayashi, T. Two-Way Multi-Label Loss. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 18–22 June 2023; pp.7476–7485.
34. Andreotti, F.; Carr, O.; Pimentel, M.A.F.; Mahdi, A.; De Vos, M. Comparing feature-based classifiers and convolutional neural networks to detect arrhythmia from short segments of ECG. In Proceedings of the IEEE 2017 Computing in Cardiology (CinC), Rennes, France, 24–27 September 2017; pp. 1–4.
35. Jiang, Z.; Lai, Y.; Zhang, J.; Zhao, H.; Mao, Z. Multi-factor operating condition recognition using 1D convolutional long short-term network. *Sensors* **2019**, *19*, 5488. [CrossRef]
36. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
37. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. In Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015—Conference Track Proceedings, San Diego, CA, USA, 7–9 May 2015; pp. 1–14.
38. Pirjatullah; Kartini, D.; Nugrahadi, D.T.; Muliadi; Farmadi, A. Hyperparameter Tuning using GridsearchCV on the Comparison of the Activation Function of the ELM Method to the Classification of Pneumonia in Toddlers. In Proceedings of the 2021 4th International Conference on Computer and Informatics Engineering: IT-Based Digital Industrial Innovation for the Welfare of Society, IC2IE 2021, Depok, Indonesia, 14–15 September 2021; pp. 390–395.