Redundancy analysis of 12-lead ECG in automatic arrhythmia recognition

Shaopeng Pang*[1], Jiahao Li[1], Fangzhou Xu†[2], and Shuwang Zhou[3]

[1]School of Information and Automation Engineering, Qilu University of Technology(Shandong Academy of Sciences), Jinan, Shandong Province, 250353, China
[2]International School for Optoelectronic Engineering, Qilu University of Technology (Shandong Academy of Sciences), Jinan, 250353, China
[3]Qilu University of Technology (Shandong Academy of Sciences), Shandong Artificial Intelligence Institute, Jinan 250014, China

Abstract

Arrhythmia is one of the most common cardiovascular diseases, which seriously endangers human health. Clinically, the 12-lead ECG is an important basis for professional doctors to diagnose the type of arrhythmias. In recent years, various neural networks with different structures have achieved great success in automatic arrhythmia recognition using the 12-lead ECG as input. However, different leads in 12-lead ECG may have different contributions to arrhythmia recognition. This inspired us to study the redundancy of lead through 12-lead ECG with different erasure combinations, where the information of a lead is erased by padding it with zeros. Then 5 deep neural networks (DNNs) with different structures are constructed to observe the effects of the absence of different lead combinations on the recognition performance. The simulation result shows that the erasure of partial leads from 12-lead ECG reduces the recognition performance of DNN models compared to using the complete 12-lead ECG as input. But the reduction is very limited, especially when only the leads I, II, and III are erased. The values of the leads I, II, and III can be derived from other leads, which may be the reason for their redundancy in the automatic arrhythmia recognition. This suggests that it is reasonable to erase appropriate leads to reduce the computational complexity and apply it to portable devices for automatic arrhythmia recognition.

Keywords: 12-lead ECG, arrhythmia recognition, deep neural network, redundancy

1 Introduction

Cardiovascular disease is one of the leading causes of increased mortality worldwide [1]. The 12-lead ECG accurately records the physiological activity of the heart over time and is one of the most common non-invasive diagnostic tools. Cardiologists diagnose a variety of arrhythmias based on the 12-lead ECG, including Atrial fibrillation (AF), First-degree atrioventricular block (I-AVB), Right bundle branch block (RBBB), and Premature ventricular contraction (PVC), etc. With the improvement of computing power, automatic arrhythmia recognition algorithms based on big data and artificial intelligence continue to emerge to assist doctors in diagnosis. The clinical measurements of the 12-lead ECG rely on a ten-electrode system. However, in recent years, portable ECG monitors such as the Holter monitors [2] and wearable limb 6-lead ECG measurement devices [3] cannot obtain complete 12-lead ECG data due to the reduction of sensors. Therefore, it is meaningful to study the redundancy of leads, that is, the effect of the absence of different leads on automatic arrhythmia recognition.

Deep learning has the property of automatically extracting complex data features. With the massive increase in the amount of clinical data, deep learning has been widely used in the diagnosis of cardiac abnormalities [4] in recent years, such as arrhythmia recognition, ECG age prediction [5], heart failure (HF) identification [6], left ventricular dysfunction prediction [7], etc. These studies demonstrate that deep learning is prospective for the diagnosis of cardiac abnormalities. In addition, a small number of works have investigated the redundancy of 12-lead ECG

*Corresponding author email: pang_shao_peng@163.com
†Corresponding author email: xfz@qlu.edu.cn

in automatic arrhythmia recognition. Chen et al. [8] developed a convolutional neural network (CNN) model to recognize arrhythmias. The simulation results showed that the performance of automatic arrhythmia recognition using only single-lead ECG was lower than using the complete 12-lead ECG. Among them, leads aVR and V1 are the most prominent. Chiou et al. [6] transformed each lead into a two-dimensional spectrum by using continuous wavelet transform, and designed a two-dimensional (2D) CNN model to recognize systolic HF. The results showed that lead V6 achieved the highest accuracy, sensitivity, specificity, and F_1 in single-lead mode, while the combination of lead V5 and lead V6 performed best in multi-lead mode. Cho et al. [9] developed a variational auto-encoder and found that 6-lead limb-based diagnosis of myocardial infarction was possible. Chen et al. [10] proposed the multi-branch convolution and residual network to recognize arrhythmias. By comparing the recognition performance of single-lead and 8-lead fusion, they found that 8-lead fusion can improve the recognition performance. Furthermore, the PhysioNet has launched a lead redundancy exploration-based competition called "Computing in Cardiology Challenge 2021" [15], which asked participants to build an algorithm that can use 12-lead, 6-lead, 4-lead, 3-lead, and 2-lead to diagnose cardiac abnormalities. Some studies in this competition are instructive. Philip et al. [11] based on Scattering-Recurrent Networks performed automatic identification of ECG with reduced leads. They found that the performance of the model decreased slightly with the reduction of the number of leads, which indicated that there was a strong correlation between leads. The classification results obtained by Jessen et al. [12] using the shared classifier showed better performance in lead reduction compared with the original 12-lead ECG, which may be related to the characteristics of their designed model. Matteo et al. [13] used AutoML to classify ECG signals from different combinations of leads, and cumulative challenge scores across multiple AutoML instances indicated the best performance for 6-leads, followed by 3-leads, 2-leads, 4-leads, and 12-leads. Niels et al. [13] used Convolutional Recurrent Neural Networks to identify the ECG information abnormalities of different combinations of leads, and found that the use of the 2-leads achieved better classification performance. The competition results show that the performance of different lead combinations in the hidden test set is quite different.

The existing works [8–15] on the redundancy analysis of 12-lead ECG usually use ECG data with different dimensions as input to train DNN models. The parameters of DNN (such as the number of filters, and convolution kernel size) need to be adjusted for different input dimensions, which leads to factors that affect the results from the input data and model parameters. Under such conditions, it is unreasonable to compare the effect of the absence of different leads on automatic arrhythmia recognition. Therefore, in this paper, we unify the input of the DNN models as 12-lead ECG and erase the information of a lead by zero padding. This can effectively guarantee the consistency of the input dimensions. Then 5 DNNs with different structures were constructed to observe the effect of the absence of different lead combinations on the recognition performance. Different from the combination of leads selected in the 2021 competition [15], the lead combinations we selected mainly depend on the properties between leads, such as removal of chest leads, removal of bipolar limb leads, removal of augmented limb leads, etc.

