

## Prediction of cardiovascular diseases by integrating multi-modal features with machine learning methods

Pengpai Li, Yongmei Hu\*, Zhi-Ping Liu\*

Department of Biomedical Engineering, School of Control Science and Engineering, Shandong University, Jinan, Shandong, 250061, China



### ARTICLE INFO

**Keywords:**  
 Cardiovascular diseases  
 Deep learning  
 Genetic algorithm  
 Multi-modal  
 Electrocardiogram  
 Phonocardiogram

### ABSTRACT

Electrocardiogram (ECG) and phonocardiogram (PCG) play important roles in early prevention and diagnosis of cardiovascular diseases (CVDs). As the development of machine learning techniques, detection of CVDs by them from ECG and PCG has attracted much attention. However, current available methods are mostly based on single source data. It is desirable to develop efficient multi-modal machine learning methods to predict and diagnose CVDs. In this study, we propose a novel multi-modal method for predicting CVDs based both on ECG and PCG features. By building up conventional neural networks, we extract ECG and PCG deep-coding features respectively. The genetic algorithm is used to screen the combined features and obtain the best feature subset. Then we employ a support vector machine to implement classifications. Experimental results demonstrate the performance of our method is superior to those of single modal methods and alternatives. Our method reaches an AUC value of 0.936 when we use multi-modal features of ECG and PCG.

### 1. Introduction

Cardiovascular diseases (CVDs) are the kind of diseases with the highest morbidity and mortality worldwide. It is estimated that 17.9 million people died of CVDs in 2017, accounting for 31% of global deaths [1]. There exist many diagnosis tools in clinics, such as cardiac auscultation (could be recorded as phonocardiogram), electrocardiograph, ultrasonic cardiogram, CT, myocardial enzyme and angiography. Cardiovascular activities are extremely complex. There are many subtypes of CVDs, such as arrhythmia, mitral valve prolapse and coronary artery disease. At present, single tool cannot effectively and comprehensively reveal the entire CVDs. Therefore, the diagnosis of CVDs based on multi-modal data has attracted widespread attention.

Due to the low cost and convenience of measurement, Electrocardiograph (ECG) and phonocardiogram (PCG) are generally used to diagnose patients with suspected CVDs [2,3]. ECG records the changes in electrical activity of the heart during each cardiac cycle. PCG is a graph that converts the vibration of heart sound into time-course vibration waves. In clinics, it is essential to consider both PCG and ECG for physicians to make diagnosis decision. According to Fig. 1, it can be observed that the waveforms of ECG and PCG are closely connected in the corresponding basic heartbeat activity. Due to the different generation mechanism of these two signals, ECG and PCG can reveal the status

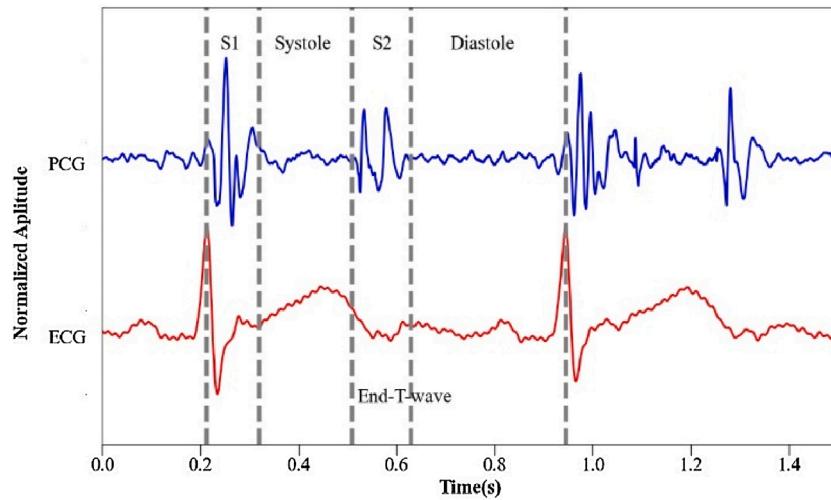
of cardiovascular activity from multiple perspectives.

