

Article

An ECG Classification Method Based on Multi-Task Learning and CoT Attention Mechanism

Quancheng Geng¹, Hui Liu¹, Tianlei Gao¹, Rensong Liu¹, Chao Chen¹, Qing Zhu^{2,*} and Minglei Shu^{1,*} ¹ Shandong Artificial Intelligence Institute, Qilu University of Technology (Shandong Academy of Sciences), Jinan 250014, China² Department of Cardiology, Qilu Hospital of Shandong University, Jinan 250012, China

* Correspondence: 198862000790@email.sdu.edu.cn (Q.Z.); shuml@sdas.org (M.S.)

Abstract: Electrocardiogram (ECG) is an efficient and simple method for the diagnosis of cardiovascular diseases and has been widely used in clinical practice. Because of the shortage of professional cardiologists and the popularity of electrocardiograms, accurate and efficient arrhythmia detection has become a hot research topic. In this paper, we propose a new multi-task deep neural network, which includes a shared low-level feature extraction module (i.e., SE-ResNet) and a task-specific classification module. Contextual Transformer (CoT) block is introduced in the classification module to dynamically model the local and global information of ECG feature sequence. The proposed method was evaluated on public CPSC2018 and PTB-XL datasets and achieved an average F1 score of 0.827 on the CPSC2018 dataset and an average F1 score of 0.833 on the PTB-XL dataset.

Keywords: ECG; SE-ResNet; multi-task deep neural network; Contextual Transformer



Citation: Geng, Q.; Liu, H.; Gao, T.; Liu, R.; Chen, C.; Zhu, Q.; Shu, M. An ECG Classification Method Based on Multi-Task Learning and CoT Attention Mechanism. *Healthcare* **2023**, *11*, 1000. <https://doi.org/10.3390/healthcare11071000>

Academic Editor: Joaquim Carreras

Received: 21 February 2023

Revised: 25 March 2023

Accepted: 26 March 2023

Published: 31 March 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Cardiovascular disease is responsible for the deaths of approximately 17.9 million people per year, accounting for 31% of all global fatalities [1]. Cardiovascular disease has become the disease with the highest mortality rate and has seriously threatened human life and health. Electrocardiogram (ECG) is the most widely used noninvasive heart disease diagnosis technology. The ECG signal represents electrical changes during the cardiac cycle and can be recorded with the surface electrodes. Analysis of ECG [2–4] signals allows doctors to gather important insights into the health condition of patients and to promptly identify heart abnormalities, thereby prolonging life and improving quality of life through appropriate treatment.

With the rapid advancement of mobile devices for ECG monitoring, automatic interpretation of ECG recordings is becoming more and more important for the early detection and diagnosis of arrhythmias [5]. In particular, many effective arrhythmia classification methods for single-lead ECG have been proposed [6–8], but single-lead ECG alone is not enough to accurately diagnose various heart diseases. 12-lead ECG can comprehensively evaluate cardiac electrical activity (including arrhythmia and myocardial infarction) and each lead reflects heart states through electrical signal changes from different perspectives. Therefore, as a standard clinical ECG examination, 12-lead ECG has attracted more and more interest from researchers [9–13].

Over the past few years, various machine learning methods have been employed to analyze ECG signals, such as decision tree [14], support vector machine [15] and hidden Markov model [16]. The most critical part of these methods is to extract discriminant information from original ECG data, i.e., feature extraction. In order to extract features, many methods have been proposed, which can be divided into two categories: manual approaches [17,18] and automatic approaches [19–21]. The manual approaches mainly rely on professional medical knowledge and the rich expertise of cardiologists. For example, Martis et al. extracted features such as RR interval, R peak amplitude and QRS duration

from the original ECG data and used decision trees to diagnose arrhythmia types [21]. In addition, wavelet transform [22], short-time Fourier transform and principal component analysis [23] are often used to extract time-frequency features. It is difficult to obtain satisfying performance for manual approaches because ECG features such as PR interval show a large diversity for different individuals.

Deep learning [24–27] is an efficient feature learning method that employs neural networks for mining useful features from a large amount of data. Due to its ability to effectively express unstructured data with non-linear representation, deep learning has been successfully applied in a variety of fields including computer vision, speech recognition and natural language processing [28–30]. Due to its powerful ability, many researchers have used deep learning-based methods for ECG classification [31–34], which show higher accuracy than traditional methods. For example, Rahhal et al. [29] proposed to use continuous wavelet transform to convert one-dimensional ECG signals into two-dimensional images and then used convolutional neural network (CNN) to extract useful features of images, yielding a high accuracy for identifying abnormal ventricular beats. Hannun et al. [34] built a 34-layer CNN network and obtained a cardiologist-level F1 score of 0.837 for classification of twelve arrhythmias. Other common deep learning models that have been used for ECG classification include Long-Short Term Memory (LSTM) network [35], Recurrent Neural Network (RNN) [36] and bidirectional LSTM network [37].

Although deep learning methods have achieved promising performance in ECG classification, they still have some problems. The feature map derived from different leads of the 12-lead ECG signal may exhibit varying levels of contributions for detecting arrhythmias. Specifically, the spatial characteristics of arrhythmias, such as the morphology of waveform, may differ between leads. For example, atrial fibrillation is most obvious in leads II and V1 [38]. In addition, previous studies mainly focus on one ECG classification task, though there are many kinds of ECG tasks (e.g., anomaly detection, heartbeat classification, ECG delineation and ECG diagnosis) in the real world and there exist correlations between tasks, which indicates that the combination of different tasks is conducive to learning better ECG representations [39].