2 Methodology

2.1 12-lead ECG

The ECG is one of the most widely used examinations in the clinic. In order to fully reflect the electrical activity of the whole heart, the 12-lead body surface ECG is often used clinically. The 12-lead ECG includes 6 limb leads and 6 chest leads, which record the electrical activity of the heart in the frontal plane and horizontal plane, respectively. The 6 limb leads include the 3 bipolar limb leads I, II and III, the 3 augmented limb leads aVR, aVL and aVF, and the 6 chest leads include V1, V2, V3, V4, V5 and V6.

The measurement of 12-lead can be realized by collecting human body signals from 10 electrodes. Fig. 1(a) shows the position of the 10 electrodes on the human body. The numerical calculation and formula deformation of the standard 12-lead ECG are shown in Fig. 1(b). It can be found that leads II, aVR, aVL and aVF can be calculated by leads I and III.

2.2 Data set

The 12-lead ECG sample used in this paper is from 4 open-access databases provided by PhysioNet [16]. All sample signals are sampled to 500Hz. The 4 databases are China Physiological Signal Challenge in 2018 database (CPSC2018) [17], CPSC2018 unused database (CPSC-Extra) [17], the Physikalisch Technische Bundesanstalt electrocardiography database (PTB-XL) [18] and Georgia 12-lead ECG challenge database (G12EC). The details of the 4 databases are in Appendix A.

The types of arrhythmias contained in these databases are not exactly the same. Referring to the 9 cate-

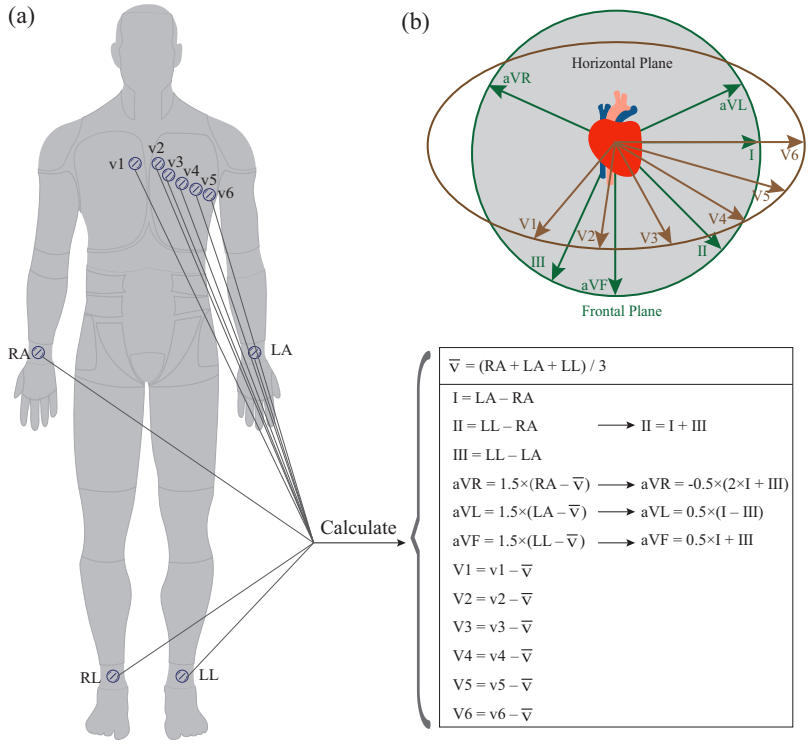

Fig. 1 Electrode position and ECG calculation formula. (a) The position of the 10 electrodes on the human body. (b) The numerical calculation and formula deformation of the 12-lead ECG. The 6 limb leads record the potential difference in the frontal plane of the heart, and the 6 chest leads record the potential difference in the horizontal plane of the heart.

Table 1 The number of normal rhythm and 8 types of arrhythmias in the 4 databases after excluding abnormal samples.

Label	Types of arrhythmia	CPSC2018	CPSC-Extra	PTB-XL	G12EC	Total
1	Normal	898	-	1732	7871	10501
2	AF	1219	40	234	525	2018
3	IAVB	721	96	338	148	1303
4	LBBB	199	9	199	365	772
5	RBBB	1671	-	388	-	2059
6	PAC	542	102	265	132	1041
7	PVC	626	-	27	469	1122
8	STD	784	29	3	-	816
9	STE	185	35	10	21	251
Total		6845	311	3196	9531	19883

gories included in CPSC2018, we extracted 12-lead ECG sample containing the same 9 categories from the other 3 databases. The 9 categories include Normal, Atrial fibrillation (AF), First-degree atrioventricular block (I-AVB), Left bundle branch block (LBBB), Right bundle branch block (RBBB), Premature atrial contraction (PAC), Premature ventricular contraction (PVC), ST-segment depression (STD) and ST-segment elevated (STE). Meanwhile, the 12-lead ECG sample with abnormal age (i.e., null) and original signal length ¡ 8s are also excluded. The effective samples of normal rhythm and 8 categories of arrhythmias are shown in Table. 1. We obtained a total of 19883 sets of 12-lead ECG samples. Each set of samples contains scalar values for the 12-lead ECG, and the corresponding age, sex, and label (arrhythmia type).

2.3 Lead Selection

According to the relationship formulas of 12 leads in Fig. 1(b), the values of the leads II, aVR, aVL, and aVF can be derived from the leads I and III. The values can be calculated indirectly, indicating that these 4 leads may have some redundancy in the automatic arrhythmia recognition. Then, based on the 7 DNN models published in the CPSC2018 [17], it was found that when erasing (padding 0) a single lead information one by one, erasing the lead aVL had the least impact on the arrhythmia recognition (see Appendix B). This inspired us to study the redundancy of the lead aVL. In addition, we investigated the effect of bipolar limb leads, augmented limb leads, and chest leads on the arrhythmia recognition after their information was erased. Therefore, we use the method of filling 0 to erase the information of one or multiple leads to explore the effect of missing one lead or different lead combinations on the automatic arrhythmia recognition under the same DNN model. Specifically, the 6 lead combinations with erased information, from less to more, are as follows:

(case-1) No lead signal is erased.

(case-2) The lead aVL is filled with 0.

(case-3) The bipolar limb lead is erased. The leads I, II and III are filled with 0.

(case-4) The augmented limb leads are erased. The leads aVR, aVL, and aVF are filled with 0.

(case-5) The leads for which values can be calculated are erased. The leads II, aVR, aVL, and aVF are filled with 0.

(case-6) The chest lead is erased. The leads V1, V2, V3, V4, V5, and V6 are filled with 0.