In the past decades, automatic prediction of CVDs based on single ECG [4–8] or single PCG [9–11] has been widely studied. However, there are still few studies based on multi-modal methods. In recent years, some studies have begun to focus on the relationship between ECG and PCG. For example, Scholzel et al. [12] studied the possibility of ECG classification being applied to PCG data. They predicted normal and abnormal cardiovascular functions based on ECG and PCG respectively. Note that the ECG and PCG data are acquired simultaneously from those suspected patients. Although the performance based on ECG signal was not very promising (F-score of 0.739), it indicates the ECG signal carries useful information of cardiovascular function. However, the model that could integrate the two kinds of signal has not been proposed. Zarrabi et al. [13] predicted the risk of myocardial infarction using multiple data sources of PCG, ECG and clinical features. They extracted features from ECG and PCG in manual encoder ways and used an SVM classifier to make final decisions. It is demonstrated that the classifier with multi-modal features shows best performance. Li et al. [14] proposed a dual-input neural network for detecting coronary artery disease by integrating ECG and PCG. The results illustrate that the dual-input model has advantages over single-input models.

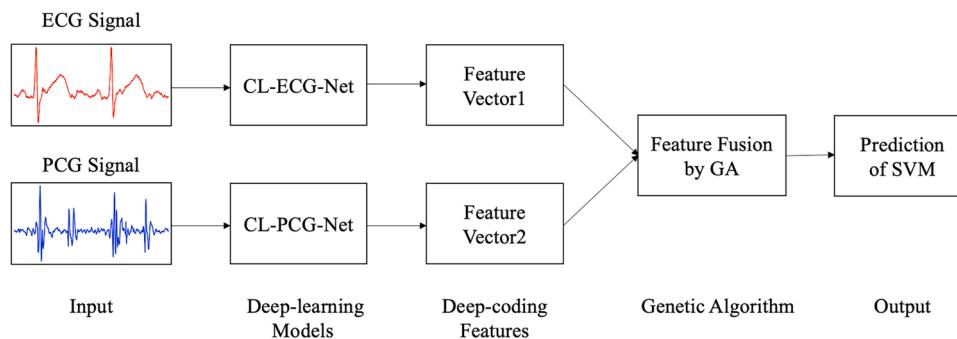
The references indicate the effectiveness of multi-modal features in the prediction of CVDs. However, feature selection steps are inevitable

\* Corresponding authors.

E-mail addresses: [huyongmei@sdu.edu.cn](mailto:huyongmei@sdu.edu.cn) (Y. Hu), [zpliu@sdu.edu.cn](mailto:zpliu@sdu.edu.cn) (Z.-P. Liu).



**Fig. 1.** The essential waveforms of ECG and PCG. The segmented regions with gray-dotted lines are the four basic states of PCG signal, i.e. S1, systole, S2, diastole, corresponding to the basic peaks of ECG signal.



**Fig. 2.** The framework of multi-modal model for predicting CVDs.

when dealing with high-dimensional features in the multi-modal problem [15,16]. Therefore, we need efficient method to solve the problem of dimension reduction. In this work, we aim to design a new model which could nicely integrate ECG and PCG to accurately predict CVDs. It is of great significance for the diagnosis and treatment of CVDs.

The main contributions of this paper are listed as follows:

- 1) A novel model based on both ECG and PCG signals is proposed to predict CVDs.
- 2) We design two networks for encoding ECG signal and PCG signal respectively. Then the boxplots and correlations for the deep-coding features of ECG and PCG are calculated and visualized.
- 3) In the process of feature dimension reduction, genetic algorithm is employed to obtain a better representation of data. Experiments on the benchmark dataset show that our model based on multi-modal features is superior to other methods.

## 2. Materials and methods

### 2.1. Framework

**Fig. 2** shows the framework of our proposed model of integrating ECG signal and PCG signal to predict CVDs. We design a deep learning network for each input channel. The output of the layer closest to the end of each network is extracted as the deep-coding features for the corresponding signal individually. Then we use a genetic algorithm embedded with SVM classifier to screen the combined features to get the optimal feature subset.