Multi-task learning has achieved great success and has performed well in many fields, such as natural language processing [40], speech recognition [41], computer vision [42] and face verification [43]. In this paper, we propose a deep multi-task learning [44] method for ECG classification. Specifically, we first construct a related auxiliary classification task with a corresponding dataset by merging similar classes (e.g., left bundle branch block and right bundle branch block) or splitting large classes. We propose a multi-task deep neural network, which includes a shared low-level feature extraction module (i.e., SE-ResNet) and a task-specific classification module. The main contributions of this study are as follows:

1. A new multi-task network consisting of a shared low-level feature extraction module and a task-specific classification module is proposed for ECG classification.
2. We propose two strategies to create auxiliary tasks, which exploit the hierarchical class information to achieve feature sharing between the main task and the auxiliary task.
3. Contextual Transformer (CoT) block is first introduced in ECG classification to solve the problem that the self-attention mechanism only focuses on the local information of the ECG sequence.

2. Materials and Methods

2.1. Data

2.1.1. CPSC2018 Dataset

The 2018 China Physiological Signal Challenge (CPSC2018) dataset [45] includes 6877 publicly accessible 12-lead ECG records (female: 3178; male: 3699) and a private test set consisting of 2954 12-lead ECG records. ECG records were collected from 11 hospitals. Record lengths range from 6 s to 60 s with a sampling rate of 500 Hz. Table 1 shows the number of records for each class and superclass. In addition, we also show the specific description of each class in Table 1.

Table 1. ECG classes and their numbers in the CPSC2018 dataset.

Class	Description	Superclass	#Records
SNR	Normal ECG	NORM	918
AF	Atrial fibrillation	AF	1098
IAVB	First-degree atrioventricular block	QRS	704
LBBB	Left bundle branch block		207
RBBB	Right bundle branch block		1695
PAC	Premature atrial contraction	V	556
PVC	Premature ventricular contraction		672
STE	ST-segment elevated	ST	825
STD	ST-segment depression		202
Total			6877

2.1.2. PTB-XL Dataset

The PTB-XL dataset [46] includes 21,837 10-second 12-lead clinical ECG records from 18,885 patients, 52% of whom were male and 48% female. The sampling frequency is 500 Hz for all records. There are three levels of ECG labels in the dataset: 5 superclasses, 23 subclasses and 44 diagnostic statements. Table 2 shows the number of records for each superclass and subclass. In this study, we used the recommended data partition scheme of training and test sets [46]. The whole data set was split into 10 folds and the ninth and tenth folds contain only ECGs that were validated by at least one cardiologist, which were recommended to be used as validation and test sets, respectively. The remaining eight folds were used for training.

Table 2. ECG classes and their numbers in the PTB-XL dataset.

Superclass	Subclass	Description	#Records
NORM	NORM	Noraml ECG	9528
CD	LAFB/LPFB	left anterior/posterior fascicular block	1803
	IRBBB	incomplete right bundle branch block	1118
	ILBBB	incomplete left bundle branch block	77
	CLBBB	complete left bundle branch block	536
	CRBBB	complete right bundle branch block	542
	_AVB	AV block	827
	IVCD	non-specific intraventricular conduction disturbance	789
	WPW	Wolff–Parkinson–White syndrome	80
HYP	LVH	left ventricular hypertrophy	2137
	RVH	right ventricular hypertrophy	126
	LAO/LAE	left atrial overload/enlargement	427
	RAO/RAE	right atrial overload/enlargement	99
	SEHYP	septal hypertrophy	30
MI	AMI	anterior myocardial infarction	3172
	IMI	inferior myocardial infarction	3263
	LMI	lateral myocardial infarction	201
	PMI	posterior myocardial infarction	17
STTC	ISCA	ischemic in anterior leads	1016
	ISCI	ischemic in inferior leads	398
	ISC_	non-specific ischemi	1275
	STTC	ST-T changes	2329
	NST_	non-specific ST changes	770
Total			21,837 [†]

[†] The total number of class labels is 30,560.

2.2. Preprocessing

In this study, data preprocessing is performed according to previous work [47]. Specifically, ECG signals are downsampled from 500 Hz to 250 Hz to accelerate the training. In addition, signals in the CPSC2018 dataset do not have the same length; therefore, they are

cropped or padded to 60 s with zero since the convolutional neural network cannot accept input of different lengths in a batch.

2.3. Creating Auxiliary Task

Previous studies mostly focus on a single task and ignore other potentially helpful information. This study proposes a multi-task model to simultaneously learn two tasks related to ECG classification, in which the feature representation learned from one task can also be utilized for the other task. Compared with single-task model, the multi-task model has multiple independent outputs on their own paths, which helps to prevent overfitting and enhances the model's generalizability.

In this study, we construct auxiliary tasks based on the relationships between ECG classes. For the CPSC2018 dataset, the classification of the original nine classes is the main task. Then we reorganize the original nine classes into five super classes (see Table 1), which are used for the auxiliary task. For the PTB-XL dataset, the main task is the classification of five classes. Then we reuse the 23 subclass-level labels (see Table 2) as the target of auxiliary task.

2.4. Multi-Task Neural Network

We design a multi-task deep neural network using hard parameter sharing. As shown in Figure 1, it mainly includes a shared neural network (SE-ResNet) and feature extraction networks for specific tasks. By using a shared network structure prior to the multi-task module, the model is able to learn both tasks concurrently using a shared representation.

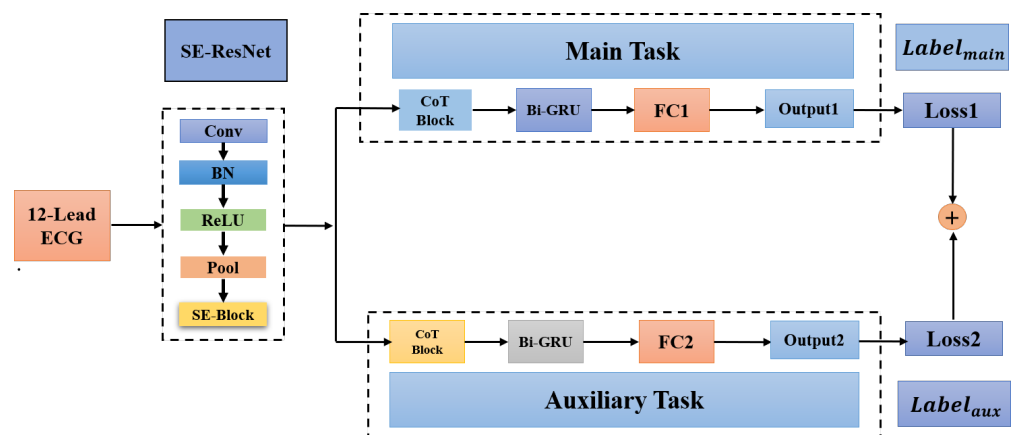


Figure 1. Multi-task neural network. FC1 and FC2 are fully connected layers used in the main task and the auxiliary task, respectively.