2.4 Block

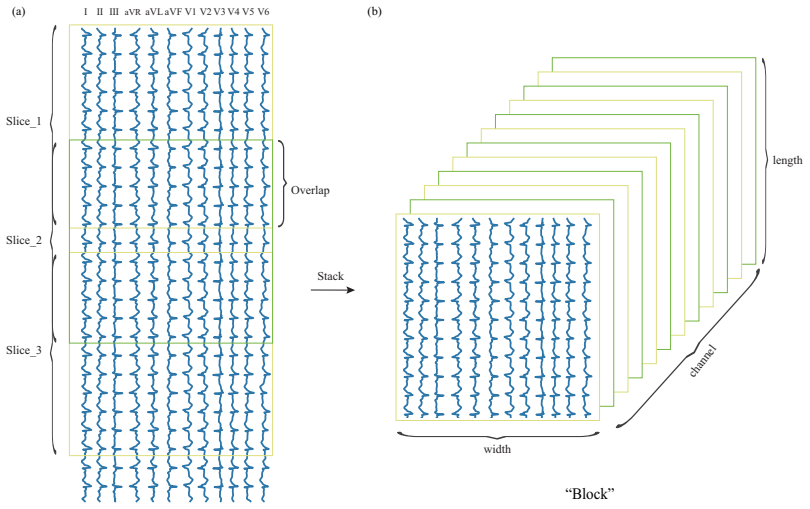

Fig. 2 Illustration of slicing and stacking. (a) The slicing method of the 12-lead ECG. (b) Blocks stacked from the slices of 12-lead ECG, which are the input of DNN models.

The lengths of the 12-lead ECG sample signals from different databases are not the same. In order to ensure that the 12-lead ECG signals feed into the DNN have the same dimension, we need manual intervention to keep the input dimension consistent for time-series data with different lengths. Common methods include padding 0 to a specified length for samples that are too short, and truncating samples that are too long to ensure that all samples have the same length. However, the waveforms of different time segments of the 12-lead ECG signal are different.

Filled samples may destroy the structure of the original 12-lead signal, and cropped samples may erase the key waveforms used to determine the type of arrhythmia. Therefore, it is not appropriate to keep the sample length consistent by padding or truncating.

We propose a segmentation method for slicing and stacking the 12-lead ECG, which guarantees the uniform input dimension. As shown in Fig. 2, we overlap-cut the 12-lead ECG with the same length and stack them into a 'Block'. Each Block has 3 dimensions: length, width, and channel, where length is the length of the slice (8.192s), width is the number of leads, and channel represents the number of slices. Each slice contains 12-lead ECG of different time periods, and the overlapping and stacking of slices can preserve the information of the original 12-lead ECG to the greatest extent.

2.5 Model

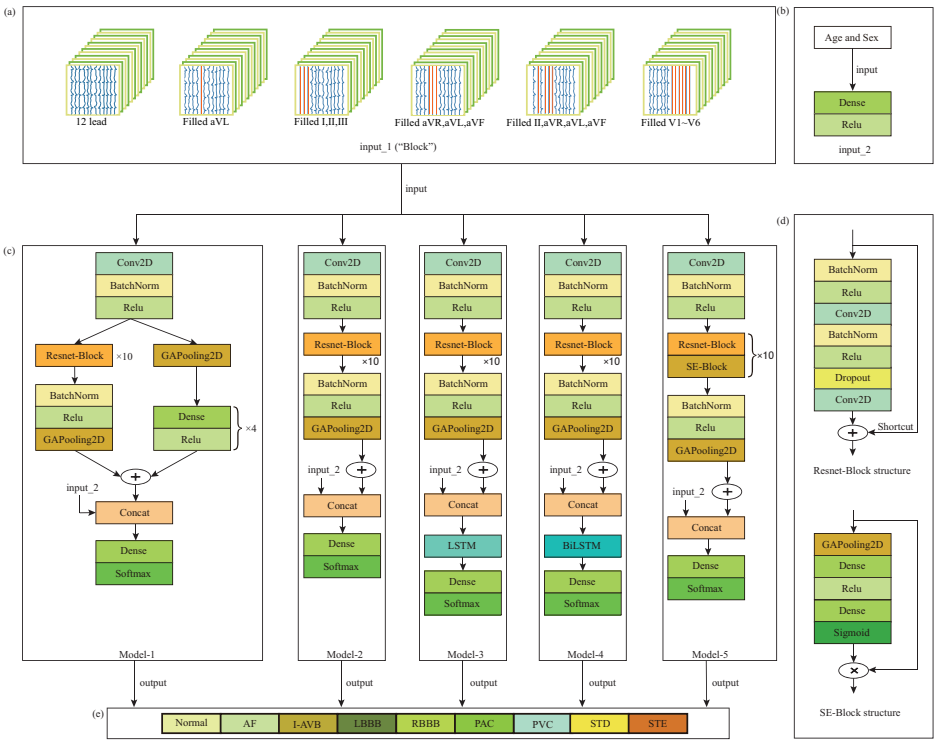

Fig. 3 Model structure. (a) The first input source of DNNs. It contains 6 different erasing methods. (b) The second input source of DNNs. The age and gender of the patient are used as the second input source of the model. (c) DNN models of different structures. These deep neural networks consist of 21 2D convolutional layers and a fully connected layer, and the mapping of classification probabilities is implemented by Softmax. (d) The constituent unit of Resnet-Block and SE-Block. (f) Classification category of DNNs. It includes normal rhythm and 8 arrhythmia categories.

Deep learning methods realize automatic feature extraction by stacking multiple layers of neural network layers with different functions, which have achieved many research results in the medical field, such as auxiliary diagnosis based on medical images [19], analysis of electronic medical records [20], pneumonia diagnosis [21], heart attack prediction [21] and autism prediction [22], etc.

With the improvement of computing power, the depth of the neural network is deeper, the extracted features are more abstract, and the feature expression ability is stronger. Among various neural networks, CNN [23] can effectively extract local features due to its local connectivity and weight sharing. Recurrent neural network (RNN) [24] is often used to process time-series data, and its variants include Long Short-Term Memory (LSTM) [25], Bidirectional Long Short-Term Memory (BiLSTM) [26], etc. In addition, Resnet [27] has become the most common framework in the field of deep learning because it overcomes the gradient vanishing and exploding problems in the deep learning process. Its advanced model SE-Resnet [28] realizes the attention mechanism of channel direction.