**Table 1**  
Data profile.

| Type     | Records | Time length (s) |      |       |        |       |
|----------|---------|-----------------|------|-------|--------|-------|
|          |         | Mean            | SD   | Min   | Median | Max   |
| Negative | 115     | 31.76           | 4.00 | 8.96  | 30.72  | 35.84 |
| Positive | 273     | 31.72           | 5.58 | 11.52 | 34.56  | 35.84 |

### 2.2. Datasets

The ECG and PCG recordings used in this study are sourced from PhysioNet/CinC Challenge 2016 [17]. The datasets have been contributed by several international institutions. According to different institutions, the dataset is divided into six subsets which are denoted as ‘training-a’ to ‘training-f’. The dataset of ‘training-a’ contains 409 recordings in which 405 ones record PCG and signal lead ECG using a Welch Allyn Meditron electronic stethoscope with a frequency response of 20 Hz – 20 kHz [18,19]. In the 405 recordings, 117 ones make up the normal control set, which are labeled as negative. The other 288 ones are collected from patients who are diagnosed with mitral valve prolapse (MVP), benign, aortic disease (AD) or other miscellaneous pathological conditions (MPC), which are labeled as positive. Both PCG signals and ECG signals are resampled to 2000 Hz.

Some recordings, or some segments of recordings, are visually noisy and hard to interpret. To avoid potential bias, 17 noise recordings are eliminated by manual cleaning as the previous studies [20]. **Table 1** lists the details of dataset.

The length of recordings varies greatly and a deep neural network

**Table 2**  
Data profile after expansion.

| Type     | Recordings | Time length (s) |
|----------|------------|-----------------|
| Negative | 1009       | 8               |
| Positive | 966        | 8               |

needs a relatively big amount of data. Thus, we expand the dataset based on a sliding window rule. Specifically, we segment the long raw signals into short recordings with a window size of 8 s. For achieving a balanced number of positive recordings and negative recordings, the stride of window used on positive recordings is 8 s while the stride of window used on negative recordings is 3 s. Finally, we change the ratio 2.4:1 (273:115) to about 1:1 after such segmentations. The expanded dataset is listed in Table 2.

It is noted that data expansion process is carried out after dividing the 388 recordings into training and validation datasets. This ensures the difference between the recordings in validation set and those in training set.

### 2.3. Single channel of feature extraction

#### 2.3.1. ECG feature extraction

The network used to encode ECG signals is shown in Fig. 3. It contains fully convolutional network and Long Short-Term Memory (LSTM)

network. Thus, it is named as CL-ECG-Net.

The original ECG signals are with the time length of 8 s and the frequency of 2000 Hz. We resample them to 2048 sample points as the input of CL-ECG-Net.

There are 8 convolutional layers and 8 pooling layers in the section of the fully convolutional neural network. The detailed parameters of convolutional and pooling layers are shown in Table 3. The boundary padding of each convolutional layer changed according to the size of its kernel, so that the length of signal remains constant after each

**Table 3**  
The detailed parameters of convolutional neural network in CL-ECG-Net.

| Index | Layer         | Index | Layer         |
|-------|---------------|-------|---------------|
| 1     | conv15_16     | 9     | conv7_64      |
| 2     | max-pooling_2 | 10    | max-pooling_2 |
| 3     | conv15_16     | 11    | conv7_64      |
| 4     | max-pooling_2 | 12    | max-pooling_2 |
| 5     | conv11_32     | 13    | conv3_128     |
| 6     | max-pooling_2 | 14    | max-pooling_2 |
| 7     | conv11_32     | 15    | conv3_128     |
| 8     | max-pooling_2 | 16    | max-pooling_2 |

Note: we denote the convolutional parameters as “conv (kernel size)\_(number of kernels)”. We denote the max-pooling parameters as “max-pooling\_(kernel size)”.