2.4.1. Squeeze-Excitation-ResNet (SE-ResNet)

We propose the SE-ResNet as the shared network by both tasks. Inspired by ResNet [48] design, SE-ResNet is composed of multiple SE blocks. As shown in Figure 2a, each SE block consists of two convolutional layers with a kernel size of 7. A batch normalization (BN) layer is applied after each convolutional layer and we use the rectified linear unit (ReLU) activation function [49] as the nonlinear transformation function. BN helps to speed up the convergence of the model during training by normalizing the data in each batch. A dropout rate of 0.2 is set to prevent overfitting of the neural network.

The last BN layer is followed by an SE module (see Figure 2b). The SE module can obtain the importance of feature information to strengthen the guiding role of channel information in the ECG classification process. The SE module includes two parts: squeeze and excitation. The squeeze operation seeks to use a value with a global receptive field to reflect the significance of each channel feature and to generate a feature map through global mean pooling of each channel. The excitation operation uses the fully connected

layer to act on the feature map and the importance weight of each channel is applied to the corresponding channel to construct the correlation between the channels.

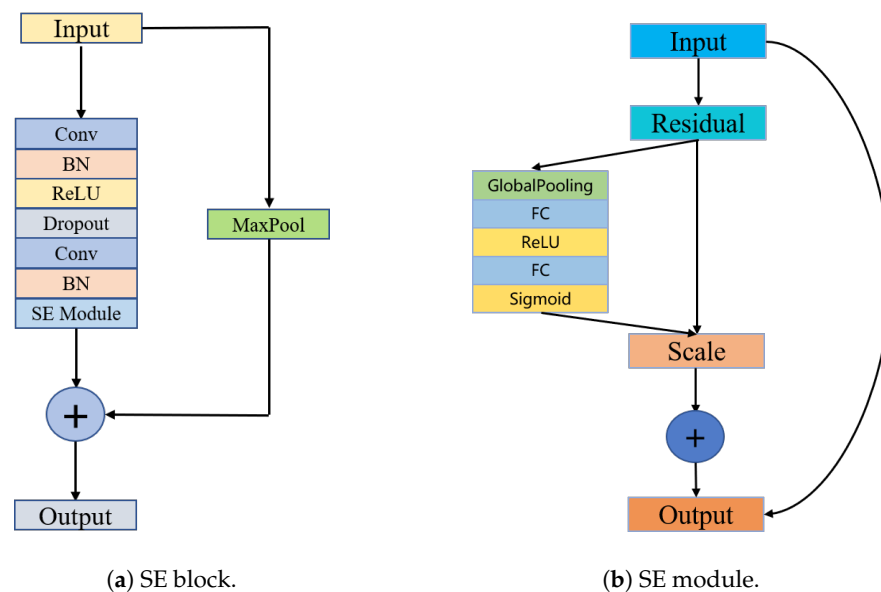


Figure 2. SE-ResNet.

2.4.2. Contextual Transformer (CoT) Block

The traditional self-attention [50] mechanism characterizes interactions between different positions of feature sequence well, only depending on the input itself. However, in the traditional self-attention mechanism, all pairs of query keyword relationships are learned independently on isolated query keyword pairs without exploring the rich context between them, which severely limits the ability for self-attention learning on ECG sequences. To alleviate this problem, we employ a new Contextual Transformer (CoT) block [51] in this study, which integrates contextual information mining and self-attention learning into a unified architecture. The CoT attention mechanism promotes self-attention learning by using additional contextual information between input keys and ultimately improves the representation characteristics of deep networks. The structure of the CoT block is shown in Figure 3.

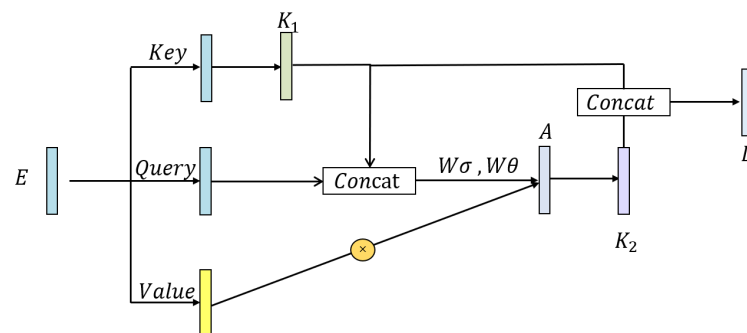


Figure 3. Structure of CoT block.

We define $Q = E$, $K = E$, $V = EW$ for the ECG signal sequence E extracted by SE-ResNet, where Q and K are the values of the original ECG signal sequence, V is the value of the feature mapping of the ECG signal and W represents an embedding matrix. Firstly, we use 3×3 convolution to statically model K to obtain with local sequence information representation. Secondly, in order to interact query information with local

sequence information K_1 , we need to concatenate K_1 and Q and then generate an attention map after continuous convolution operation. The formula is as follows:

$$A = [K_1, Q]W_\sigma W_\theta \quad (1)$$

We then need to multiply A and V to get the sequence K_2 with global sequence information by:

$$K_2 = V * A \quad (2)$$

Finally, we concatenate the local sequence information K_1 and the global sequence information K_2 to obtain the final output result D .

$$D = [K_1, K_2] \quad (3)$$

2.4.3. Bi-GRU Module

The GRU [52] layer can model the relationship of long-term context, but one drawback is that it can only read the input ECG signal sequence information from a single direction. In this study, we use the Bi-GRU layer to capture the characteristics of the input ECG feature map from the positive and negative directions, respectively.