The DNN models with 5 different structures are established in this paper to realize automatic arrhythmia recognition based on the model proposed by our previous studies [30]. Meanwhile, we change the one-dimensional convolution layer to two-dimensional convolution layer and modify the parameters of the DNN models. The structure diagram of the 5 DNN models is shown in Fig. 3.

3 Experimentation Details and Evaluation

The hardware equipment needed in this experiment is a high-performance computer. The experiment is based on Keras 2.2.4 to complete the construction, training and testing of the DNN model. The specific experimental details are divided into 4 parts: data preprocessing, DNNs structure, training and testing, and evaluation.

3.1 Data Preprocessing

In order to reduce the computational complexity, the 12-lead ECG of the selected samples is down-sampled to 250Hz. Table. 1 shows the number of samples labeled as normal rhythm and 8 types of arrhythmia in 4 databases.

On the premise of ensuring that the information of the original 12-lead ECG signal is completely preserved, we slice the 12-lead ECG to unify the data length, and stack them into Blocks. Specifically, each sample was cut into 12 slices, where the slice length is 2048. The overlap length is determined by the sample signal length S. When the sample signal length $S \in (2048, 12 \times 2048]$, overlap length is Overlap $= \lfloor (12 \times 2048 - S)/11 \rfloor$, where $\lfloor x \rfloor$ is the largest integer less than x. When the sample signal length $S > 12 \times 2048$, Overlap $= 0$, we intercept $[0, 12 \times 2048)$ of this sample as the effective signal part, and discard the rest. Each ECG segment with a length of 2048 is used as a slice.

By stacking 12 slices, we give the Block with length 2048, width 12, and channel 12. We have 19883 samples in total, of which different samples are from different normal individuals or patients. Each sample contains "Block", age, male and corresponding labels. The samples are divided into 3 sets: training set, validation set, and test set. Then the samples of each category are allocated 77% to the training set, 8% to the validation set, and 15% to the test set. Note that the samples in the training set are randomly scrambled, and the samples in the test set do not participate in the training process. In addition, we input the age and gender of the samples as auxiliary features into the DNN models for training.

3.2 DNN model

The structures of the 5 DNN models we built are shown in Fig. 3, and their details are as follows:

(model-1) Resnet with global pooling. Model-1 consists of three parts: Resnet is used to extract local features of 12-lead ECG, global pooling is used to extract global features of 12-lead ECG, and auxiliary features (age and gender) are used to improve classification performance. Resnet realizes the identity mapping of features through unique shortcut connections to solve the gradient disappearance and explosion problems of DNN models. In this paper, each Resnet-Block is composed of two paths. A path is composed of BatchNorm layers, Relu layers, two-dimensional Convolution layers, and Dropout layer. Another path is the shortcut connection, which is used to standardize the dimension of input features and transfer the original features.

(model-2) Resnet. Model-2 is the most common model. The structure of model-2 is taken from part of model-1.

(model-3) Resnet+LSTM. LSTM is a variant of RNN, which solves the problem that RNN cannot deal with long-term dependence by introducing mechanisms (input gate, forget gate, and output gate) and cell states. LSTM preserves important features through a variety of gate functions to ensure that they are not lost during long-term propagation. Due to its unique characteristics, LSTM is suitable for modeling time-series data such as 12-lead ECG.

(model-4) Resnet+BiLSTM. BiLSTM is a variant based on LSTM, which is composed of the forward LSTM and backward LSTM. Its advantage is that it can capture the forward information and backward information of 12-lead ECG at the same time.

(model-5) SE-Resnet. SE implements the attention mechanism in the channel dimension. In this paper, the channel dimension represents the slice number in the Block. The SE can improve the performance of the Resnet model by increasing the weight of important slices and weakening the weight of secondary slices.

3.3 Training and testing

The data fed into the DNN model are Blocks and auxiliary features, respectively. Auxiliary features include gender and age, where males and females are represented as 1 and 0, respectively, and ages are integers. The labels for the samples use 0 to 8, where 0 represents the normal rhythm and 1 to 8 represents 8 different types of arrhythmias. We train each DNN model based on the 12-lead ECG with 6 different erasure combinations. After training, we get 30 models and test them through the test set.

The relevant hyper-parameters in the training process include the stochastic gradient descent (SGD) with momentum, in which the momentum is 0.2, each model is trained for 50 epochs, dropout rate is 0.4, batch-size is

32, and the learning rate is 0.1 in the first 30 epochs and 0.01 in the last 20 epochs. The number of units of LSTM and BiLSTM is 64. Convolutional layer parameters of the 5 DNN models are shown in Appendix C.

3.4 Evaluation

The classification performance of the DNN model can be comprehensively evaluated by *Precision*, *Recall*, F_1, Receiver Operating Characteristic Curve (ROC), and Area Under Curve (AUC). The *Precision* of the ith category is

$$Precision_i = \frac{TP_i}{TP_i + FP_i}, \tag{1}$$

where TP is the number of samples that are actually positive and predicted to be positive, and FP is the number of samples that are actually negative and predicted to be positive. Then the 9-class average *Precision* of the erasure combination for a DNN model is

$$Precision = \frac{1}{9}\sum_{i=1}^{9} Precision_i. \tag{2}$$

The *Recall* of the ith category is

$$Recall_i = \frac{TP_i}{TP_i + FN_i}, \tag{3}$$

where FN is the number of samples that are actually positive and predicted to be negative. Then the 9-class average *Recall* of the erasure combination for a DNN model is

$$Recall = \frac{1}{9}\sum_{i=1}^{9} Recall_i. \tag{4}$$

Based on *Precision* and *Recall*, F_1 of a DNN model can be obtained as

$$F_1 = \frac{1}{9}\sum_{i=1}^{9} \frac{2(Precision_i \times Recall_i)}{Precision_i + Recall_i}. \tag{5}$$

ROC analysis and AUC metrics have been widely used in medical decision-making systems [31]. For multi-class problems with imbalanced number of samples, the macro-average ROC [18] measures the performance of the model by plotting the average quantitative relationship between the sensitivity and specificity of multi-class. AUC is the area under the ROC curve. When the ROC curve is close to the upper left corner and the AUC is close to 1, the DNN model has better classification performance.