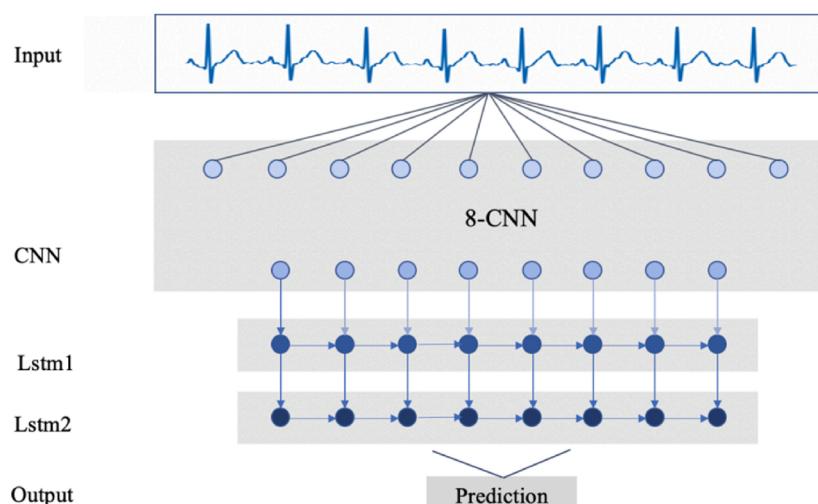


Fig. 3. The network architecture of CL-ECG-Net.

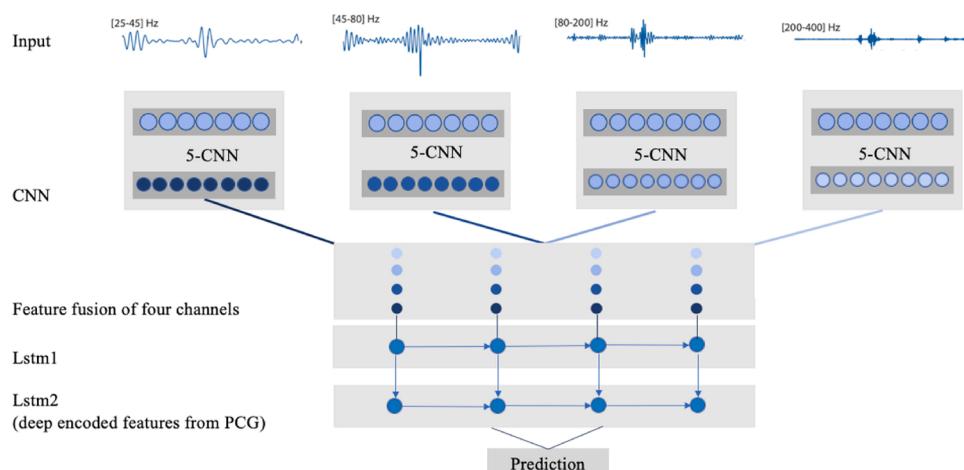


Fig. 4. The network architecture of CL-PCG-Net.

**Table 4**

The detailed parameters of convolutional neural network in CL-PCG-Net.

| Index | Layer         | Index | Layer         |
|-------|---------------|-------|---------------|
| 1     | conv5_16      | 6     | max-pooling_5 |
| 2     | max-pooling_5 | 7     | conv5_32      |
| 3     | conv5_16      | 8     | max-pooling_4 |
| 4     | max-pooling_5 | 9     | conv5_32      |
| 5     | conv5_16      | 10    | max-pooling_4 |

Note: we denote the convolutional parameters as “conv (kernel size)\_(number of kernels)” and we denote the max-pooling parameters as “max-pooling\_(kernel size)”.

convolutional layer. The kernel size of all pooling layers is 2, and the stride is also 2. Therefore, the length of the signal is reduced by 2 times after each pooling layer. Each convolutional layer is followed by a batch normalization layer and a rectified linear unit (ReLU). Batch normalization layer [21] is a trainable layer, which is used to renormalize the data distribution and help to reduce the difficulty of learning rate adjustment and speed up training. In a deep neural network, ReLU may help alleviate the problem of gradient vanishing [22].

Following the convolutional neural network, two LSTM layers are used in CL-ECG-Net. The core idea of LSTM is the cellular state. An LSTM layer can be regarded as a cell that replicates itself, forming a sequence of cells along the timeline. There are one-way connections between these cells, passing information down in the direction of time. The input size of LSTM cells in this paper depended on the number of feature maps generated by the last convolutional layer. Its value is 128. The output size of both LSTM cells is 64. The output of LSTM2 is extracted as the ECG deep-coding features.