We will get the matrix D through the CoT attention layer as the input of Bi-GRU. In the forward layer of the Bi-GRU layer, the input vectors (from x_t to x_{t+1}) are read in a positive order and the forward hidden layer state corresponding to each vector is calculated, namely $(\vec{h}_1, \dots, \vec{h}_t, \dots, \vec{h}_T)$. Similarly, in the backward layer of the Bi-GRU layer, the input vectors (from x_t to x_{t-1}) are read in reverse order and the reverse hidden layer state corresponding to each vector is calculated. The structure of the Bi-GRU layer is shown in Figure 4.

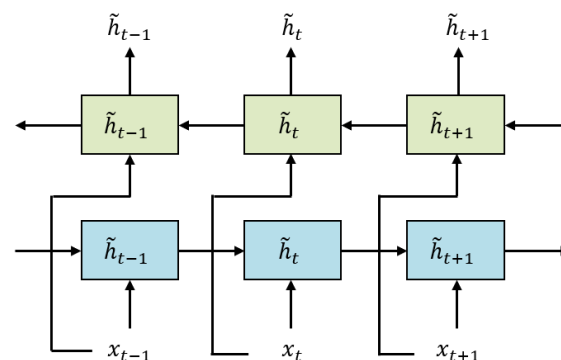


Figure 4. Structure of Bi-GRU layer.

2.5. Loss Function

Feature vectors output by the Bi-GRU layer are sent to the last fully connected layer, which generates the class distribution of the two tasks. For two tasks, we calculate the loss using the cross entropy function. Then two loss values are added with different weights. The final loss is as follows:

$$L^{(total)} = \lambda L^{(main)} + (1 - \lambda) L^{(aux)} \quad (4)$$

where λ is the weighting parameter, L^{main} , L^{aux} represent the loss of the main task and auxiliary task, respectively.

3. Experimental Setup

3.1. Parameter Setting

For both datasets, we use the Adam optimizer and set the initial learning rate to 0.0005. The learning rate is reduced by a factor of 10 every 10 epochs. We set the batch size to 32. Ten-fold cross-validation is used in this work, where the validation set is used

for model and parameter selection and the test set is used to test the effectiveness of the network. In order to alleviate the over-fitting during the training process, we also adopted an early stop mechanism, i.e., we will end the training when the loss is not decreasing for 10 consecutive epochs.

3.2. Evaluation Metrics

In this study, the F1 score is used to measure the model classification performance for each class. It is a harmonic mean of precision and recall, defined as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (5)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (6)$$

$$F1 = \frac{2 \times (\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}} \quad (7)$$

The terms TP, FP and FN refer to the number of true positive, false positive and false negative samples, respectively. For the overall evaluation, we use the macro F1 score, i.e., the arithmetic mean of F1 scores of all classes and the area under the ROC curve (AUC).

4. Experiment and Results

4.1. Results on the PTB-XL Dataset

The results are reported in Table 3. The average F1 score reaches 0.833 and the average AUC value reaches 0.925. The precision values of all classes are larger than 0.810 and the recall values of all classes except the MI are also larger than 0.810. This is because the MI superclass mainly includes two types of subclasses: AMI and IMI. AMI is mainly caused by V1-V6 leads and IMI is mainly caused by II,III,aVF. The difference between the two leads affects the judgment of our model. Among all classes, the model performs best for CD and HYP with AUC scores exceeding 0.930. Figure 5 shows the overall classification performance of this method and seven previous studies, which are all single-task classification. The highest F1 score produced by other studies reaches 0.823, which is lower than 0.833.

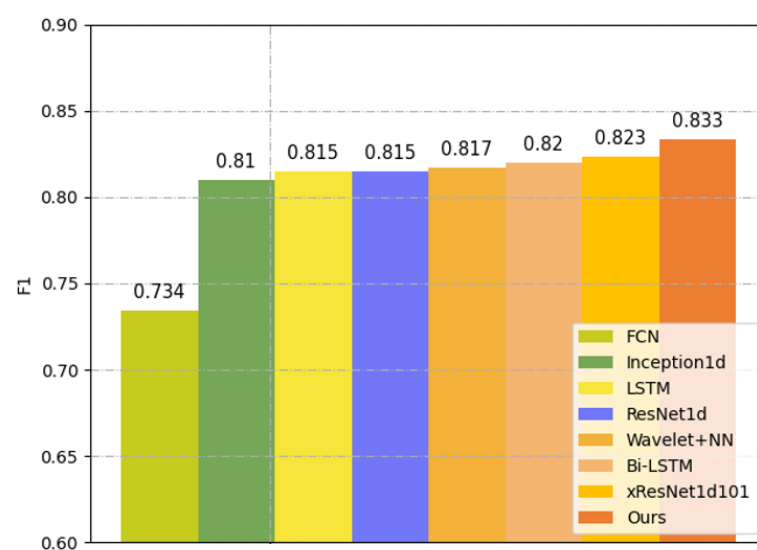


Figure 5. Classification performance of the proposed method and seven other methods (FCN [11], Inception1d [12], LSTM [35], ResNet1d [25], Wavelet+NN [13], Bi-LSTM [37], and xResNet1d101 [26]) on the PTB-XL dataset.

Table 3. Classification performance of the proposed method on the PTB-XL dataset.