4 Result

The details of *Precision*, *Recall* and F_1 for 5 DNN models based on the 12-lead ECG with 6 different erasure combinations are shown in Table. 2. We found that the best results for each DNN model usually occurred in the case-1, especially *Recall* and F_1. This shows that when all 12-lead ECG data are used for the training of the DNN models, the recognition performance of the models is usually improved. It is worth noting that the best performances of some DNN models occur in other cases, for example, case-2 achieves the best results for $Precision = 0.797$ and $F_1 = 0.758$ ($Precision = 0.795$ and $F_1 = 0.756$) in model-1 (model-3), case-3 and case-5 achieve the best results for $Recall = 0.759$ in model-2, and case-3 achieves the best result for $Precision = 0.788$ in model-4. This indicates that the absence of some leads may improve the *Precision*, *Recall*, or F_1 of some DNN models.

The average *Precision*, *Recall* and F_1 based on the 12-lead ECG with 6 different erasure combinations are shown in Fig. 4(a). The statistical results show that the optimal average $F_1 = 0.761$ and $Recall = 0.747$ are obtained in the case-1. Counterintuitively, we find that the optimal average $Precision = 0.785$ is obtained in case-3. Meanwhile, the average *Recall* and F_1 of the case-3 are both suboptimal after that of the case-1. Then we give the ROC of each DNN model for 6 cases based on the test set in Fig. 4(b)-(f). The comparison found that case-3 achieved the highest AUC in model-3, and achieved the second highest AUC in most models. Taken together, this shows that the absence of the leads I, II and III has the least impact on the recognition performance of the DNN models.

Table 2 Arrhythmia recognition performance of DNN models.

Precision	Case-1	Case-2	Case-3	Case-4	Case-5	Case-6
Model-1	0.770	**0.797**	0.794	0.780	0.761	0.735
Model-2	**0.816**	0.781	0.772	0.747	0.767	0.780
Model-3	0.755	**0.795**	0.771	0.677	0.771	0.708
Model-4	0.774	0.731	**0.788**	0.769	0.774	0.741
Model-5	**0.802**	0.692	0.801	0.758	0.798	0.739
Average score	0.783	0.759	**0.785**	0.746	0.774	0.741
Recall	Case-1	Case-2	Case-3	Case-4	Case-5	Case-6
Model-1	**0.746**	0.732	0.709	0.700	0.717	0.719
Model-2	0.737	0.730	**0.759**	0.707	**0.759**	0.674
Model-3	**0.737**	0.730	0.759	0.707	0.759	0.674
Model-4	**0.753**	0.694	0.707	0.675	0.701	0.734
Model-5	**0.760**	0.689	0.725	0.715	0.744	0.743
Average score	**0.747**	0.715	0.725	0.687	0.724	0.715
F_1	Case-1	Case-2	Case-3	Case-4	Case-5	Case-6
Model-1	0.752	**0.758**	0.744	0.727	0.732	0.717
Model-2	**0.768**	0.750	0.763	0.719	0.761	0.707
Model-3	0.743	**0.756**	0.739	0.654	0.727	0.707
Model-4	**0.761**	0.710	0.739	0.714	0.731	0.734
Model-5	**0.779**	0.683	0.740	0.734	0.766	0.738
Average score	**0.761**	0.731	0.745	0.710	0.743	0.721

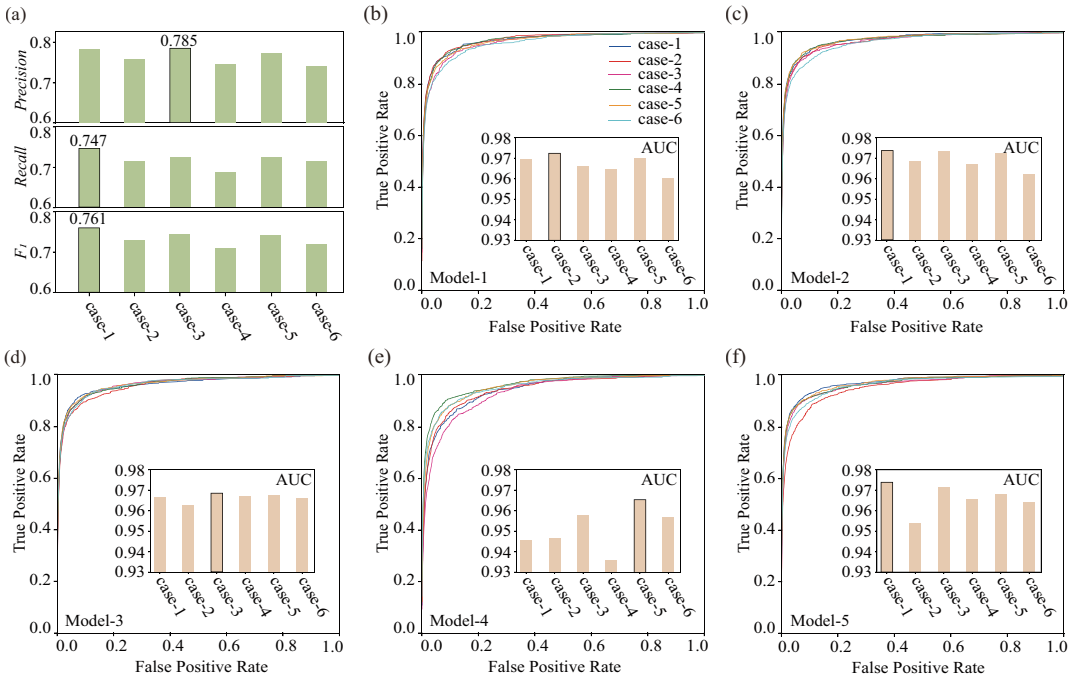

Fig. 4 (a) The average *Precision*, *Recall* and F_1 for 6 cases. (b)-(f) The ROCs and AUCs of 5 DNN models for 6 cases.

The values of the leads I, II and III can be derived from other leads, which are

$$\begin{cases} \mathrm{I} = \frac{2}{3}\mathrm{aVF} + \frac{4}{3}\mathrm{aVL}, \\ \mathrm{II} = \frac{4}{3}\mathrm{aVF} + \frac{2}{3}\mathrm{aVL}, \\ \mathrm{III} = \frac{2}{3}(\mathrm{aVF} - \mathrm{aVL}). \end{cases} \tag{6}$$

The substitutability of the leads I, II and III may account for their redundancy in automatic arrhythmia recognition of multiple DNN models. Therefore, compared with 12-lead ECG, it is reasonable to delete appropriate leads, such as the leads I, II and III, to reduce the computational complexity and apply it to portable devices for automatic arrhythmia recognition.