### 2.3.2. PCG feature extraction

Compared to ECG signals, PCG signals relatively contain a wider frequency bandwidth. According to previous studies, if we train neural network models using raw PCG signals, it tends to produce unsatisfactory results. Therefore, it often pre-codes PCG signals by frequency domain analysis before applying neural networks [23,24]. First, the PCG signal is resampled to 1000 Hz. Then, according to [25], the frequency of the PCG signal is decomposed into four frequency bands: 25–45 Hz, 45–80 Hz, 80–200 Hz and 200–400 Hz.

The network used to encode PCG signal is shown in Fig. 4. It also contains fully convolutional network and LSTM network. Therefore, it is named as CL-PCG-Net. We set up four convolutional channels for CL-PCG-Net corresponding to the four frequency bands of PCG signals. The four convolutional channels have exactly the same structure. The detailed parameters of the four convolutional channels are shown in Table 4. The same as CL-ECG-Net, the feature length remains constant after each convolutional layer. The length of signal is controlled by pooling layers. After five pooling layers, the length of signal in each channel is reduced by 2000 times. Thus, after passing through the convolutional layers, we get a 4\*32 matrix in each channel.

In the multi-channel feature fusion layer, the feature graphs of the four channels are connected. After the fusion layer, a fused feature matrix of 4\*128 (128=4\*32) is obtained.

We also design two LSTM layers after the multi-channel feature fusion layer in CL-PCG-Net. The first LSTM layer has 128 units and returns the output of each cell unit. The second LSTM has 64 units and only returns the output of the last cell unit. The output of LSTM2 is extracted as the PCG deep-coding features.

### 2.4. Genetic algorithm

Genetic algorithm (GA) is an embedded feature searching approach that simulates the natural selection and genetic mechanism of Darwinian evolution. It has been reported that GA is able to obtain competitive results in a wide variety of real-world classification problems compared with other non-evolutionary, widely-used machine

learning paradigms [26–28]. In this paper, GA is employed to reduce the feature dimension and find the optimal feature subset.

The main procedure of GA is shown in Algorithm 1. First, a population composed of feature subset is randomly generated. Then we calculate the fitness of each feature subset. The feature subset with better fitness will be reserved, while those feature subsets with bad fitness will be rejected. After that, in the crossover process, we generate a new population by using the population reserved to generate new feature subset. Then, in the variation process, we choose some feature subsets in the new population, add and eliminate some features in these feature subsets. Finally, the above procedure will be terminated until the stopping criterion is met.

**Algorithm 1.** Basic procedure of genetic algorithm.

```

1    $t \leftarrow 0$ 
2    $P_t \leftarrow 0$  Generate a random population of feature subset
3   Calculate the fitness of feature subset in  $P_t$ 
4   Repeat
5      $P' \leftarrow$  Select promising feature subset in  $P_t$ 
6      $C \leftarrow$  Cross operator on  $P'$ 
7      $V \leftarrow$  Variation operator on  $C$ 
8     Calculate the fitness of feature subset in  $V$ 
9      $P_{t+1} \leftarrow 0$  Replace  $P_t$  with  $V$ 
10     $t \leftarrow t + 1$ 
11 Until Stopping criterion is met
12 The last  $P_t$  is treated as the optimal feature subset

```

### 2.5. Classification algorithm

The feature subsets selected by GA are fed into an SVM classifier [29]. We choose the linear kernel function ‘linear’ as the kernel function of SVM. Besides, at each evolutionary iteration of GA, the SVM classifier is embedded to train each feature subset. The AUC criteria of the trained SVM classifier is used to select promising feature subsets.