Type	AUC	Accuracy	Precision	Recall	F1
NORM	0.924	0.905	0.877	0.849	0.863
MI	0.898	0.912	0.814	0.734	0.772
STTC	0.926	0.869	0.834	0.813	0.823
CD	0.946	0.868	0.867	0.872	0.869
HYP	0.933	0.883	0.852	0.819	0.835
AVG	0.925	0.887	0.849	0.817	0.833

4.2. Results on the CPSC2018 Dataset

The results on the CPSC2018 dataset are reported in Table 4. Overall, the proposed method yields an average AUC value of 0.977, an average accuracy of 0.966, an average F1 score of 0.827, an average precision of 0.852 and an average recall of 0.808. Among all arrhythmias, the model performs best for LBBB and RBBB with F1 scores exceeding 0.930.

Table 4. The classification performance of the proposed method in CPSC2018.

Type	AUC	Accuracy	Precision	Recall	F1
SNR	0.973	0.948	0.824	0.824	0.824
AF	0.981	0.962	0.885	0.968	0.925
IABV	0.981	0.977	0.923	0.845	0.882
LBBB	0.999	0.996	0.957	0.917	0.937
RBBB	0.994	0.965	0.944	0.934	0.939
PAC	0.961	0.951	0.712	0.758	0.734
PVC	0.962	0.959	0.885	0.676	0.767
STD	0.984	0.962	0.868	0.805	0.835
STE	0.960	0.977	0.667	0.545	0.600
AVG	0.977	0.966	0.852	0.808	0.827

Table 5 shows per-class and overall classification performance of this method and five previous studies. The highest F1 score produced by other studies reaches 0.813, which is lower than 0.827. Our model is superior to other models in the diagnosis of SNR, IABV, LBBB, RBBB, STD and STE.

Table 5. Comparison for classification performance of previous works and ours evaluated on the recommended test set of CPSC2018.

Model, Year	F1									
	SNR	AF	IABV	LBBB	RBBB	PAC	PVC	STD	STE	AVG
CNN + LSTM [30], 2018	0.753	0.900	0.809	0.874	0.922	0.638	0.832	0.762	0.462	0.772
CNN + LSTM [31], 2020	-	-	-	-	-	-	-	-	-	0.806
CNN + Attention [32], 2019	0.790	0.930	0.850	0.860	0.930	0.750	0.850	0.800	0.560	0.813
CNN + LSTM + Attention [33], 2020	0.789	0.920	0.850	0.872	0.933	0.736	0.861	0.789	0.556	0.812
Interpretable ResNet [47], 2021	0.805	0.919	0.864	0.866	0.926	0.735	0.851	0.814	0.535	0.813
Ours	0.824	0.925	0.882	0.937	0.939	0.734	0.767	0.835	0.600	0.827

4.3. Effect of Random Auxiliary Task

We use two strategies when creating auxiliary tasks. The first is to merge similar classes into superclasses (i.e., CPSC2018 dataset) and the second is to split one class into more subclasses (i.e., PTB-XL dataset). Obviously, the classes of auxiliary task are still organized hierarchically.

In order to verify the effectiveness of the proposed strategies, we conduct the experiments with random auxiliary tasks on the CPSC2018 dataset and the PTB-XL dataset, respectively, e.g., we randomly merge classes into superclasses which do not make sense from a clinic perspective. As shown in Tables 6 and 7, the average F1 score for CPSC2018 and PTB-XL datasets reaches 0.807 and 0.818, respectively, which is comparable to that produced by previous single-task classification and is still lower than that produced with proposed auxiliary task.

Table 6. Classification performance with the random auxiliary task on the CPSC2018 dataset.

Type	AUC	Accuracy	Precision	Recall	F1
SNR	0.974	0.952	0.848	0.824	0.836
AF	0.980	0.965	0.866	0.915	0.890
IABV	0.983	0.975	0.897	0.859	0.878
LBBB	0.999	0.996	0.957	0.917	0.937
RBBB	0.993	0.969	0.936	0.960	0.948
PAC	0.952	0.948	0.671	0.823	0.739
PVC	0.969	0.962	0.809	0.765	0.786
STD	0.980	0.956	0.777	0.890	0.830
STE	0.956	0.968	0.500	0.318	0.389
AVG	0.976	0.966	0.807	0.808	0.807

Table 7. Classification performance with the random auxiliary task on the PTB-XL dataset.

Type	AUC	Accuracy	Precision	Recall	F1
NORM	0.917	0.889	0.845	0.839	0.842
MI	0.883	0.907	0.807	0.699	0.749
STTC	0.923	0.863	0.840	0.781	0.809
CD	0.935	0.859	0.861	0.865	0.863
HYP	0.925	0.879	0.851	0.806	0.828
AVG	0.917	0.879	0.841	0.798	0.818

4.4. Ablation Experiment

In order to prove the effectiveness of the CoT attention mechanism and the Bi-GRU layer, we conduct ablation experiments about these two modules. Results produced on the PTB-XL and CPSC2018 datasets by our model without the CoT block are shown in Tables 8 and 9. It can be observed that metrics have decreased overall. The F1 score decreases from 0.833 to 0.819 on the PTB-XL dataset and from 0.827 to 0.761 on the CPSC2018 dataset.

Table 8. Results produced by the proposed model without the CoT block on the PTB-XL dataset.

Type	AUC	Accuracy	Precision	Recall	F1
NORM	0.908	0.893	0.870	0.815	0.842
MI	0.881	0.904	0.782	0.735	0.758
STTC	0.917	0.865	0.824	0.816	0.820
CD	0.935	0.859	0.858	0.861	0.859
HYP	0.911	0.869	0.827	0.808	0.817
AVG	0.910	0.878	0.832	0.807	0.819

Table 9. Results produced by the proposed model without the CoT block on the CPSC2018 dataset.