5 Discussion

With the development of social technology, more and more people begin to care about their physical condition. Various auxiliary detection equipment such as sports watches, heart rate detectors, and portable ECG detection bracelets have been developed and used. Common portable ECG detection devices include 6 limb leads detection devices and 2-lead detection devices. The disadvantage of these devices is that they cannot obtain complete information of the 12-lead ECG and therefore cannot obtain a comprehensive cardiac status.

We study the redundancy of lead through 12-lead ECG with different erasure combinations. The 12-lead ECG dataset used in this paper is a mixture of 4 open-access datasets from China, Germany, and the United States, which to some extent represent a collection of samples from all over the world with arrhythmias. Therefore, the trained model has good generalization performance. The 5 DNN models with different structures were constructed to observe the effect of the absence of different lead combinations on the recognition performance. It is found that the *Precision*, *Recall*, and F_1 obtained by the 5 DNN models are basically distributed in the interval $[0.7, 0.8]$, which is lower than our expected results. We believe that the main reason is the wide geographic coverage of the open-access datasets currently used, resulting in large differences in samples of the same arrhythmia type. The simulation results based on 5 DNN models found that the erasure of partial leads from 12-lead ECG reduces the recognition performance of DNN models. compared to using the complete 12-lead ECG as input, the average F_1 decreased by up to 0.016 when the 12-lead ECG with 5 different erasure combinations designed in this paper were used as inputs. But the reduction is very limited, especially when only the leads I, II and III are erased. This may be due to the fact that their values can be derived from other leads.

In our original assumption case-5 (lead II, aVL, aVR, and aVF erased) should have the lowest impact compared to other erasures. Because these four leads can be calculated from the rest of the leads. The actual results show that case-3 (leads I, II and III erased) is the best choice except for using 12-lead data. We believe that the reasons for this situation come from three aspects. First, the model may largely determine the classification performance. Second, for some specific arrhythmia types, the participation of specific leads may be required, such as ST types. The judgment always depends on all 12-lead, and finally based on this way of erasing lead information may make the model learn unreasonable features. The research of Alday et al. [33] shows that using deep learning for arrhythmia classification, the classification error is not limited to age, race, gender, and there may also be factors such as model structure and optimization function. In this study, the model we have chosen has minimized the holdover of the main body and only changed the structure of some of the layers. Obviously, the results show that the classification performance obtained by these slightly different models in the case of the same lead combination is not very different.

Deleting part of leads can effectively reduce measurement data and computational complexity. This enables the portable device to improve the performance of automatic arrhythmia recognition and quick feedback to the user, which is of great significance for timely treatment of arrhythmias. In conclusion, on the premise of basically maintaining the performance of automatic arrhythmia recognition, it is feasible to erase the appropriate lead information to reduce the computational complexity. Future work can be extended to study the impact of all possible lead combinations on automatic arrhythmia recognition, study the lead combinations required to recognize a specific arrhythmia, and improve the recognition performance of portable devices.

6 Acknowledgements

This work was supported by National Nature Science Foundation of China under Grant No. 61903208, Young doctorate Cooperation Fund Project of Qilu University of Technology (Shandong Academy of Sciences) under Grant No. 2019BSHZ0014, Introduce Innovative Teams of 2021 "New High School 20 Items" Project under Grant No.2021GXRC071, the Program for Youth Innovative Research Team in the University of Shandong Province in

China under Grant No. 2019KJN010, the Graduate Education and Teaching Reform Research Project of Qilu University of Technology in 2019 under Grant No. YJG19007.

7 Ethics statement and consent to participate

The database used in the study was an open access database. Therefore, no ethics statement and informed consent is required for this study. All methods in this study were carried out in accordance with relevant guidelines and regulations.

8 Competing interests

The authors declare no competing interests.

9 Data availability

The datasets used during the current study available in the PhysioNet/Computing in Cardiology Challenge 2020, `http://moody-challenge.physionet.org/2020`

Appendix A. Details of the 4 databases

Table 3 Details of the 4 databases.

Database	Total	Male	Female	Signal duration	Sample rate
CPSC2018	6877	3699	3178	6s-144s	500Hz
CPSC-Extra	3453	1843	1610	8s-98s	500Hz
PTB-XL	21837	11379	10458	10s	500Hz
G12EC	10344	5551	4793	10s	500Hz

Appendix B. The selection process of lead aVL

We debugged 7 DNN models in the top 25 models of the CPSC2018 challenge [32], including CPSC0223, CPSC0183, CPSC0235, CPSC0228, CPSC0204, CPSC0212, and CPSC0166. Copy the samples of the CPSC2018 database into 13 copies, one is the original data, and the other 12 copies delete one lead one by one. Because CPSC2018 challenge only discloses the test code of these teams, it does not include the training code. Therefore we used these model that these teams trained on the full 12-lead data. And these models are used to test the data of single lead erased one by one. The average F_1 of 7 DNN models trained based on 13 copies is shown in Fig. 5. The simulation results show that the average F_1 is the highest when the lead aVL is erased, leading us to speculate that lead aVL may be redundant in automatic arrhythmia recognition.

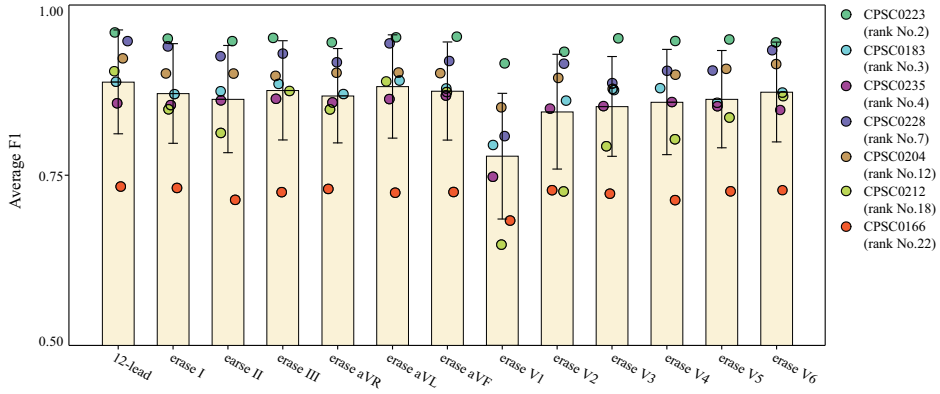

Fig. 5 The average F_1 of 7 DNN models trained based on 13 copies

Appendix C. Parameter description of 5 DNN models

Table 4 Description of convolutional layer parameters. Each "Conv2D" represents a 2D convolutional layer. All the Conv2D-i are located in Resnet-Block except the Conv2D-1.