### 2.6. Evaluation criteria

To quantify the classification results, we employ five evaluation metrics, i.e., sensitivity (SN), specificity (SP), F-score, accuracy (ACC), and receiver operating characteristic (ROC) curve, for performance comparison among cases. They are defined as follows:

$$SN = \frac{TP}{TP + FN}$$

$$SP = \frac{TN}{TN + FP}$$

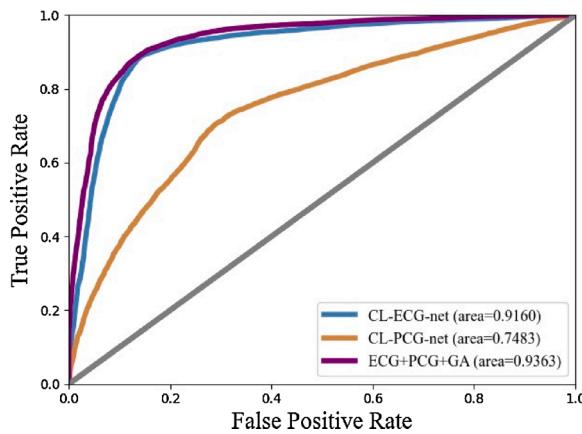
$$F\text{-score} = \frac{2 \times P \times SN}{P + SN}$$

$$ACC = \frac{TP + TN}{TP + FP + TN + FN}$$

where TP, FN, FP, TN and P represent the numbers of true positive, false negative, false positive, true negative, and precision respectively. ROC is a valuable tool to evaluate diagnostic tests and predictive models [30, 31]. The area under the ROC curve (AUC) represents the tradeoff between SN and SP.

## 3. Experiments

We apply the above methods to the datasets respectively. The trained CL-ECG-Net and CL-PCG-Net are also used to predict the CVDs to prove the advantage of our multi-modal method. Moreover, we use principal component analysis (PCA) [28] in the process of features fusion to prove the effectiveness of GA. Five-fold cross validations are used for



**Fig. 5.** ROC curve for each case introduced in experiments. GA indicates applying genetic algorithm to optimize the corresponding feature set.

validation in each case. Each case is repeated for ten times.

### 3.1. Training setting of network

The weights of convolutional layers and fully connected layers are initialized by He-normal algorithm [32]. And LSTM layers are initialized by Glorot-normal algorithm [33]. We use the optimizer of Adam [34] to update the network parameters. In the training process of CL-ECG-Net, there are a total of 150 epochs. The learning rate is initialized to 0.001 and multiplied by 0.1 every 50 epochs. In the training process of CL-PCG-Net, there are a total of 160 epochs. Its learning rate is initialized to 0.001 and multiplied by 0.1 every 80 epochs.

### 3.2. Parameters setting of genetic algorithm

The total number of feature subsets is initialized to be 400. The population of feature subsets is initialized as 400. In the process of crossover, 100 optimal feature sets (25% of the whole set) are reserved and used to generate 300 new feature sets. In the process of variation, 20% of the feature sets are involved, and about 20% of the features in each feature set involved are changed. The totally number of evolutional iterations is set to be 100.

## 4. Results and discussion

### 4.1. Classification performance

Fig. 5 shows the ROC curve of the five-fold cross validation results for each experiment case. And the detailed performances of different cases are shown in Table 5. We find the last case which combines the deep-coding features from ECG and PCG and GA obtains the best classification performance, namely SN of 0.903, SP of 0.845, F-score of 0.874, ACC of 0.873, and AUC of 0.936. The performance of multi-modal method outperforms the other cases with single model based on ECG or PCG. Compared with PCA, genetic algorithm is more suitable for feature fusion and dimension reduction dealing with these deep-coding features.

**Table 5**

The mean and standard variance (between parenthesis) criteria value of the 10 repeated experiments. LSTMs represent the output of LSTM layers in CL-ECG-Net and CL-PCG-Net. GA represents the genetic algorithm embedded with a classifier of SVM.

| Case        | SN                  | SP           | F-score             | ACC                 | AUC                 |
|-------------|---------------------|--------------|---------------------|---------------------|---------------------|
| CL-ECG-Net  | 0.898(0.014)        | 0.847(0.021) | 0.873(0.012)        | 0.872(0.012)        | 0.916(0.011)        |
| CL-PCG-Net  | 0.711(0.022)        | 0.707(0.020) | 0.705(0.013)        | 0.709(0.011)        | 0.748(0.014)        |
| LSTMs + PCA | 0.903(0.008)        | 0.837(0.018) | 0.871(0.016)        | 0.871(0.012)        | 0.918(0.013)        |
| LSTMs + GA  | <b>0.903(0.006)</b> | 0.845(0.018) | <b>0.874(0.010)</b> | <b>0.873(0.010)</b> | <b>0.936(0.011)</b> |