Type	AUC	Accuracy	Precision	Recall	F1
SNR	0.973	0.946	0.835	0.794	0.814
AF	0.979	0.958	0.860	0.868	0.864
IABV	0.979	0.971	0.905	0.803	0.851
LBBB	0.998	0.988	0.750	0.900	0.818
RBBB	0.992	0.964	0.918	0.960	0.939
PAC	0.956	0.948	0.741	0.645	0.690
PVC	0.960	0.961	0.918	0.662	0.769
STD	0.983	0.951	0.809	0.878	0.842
STE	0.941	0.957	0.444	0.182	0.258
AVG	0.973	0.960	0.798	0.744	0.761

Results produced on the PTB-XL and CPSC2018 datasets by our model without the Bi-GRU layer are shown in Tables 10 and 11. It can be also observed that metrics decrease overall. The F1 score decreases from 0.833 to 0.820 on the PTB-XL dataset and decreases from 0.827 to 0.807 on the CPSC2018 dataset, though the accuracy decrease is not as much as that produced by removing the CoT block.

Table 10. Results produced by the proposed model without the Bi-GRU module on the PTB-XL dataset.

Type	AUC	Accuracy	Precision	Recall	F1
NORM	0.904	0.892	0.855	0.832	0.843
MI	0.877	0.899	0.766	0.749	0.757
STTC	0.914	0.869	0.835	0.810	0.822
CD	0.931	0.861	0.859	0.860	0.859
HYP	0.911	0.874	0.850	0.789	0.818
AVG	0.907	0.879	0.833	0.808	0.820

Table 11. Results produced by the proposed model without the Bi-GRU module on the CPSC2018 dataset.

Type	AUC	Accuracy	Precision	Recall	F1
SNR	0.964	0.939	0.768	0.843	0.804
AF	0.989	0.965	0.918	0.849	0.882
IABV	0.982	0.975	0.922	0.831	0.874
LBBB	0.998	0.988	0.750	0.900	0.818
RBBB	0.991	0.955	0.935	0.944	0.939
PAC	0.963	0.948	0.716	0.774	0.744
PVC	0.952	0.952	0.893	0.735	0.806
STD	0.982	0.965	0.890	0.793	0.839
STE	0.961	0.964	0.571	0.545	0.558
AVG	0.976	0.961	0.818	0.802	0.807

5. Discussion

Experimental results demonstrate the effectiveness of our proposed method. As shown in Table 5 and Figure 5, our method produced an overall F1 score of 0.833 and 0.827 on the PTB-XL and CPSC2018 datasets, respectively, which are higher than the optimal F1 scores produced by other methods by 0.01 and 0.014. Among different classes of the CPSC2018 dataset, as shown in Table 5, our method shows advantages for SNR, LBBB, STD, STE and is inferior to other methods for AF, PVC, etc. In addition, as shown in Table 3, the F1 score of MI is lower than the other four superclasses. Therefore, we need to try different preprocessing techniques (including different sampling rates or noise reduction methods, etc.) to improve the classification performance of MI in the next work.

We find that the auxiliary task has an impact on the main task. For example, LBBB and RBBB belong to the same superclass in the auxiliary task and our method yields similar scores (0.937 and 0.939) for them, while there is a large gap between the two scores for other methods. When the correlation between the main task and auxiliary task is high, the shared SE-ResNet can be well trained in parallel by two tasks. As shown in Tables 6 and 7, by contrast, the correlation between the main task and random auxiliary task is relatively small, the advantage of multi-task learning cannot be exhibited, indicating the effectiveness of the proposed strategy for creating auxiliary tasks.

Tables 8–11 demonstrate the effectiveness of the CoT attention mechanism and the Bi-GRU layer. CoT block can simultaneously extract key features of static local information and global dynamic information from ECG sequences. Although the Bi-GRU module also plays an important role in processing sequence information, the accuracy drop caused by removing the Bi-GRU layer is not as much as that caused by removing the CoT block.

In addition, to make the results more distinct, we use three decimals consistently for all results including F1 scores in this study. Two decimals may lead to indistinguishable comparison, e.g., three methods would have the same F1 score of 0.81 in Table 5. Although a value at the third decimal may be insignificant in the statistical sense, many studies still presented F1 scores with three decimals for better comparison of different methods [11,30,47].

6. Conclusions

In this work, a new multi-task deep neural network, which includes a shared low-level feature extraction module (i.e., SE-ResNet) and a task-specific classification module, is proposed. Contextual Transformer (CoT) block is introduced in the classification module to solve the problem that the self-attention mechanism only focuses on the local information

of the ECG feature sequence. Thus, the proposed method can dynamically model the local and global information of the ECG feature sequence. In addition, we propose to create auxiliary tasks by merging similar classes or splitting large classes, which exploit the hierarchical class information. The proposed method produced an overall F1 score of 0.833 and 0.827 on the PTB-XL and CPSC2018 datasets, respectively, which are higher than the optimal F1 scores produced by other methods by 0.01 and 0.014, suggesting the effectiveness of our method. However, our multi-task method still has some limitations. Specifically, our model shows bad performance in detecting some abnormal classes, such as detecting the superclass MI on the PTB-XL dataset. In addition, we only use hard parameter sharing to achieve multi-tasking without considering the soft parameter implementation. Finally, we hope that the arrhythmia detection method proposed in this paper can play a role in the process of early diagnosis of cardiovascular diseases. In the future, we will consider integrating more auxiliary tasks for improving the performance of the main task and investigate other approaches to creating auxiliary tasks. With more tasks, we also need to improve the network structure for computational efficiency.