Layer	Kernel size	Stride	Filter	Output size
Conv2D-1	12×12	1×1	12	$2048 \times 12 \times 12$
Conv2D-2	32×1	1×1	12	$2048 \times 12 \times 12$
Conv2D-3	32×1	1×1	12	$2048 \times 12 \times 12$
Conv2D-4	32×1	2×1	12	$1024 \times 12 \times 12$
Conv2D-5	32×1	1×1	12	$1024 \times 12 \times 12$
Conv2D-6	32×1	1×1	24	$1024 \times 12 \times 24$
Conv2D-7	32×1	1×1	24	$1024 \times 12 \times 24$
Conv2D-8	32×1	2×1	24	$512 \times 12 \times 24$
Conv2D-9	32×1	2×1	24	$256 \times 12 \times 24$
Conv2D-10	32×1	1×1	48	$256 \times 12 \times 48$
Conv2D-11	32×1	1×1	48	$256 \times 12 \times 48$
Conv2D-12	32×1	2×1	48	$128 \times 12 \times 48$
Conv2D-13	32×1	1×1	48	$128 \times 12 \times 48$
Conv2D-14	32×1	1×1	96	$128 \times 12 \times 96$
Conv2D-15	32×1	1×1	96	$128 \times 12 \times 96$
Conv2D-16	32×1	2×1	96	$64 \times 12 \times 96$
Conv2D-17	32×1	1×1	96	$64 \times 12 \times 96$
Conv2D-18	32×1	1×1	192	$64 \times 12 \times 192$
Conv2D-19	32×1	1×1	192	$64 \times 12 \times 192$
Conv2D-20	32×1	2×1	192	$32 \times 12 \times 192$
Conv2D-21	32×1	1×1	192	$32 \times 12 \times 192$

References

[1] Benjamin, Emelia J and Muntner, Paul and Alonso, Alvaro and Bittencourt, Marcio S and Callaway, Clifton W and Carson, April P and Chamberlain, Alanna M and Chang, Alexander R and Cheng, Susan and Das, Sandeep R. Heart disease and stroke statistics—2019 update: a report from the American Heart Association. *Circulation* **139(10)**, e56-e528. (2019).

[2] Nikolic, G. and Bishop, R. L. and Singh, J. B. Sudden death recorded during Holter monitoring. *Circulation* **66(1)**, 218-25. https://doi.org/10.1161/01.cir.66.1.218 (1982).

[3] Walsh III, Joseph A and Topol, Eric J and Steinhubl, Steven R. Novel wireless devices for cardiac monitoring. *Circulation* **130(7)**, 573-581. (2014).

[4] Esteva, A. and Robicquet, A. and Ramsundar, B. and Kuleshov, V. and DePristo, M. and Chou, K. and Cui, C. and Corrado, G. and Thrun, S. and Dean, J. A guide to deep learning in healthcare. *Nat Med* **25(1)**, 24-29. https://doi.org/10.1038/s41591-018-0316-z (2019).

[5] Lima, E. M. and Ribeiro, A. H. and Paixao, G. M. M. and Ribeiro, M. H. and Pinto-Filho, M. M. and Gomes, P. R. and Oliveira, D. M. and Sabino, E. C. and Duncan, B. B. and Giatti, L. and Barreto, S. M. and Meira, W. and Schon, T. B. and Ribeiro, A. L. P. Deep neural network-estimated electrocardiographic age as a mortality predictor. *Nature Communications* **12(1)**, 1-10. https://doi.org/10.1038/s41467-021-25351-7 (2021).

[6] Chiou, Y. A. and Syu, J. Y. and Wu, S. Y. and Lin, L. Y. and Yi, L. T. and Lin, T. T. and Lin, S. F. Electrocardiogram lead selection for intelligent screening of patients with systolic heart failure. *Scientific Reports* **11(1)**, 1948. https://doi.org/10.1038/s41598-021-81374-6 (2021).

[7] Attia, Zachi I and Kapa, Suraj and Lopez-Jimenez, Francisco and McKie, Paul M and Ladewig, Dorothy J and Satam, Gaurav and Pellikka, Patricia A and Enriquez-Sarano, Maurice and Noseworthy, Peter A and Munger, Thomas M. Screening for cardiac contractile dysfunction using an artificial intelligence–enabled electrocardiogram. *Nature medicine* **25(1)**, 70-74. (2019).

[8] Chen, T. M. and Huang, C. H. and Shih, E. S. C. and Hu, Y. F. and Hwang, M. J. Detection and Classification of Cardiac Arrhythmias by a Challenge-Best Deep Learning Neural Network Model. *iScience* **23(3)**, 100886. `https://doi.org/10.1016/j.isci.2020.100886` (2020).

[9] Cho, Y. and Kwon, J. M. and Kim, K. H. and Medina-Inojosa, J. R. and Jeon, K. H. and Cho, S. and Lee, S. Y. and Park, J. and Oh, B. H. Artificial intelligence algorithm for detecting myocardial infarction using six-lead electrocardiography. *Scientific Reports* **10(1)**, 1-10. `https://doi.org/10.1038/s41598-020-77599-6` (2020).

[10] Chen, Bin and Guo, Wei and Li, Bin and Teng, Rober KF and Dai, Mingjun and Luo, Jianping and Wang, Hui. A Study of Deep Feature Fusion based Methods for Classifying Multi-lead ECG. *arXiv preprint arXiv:1808.01721* (2018).

[11] Warrick, P. A. and Lostanlen, V. and Eickenberg, M. and Homsi, M. N. and Rodríguez, A. C. and Andén, J. Arrhythmia classification of reduced-lead electrocardiograms by scattering-recurrent networks. *In 2021 Computing in Cardiology (CinC)*, **48**, 1-4. (2021).

[12] Jessen, H. T. and van de Leur, R. R. and Doevendans, P. A. and van Es, R. Automated diagnosis of reduced-lead electrocardiograms using a shared classifier. *In 2021 Computing in Cardiology (CinC)*, **48**, 1-4. (2021).

[13] Bodini, M. and Rivolta, M. W. and Sassi, R. Classification of ECG signals with different lead systems using AutoML. *In 2021 Computing in Cardiology (CinC)*, **48**, 1-4. (2021).