### 4.2. Analysis of deep-coding features

To show the generalization performance of CL-ECG-Net and CL-PCG-Net, we give the loss and accuracy *versus* iteration plots for autoencoder as shown in Fig. 6. The loss and accuracy curve in training dataset shows an adequate training of model, shown in Fig. 6(a) and (c). Moreover, the trend of accuracy curve in validation data is roughly the same as the trend of the accuracy curve in training, while the loss function curve in validation does not show a continuous downward trend. In our analysis, the fact reflects some severe noise in our dataset because the trend of loss and accuracy in the validation dataset are slightly opposite.

To assess the quality of features obtained from CL-ECG-Net and CL-PCG-Net, we calculated the Pearson's Correlation Coefficient (PCC) of the deep-coding features, which is showed in Fig. 7. The PCC is used to measure the correlations between deep-coding features and sample categories. The formula of PCC is given by:

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2) - E^2(X)} \sqrt{E(Y^2) - E^2(Y)}}$$

Fig. 7(a) shows the PCC between ECG deep-coding features and target values. The red dotted line 0.6 represents the boundary between moderate correlation and high correlation. The PCC of deep-coding features mostly locates between 0.6 and 0.8, indicating a significant relationship between them. Fig. 7(b) shows the PCC between PCG deep-coding features and target values. Compared with ECG features, the quality of PCG features is relatively poor. However, PCG still provides useful information for the final decision through feature fusion.

To show the distribution characteristics of these deep-coding features, we boxplot the positive and negative deep-coding features obtained from ECG and PCG signals as shown in Fig. 8. The blue boxes show values between first and third quartile, while the red crosses show values beyond the minimum and maximum valid value [35].

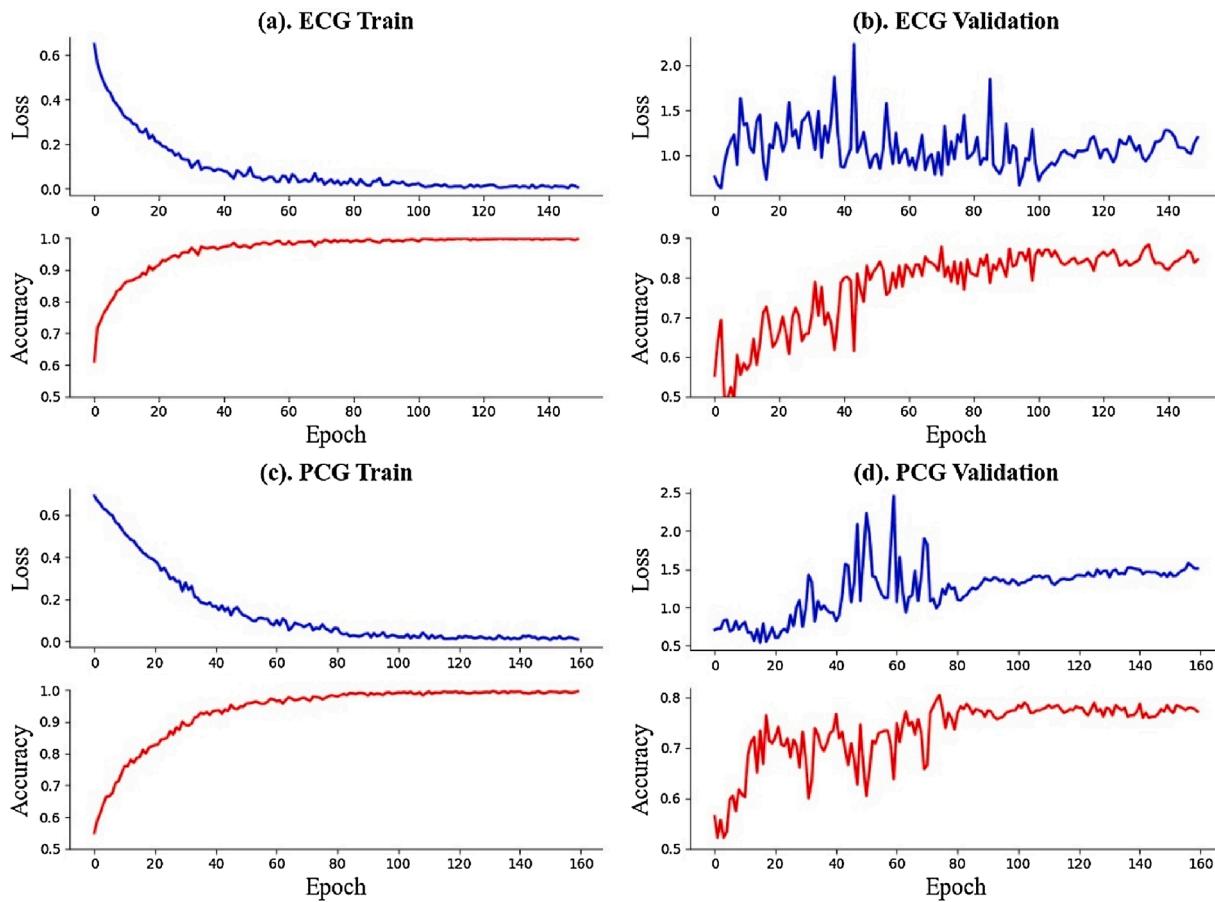
Each deep-coding feature in the positive class and negative class illustrates a large distribution difference. For example, the first deep coding feature of ECG in Fig. 8(a) and (b) has a positive class feature value of  $-0.95 \sim -0.75$  and a negative class feature value of  $0.25 \sim 0.40$ .