Author Contributions: Conceptualization, Q.G. and H.L.; methodology, Q.G., H.L. and T.G.; validation, Q.G., H.L. and T.G.; formal analysis, R.L., C.C. and M.S.; investigation, T.G. and Q.Z.; resources, T.G. and Q.Z.; data curation, M.S.; writing—original draft preparation, Q.G. and H.L.; writing—review and editing, Q.G. and H.L.; visualization, Q.G. and H.L.; supervision, M.S.; project administration, M.S.; funding acquisition, R.L. and C.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Shandong Provincial Natural Science Foundation ZR2020QF020 and ZR2021QF125.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The ECG signal data used to support the findings of this study have been deposited in the PTB-XL repository (<https://www.physionet.org/content/ptb-xl/1.0.3/>, accessed on 11 January 2023) and CPSC 2018 (<http://2018.icbeb.org/Challenge.html>, accessed on 11 January 2023).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Reddy, K.S. Cardiovascular diseases. In *The Nutrition Transition*; Elsevier: Amsterdam, The Netherlands, 2002; pp. 191–203.
2. Luo, K.; Li, J.; Wang, Z.; Cuschieri, A. Patient-specific deep architectural model for ECG classification. *J. Healthc. Eng.* **2017**, *2017*, 4108720. [[CrossRef](#)] [[PubMed](#)]
3. De Chazal, P.; O'Dwyer, M.; Reilly, R.B. Automatic classification of heartbeats using ECG morphology and heartbeat interval features. *IEEE Trans. Biomed. Eng.* **2004**, *51*, 1196–1206. [[CrossRef](#)] [[PubMed](#)]
4. Ye, C.; Vijaya Kumar, B.V.K.; Coimbra, M.T. Heartbeat Classification Using Morphological and Dynamic Features of ECG Signals. *Biomed. Eng. IEEE Trans.* **2012**, *59*, 2930–2941.
5. Glass, L.; Micheli-Tzanakou, E. *Cardiac Oscillations and Arrhythmia Analysis*; Springer: New York, NY, USA, 2006.
6. Jambukia, S.H.; Dabhi, V.K.; Prajapati, H.B. Classification of ECG signals using machine learning techniques: A survey. In Proceedings of the 2015 International Conference on Advances in Computer Engineering and Applications, Ghaziabad, India, 19–20 March 2015; IEEE: New York, NY, USA, 2015; pp. 714–721.
7. Kiranyaz, S.; Ince, T.; Gabbouj, M. Real-time patient-specific ECG classification by 1-D convolutional neural networks. *IEEE Trans. Biomed. Eng.* **2015**, *63*, 664–675. [[CrossRef](#)]
8. Kachuee, M.; Fazeli, S.; Sarrafzadeh, M. Ecg heartbeat classification: A deep transferable representation. In Proceedings of the 2018 IEEE International Conference on Healthcare Informatics (ICHI), New York, NY, USA, 4–7 June 2018; IEEE: Los Alamitos, CA, USA, 2018; pp. 443–444.
9. Baloglu, U.B.; Talo, M.; Yildirim, O.; San Tan, R.; Acharya, U.R. Classification of myocardial infarction with multi-lead ECG signals and deep CNN. *Pattern Recognit. Lett.* **2019**, *122*, 23–30. [[CrossRef](#)]
10. Liu, W.; Huang, Q.; Chang, S.; Wang, H.; He, J. Multiple-feature-branch convolutional neural network for myocardial infarction diagnosis using electrocardiogram. *Biomed. Signal Process. Control* **2018**, *45*, 22–32. [[CrossRef](#)]
11. Strodthoff, N.; Wagner, P.; Schaeffter, T.; Samek, W. Deep learning for ECG analysis: Benchmarks and insights from PTB-XL. *IEEE J. Biomed. Health Inform.* **2020**, *25*, 1519–1528. [[CrossRef](#)]