[14] Osnabrugge, N. and Rustemeyer, F.and Kaparakis, C.and Battipaglia, F.and Bonizzi, P. and Karel, J. Multi-label classification on 12, 6, 4, 3 and 2 lead electrocardiography signals using convolutional recurrent neural networks. *In 2021 Computing in Cardiology (CinC)*, **48**, 1-4. (2021).

[15] Reyna, Matthew A and Sadr, Nadi and Alday, Erick A Perez and Gu, Annie and Shah, Amit J and Robichaux, Chad and Rad, Ali Bahrami and Elola, Andoni and Seyedi, Salman and Ansari, Sardar. Will two do? Varying dimensions in electrocardiography: the PhysioNet/Computing in Cardiology Challenge 2021. *In 2021 Computing in Cardiology (CinC)*, **48**, 1-4. (2021).

[16] Goldberger, A. L. and Amaral, L. A. and Glass, L. and Hausdorff, J. M. and Ivanov, P. C. and Mark, R. G. and Mietus, J. E. and Moody, G. B. and Peng, C. K. and Stanley, H. E. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation* **101(23)**, E215-20. `https://doi.org/10.1161/01.cir.101.23.e215` (2000).

[17] Liu, F. F. and Liu, C. Y. and Zhao, L. N. and Zhang, X. Y. and Wu, X. L. and Xu, X. Y. and Liu, Y. L. and Ma, C. Y. and Wei, S. S. and He, Z. Q. and Li, J. Q. and Kwee, E. N. Y. An Open Access Database for Evaluating the Algorithms of Electrocardiogram Rhythm and Morphology Abnormality Detection. *Journal of Medical Imaging and Health Informatics* **8(7)**, 1368-1373. `https://doi.org/10.1166/jmihi.2018.2442` (2018).

[18] IWagner, P. and Strodthoff, N. and Bousseljot, R. D. and Kreiseler, D. and Lunze, F. I. and Samek, W. and Schaeffter, T. PTB-XL, a large publicly available electrocardiography dataset. *Sci Data* **7(1)**, 154. `https://doi.org/10.1038/s41597-020-0495-6` (2020).

[19] Wang, G. T. and Li, W. Q. and Zuluaga, M. A. and Pratt, R. and Patel, P. A. and Aertsen, M. and Doel, T. and David, A. L. and Deprest, J. and Ourselin, S. and Vercauteren, T. Interactive Medical Image Segmentation Using Deep Learning With Image-Specific Fine Tuning. *Ieee Transactions on Medical Imaging* **37(7)**, 1562-1573. `https://doi.org/10.1109/Tmi.2018.2791721` (2018).

[20] Shickel, B. and Tighe, P. J. and Bihorac, A. and Rashidi, P. Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis. *IEEE J Biomed Health Inform* **22(5)**, 1589-1604. `https://doi.org/10.1109/JBHI.2017.2767063` (2018).

[21] Yang, H Mehta and Duan, T and Ding, D and Bagul, A and Langlotz, C and Shpanskaya, K. CheXNet: radiologist-level pneumonia detection on chest x-rays with deep learning. *arXiv preprint arXiv:1711.05225* (2017).

[22] Hazlett, H. C. and Gu, H. B. and Munsell, B. C. and Kim, S. H. and Styner, M. and Wolff, J. J. and Elison, J. T. and Swanson, M. R. and Zhu, H. T. and Otteron, K. N. B. and Collins, D. L. and Constantino, J. N. and Dager, S. R. and Estes, A. M. and Evans, A. C. and Fonov, V. S. and Gerig, G. and Kostopoulos, P. and McKinstry, R. C. and Pandey, J. and Paterson, S. and Pruett, J. R. and Schultz, R. T. and Shaw, D. W. and Zwaigenbaum, L. and Piven, J. and Network, IBIS. Early brain development in infants at high risk for autism spectrum disorder. *Nature* **542**, `https://doi.org/10.1038/nature21369` (2017).

[23] Albawi, Saad and Mohammed, Tareq Abed and Al-Zawi, Saad. Understanding of a convolutional neural network. *2017 international conference on engineering and technology (ICET)*, 1-6.

[24] Medsker, Larry R and Jain, LC. Recurrent neural networks. *Design and Applications* **5**, 64-67. (2001).

[25] Hochreiter, S. and Schmidhuber, J. Long short-term memory. *Neural Comput* **9(8)**, 1735-80. `https://doi.org/10.1162/neco.1997.9.8.1735` (1997).

[26] Zhou, Peng and Shi, Wei and Tian, Jun and Qi, Zhenyu and Li, Bingchen and Hao, Hongwei and Xu, Bo. Attention-based bidirectional long short-term memory networks for relation classification. *Proceedings of the 54th annual meeting of the association for computational linguistics (volume 2: Short papers)*, 207-212.

[27] He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian. Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770-778.

[28] Zhu, Zhaowei and Wang, Han and Zhao, Tingting and Guo, Yangming and Xu, Zhuoyang and Liu, Zhuo and Liu, Siqi and Lan, Xiang and Sun, Xingzhi and Feng, Mengling. Classification of Cardiac Abnormalities From ECG Signals Using SE-ResNet. *2020 Computing in Cardiology*, 1-4.

[29] Hannun, Awni Y and Rajpurkar, Pranav and Haghpanahi, Masoumeh and Tison, Geoffrey H and Bourn, Codie and Turakhia, Mintu P and Ng, Andrew Y. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature medicine* **25(1)**, 65-69. (2019).

[30] Li, J. and Pang, S. P. and Xu, F. and Ji, P. and Zhou, S. and Shu, M. Two-dimensional ECG-based cardiac arrhythmia classification using DSE-ResNet. *Scientific Reports,* **12(1)**, 1-13. (2022).

[31] Zweig, M. H. and Campbell, G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem* **39(4)**, 561-77.

[32] The 7th International Conference on Biomedical Engineering and Biotechnology. The China Physiological Signal Challenge 2018: Automatic identification of the rhythm/morphology abnormalities in 12-lead ECGs. *Website* `http://2018.icbeb.org/Challenge.html` (2018).

[33] Alday, E.A.P. and Rad, A.B. and Reyna, M.A. and Sadr, N. and Gu, A. and Li, Q. and Dumitru, M. and Xue, J. and Albert, D. and Sameni, R. and Clifford, G.D. Age, sex and race bias in automated arrhythmia detectors. *Journal of Electrocardiology* **74**, 5-9.(2022).