12. Ismail Fawaz, H.; Lucas, B.; Forestier, G.; Pelletier, C.; Schmidt, D.F.; Weber, J.; Webb, G.I.; Idoumghar, L.; Muller, P.A.; Petitjean, F. Inceptiontime: Finding alexnet for time series classification. *Data Min. Knowl. Discov.* **2020**, *34*, 1936–1962. [\[CrossRef\]](#)
13. Sharma, L.D.; Sunkaria, R.K. Inferior myocardial infarction detection using stationary wavelet transform and machine learning approach. *Signal Image Video Process.* **2018**, *12*, 199–206. [\[CrossRef\]](#)
14. Martis, R.J.; Acharya, U.R.; Prasad, H.; Chua, C.K.; Lim, C.M.; Suri, J.S. Application of higher order statistics for atrial arrhythmia classification. *Biomed. Signal Process. Control* **2013**, *8*, 888–900. [\[CrossRef\]](#)
15. Osowski, S.; Hoai, L.T.; Markiewicz, T. Support vector machine-based expert system for reliable heartbeat recognition. *IEEE Trans. Biomed. Eng.* **2004**, *51*, 582–589. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Frénay, B.; De Lannoy, G.; Verleysen, M. Improving the transition modelling in hidden Markov models for ECG segmentation. In Proceedings of the ESANN, Bruges, Belgium, 22–24 April 2009.
17. Afkhami, R.G.; Azarnia, G.; Tinati, M.A. Cardiac arrhythmia classification using statistical and mixture modeling features of ECG signals. *Pattern Recognit. Lett.* **2016**, *70*, 45–51. [\[CrossRef\]](#)
18. Chen, S.; Hua, W.; Li, Z.; Li, J.; Gao, X. Heartbeat classification using projected and dynamic features of ECG signal. *Biomed. Signal Process. Control* **2017**, *31*, 165–173. [\[CrossRef\]](#)
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. [\[CrossRef\]](#)
20. Jun, T.J.; Park, H.J.; Minh, N.H.; Kim, D.; Kim, Y.H. Premature ventricular contraction beat detection with deep neural networks. In Proceedings of the 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA), Anaheim, CA, USA, 18–20 December 2016; IEEE: New York, NY, USA, 2016; pp. 859–864.
21. Sellami, A.; Hwang, H. A robust deep convolutional neural network with batch-weighted loss for heartbeat classification. *Expert Syst. Appl.* **2019**, *122*, 75–84. [\[CrossRef\]](#)
22. Banerjee, S.; Mitra, M. ECG feature extraction and classification of anteroseptal myocardial infarction and normal subjects using discrete wavelet transform. In Proceedings of the 2010 International Conference on Systems in Medicine and Biology, Kharagpur, India, 16–18 December 2010; IEEE: New York, NY, USA, 2010; pp. 55–60.
23. Monasterio, V.; Laguna, P.; Martinez, J.P. Multilead analysis of T-wave alternans in the ECG using principal component analysis. *IEEE Trans. Biomed. Eng.* **2009**, *56*, 1880–1890. [\[CrossRef\]](#)
24. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [\[CrossRef\]](#)
25. Wang, Z.; Yan, W.; Oates, T. Time series classification from scratch with deep neural networks: A strong baseline. In Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, USA, 14–19 May 2017; IEEE: New York, NY, USA, 2017; pp. 1578–1585.
26. He, T.; Zhang, Z.; Zhang, H.; Zhang, Z.; Xie, J.; Li, M. Bag of tricks for image classification with convolutional neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 558–567.
27. Zeiler, M.D.; Fergus, R. Visualizing and understanding convolutional networks. In *Proceedings of the Computer Vision—ECCV 2014: 13th European Conference, Zurich, Switzerland, 6–12 September 2014, Proceedings, Part I 13*; Springer: Berlin, Germany, 2014; pp. 818–833.
28. Sharif Razavian, A.; Azizpour, H.; Sullivan, J.; Carlsson, S. CNN features off-the-shelf: An astounding baseline for recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Columbus, OH, USA, 23–28 June 2014; pp. 806–813.
29. Al Rahhal, M.M.; Bazi, Y.; Al Zuair, M.; Othman, E.; BenJdira, B. Convolutional neural networks for electrocardiogram classification. *J. Med. Biol. Eng.* **2018**, *38*, 1014–1025. [\[CrossRef\]](#)
30. Yao, Q.; Fan, X.; Cai, Y.; Wang, R.; Yin, L.; Li, Y. Time-incremental convolutional neural network for arrhythmia detection in varied-length electrocardiogram. In Proceedings of the 2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech), Athens, Greece, 12–15 August 2018; IEEE: New York, NY, USA, 2018; pp. 754–761.
31. He, R.; Liu, Y.; Wang, K.; Zhao, N.; Yuan, Y.; Li, Q.; Zhang, H. Automatic cardiac arrhythmia classification using combination of deep residual network and bidirectional LSTM. *IEEE Access* **2019**, *7*, 102119–102135. [\[CrossRef\]](#)
32. Wang, R.; Yao, Q.; Fan, X.; Li, Y. Multi-class arrhythmia detection based on neural network with multi-stage features fusion. In Proceedings of the 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), Bari, Italy, 6–9 October 2019; IEEE: New York, NY, USA, 2019; pp. 4082–4087.
33. 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. [\[CrossRef\]](#)
34. Hannun, A.Y.; Rajpurkar, P.; Haghpanahi, M.; Tison, G.H.; Bourn, C.; Turakhia, M.P.; Ng, A.Y. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat. Med.* **2019**, *25*, 65–69. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [\[CrossRef\]](#) [\[PubMed\]](#)

36. Übeyli, E.D. Recurrent neural networks employing Lyapunov exponents for analysis of ECG signals. *Expert Syst. Appl.* **2010**, *37*, 1192–1199. [CrossRef]
37. Yildirim, Ö. A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification. *Comput. Biol. Med.* **2018**, *96*, 189–202. [CrossRef]
38. Zhang, J.; Liu, A.; Gao, M.; Chen, X.; Zhang, X.; Chen, X. ECG-based multi-class arrhythmia detection using spatio-temporal attention-based convolutional recurrent neural network. *Artif. Intell. Med.* **2020**, *106*, 101856. [CrossRef]
39. Ji, J.; Chen, X.; Luo, C.; Li, P. A deep multi-task learning approach for ECG data analysis. In Proceedings of the 2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), Las Vegas, NV, USA, 4–7 March 2018; IEEE: New York, NY, USA, 2018; pp. 124–127.
40. Collobert, R.; Weston, J. A unified architecture for natural language processing: Deep neural networks with multitask learning. In Proceedings of the 25th International Conference on Machine Learning, Helsinki, Finland, 5–9 July 2008; pp. 160–167.
41. Deng, L.; Hinton, G.; Kingsbury, B. New types of deep neural network learning for speech recognition and related applications: An overview. In Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 26–31 May 2013; IEEE: New York, NY, USA, 2013; pp. 8599–8603.
42. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
43. Taigman, Y.; Yang, M.; Ranzato, M.; Wolf, L. Deepface: Closing the gap to human-level performance in face verification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 24–27 June 2014; pp. 1701–1708.
44. Ruder, S. An overview of multi-task learning in deep neural networks. *arXiv* **2017**. arXiv:1706.05098.
45. Liu, F.; Liu, C.; Zhao, L.; Zhang, X.; Wu, X.; Xu, X.; Liu, Y.; Ma, C.; Wei, S.; He, Z.; et al. An open access database for evaluating the algorithms of electrocardiogram rhythm and morphology abnormality detection. *J. Med. Imaging Health Inform.* **2018**, *8*, 1368–1373. [CrossRef]
46. Wagner, P.; Strodthoff, N.; Bousseljot, R.; Samek, W.; Schaeffter, T. PTB-XL, a Large Publicly Available Electrocardiography Dataset (Version 1.0.3), Physionet. 2020. Available online: <https://doi.org/10.13026/x4td-x982> (accessed on 11 January 2023)
47. Zhang, D.; Yang, S.; Yuan, X.; Zhang, P. Interpretable deep learning for automatic diagnosis of 12-lead electrocardiogram. *Iscience* **2021**, *24*, 102373. [CrossRef]
48. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
49. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
50. 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*, 5998–6008.
51. Li, Y.; Yao, T.; Pan, Y.; Mei, T. Contextual transformer networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 1489–1500. [CrossRef] [PubMed]
52. Lynn, H.M.; Pan, S.B.; Kim, P. A deep bidirectional GRU network model for biometric electrocardiogram classification based on recurrent neural networks. *IEEE Access* **2019**, *7*, 145395–145405. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.