### 4.3. Genetic algorithm visualization

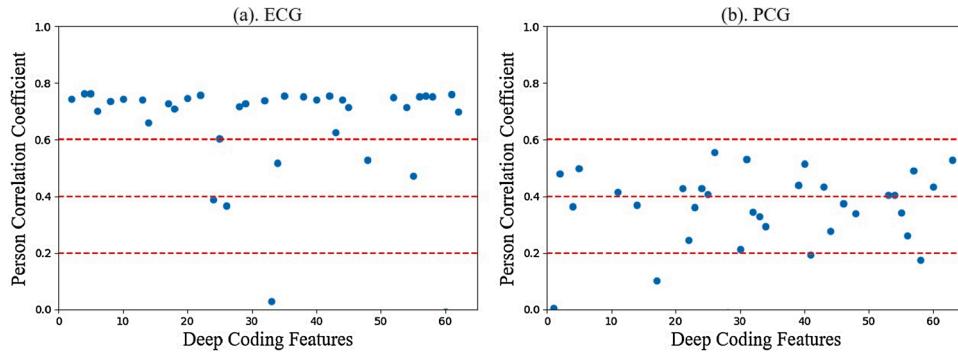
The genetic algorithm greatly improves the quality of the combined features obtained from ECG and PCG. Fig. 9 shows a schematic diagram of the population evolution. In Fig. 9(b), the X-axis represents time, that is, the time of population iterations. The Y-axis represents 400 feature subsets in the population. Feature subsets are ranked according to those fitness value. And  $y = 0$  represents the best feature subset after each evolution. The Z-axis represents the fitness value (AUC). To better highlight the changing fitness values after each evolution, the fitness values underwent a z-score normalization. As shown in Fig. 9(a), we extracted the curve at  $y = 0$ , which is the evolution curve of the optimal feature set. It can be seen that in this experiment, the genetic algorithm model improves the performance of the initial feature set from 0.890 to 0.934.

## 5. Conclusions

In this study, we proposed a multi-modal machine learning method to predict CVDs by integrating ECG and PCG. We extracted ECG deep-



**Fig. 6.** The loss and accuracy vs iteration plots for training CL-ECG-Net and CL-PCG-Net.



**Fig. 7.** The correlation between deep-coding features and labels. (a) ECG and (b) PCG.

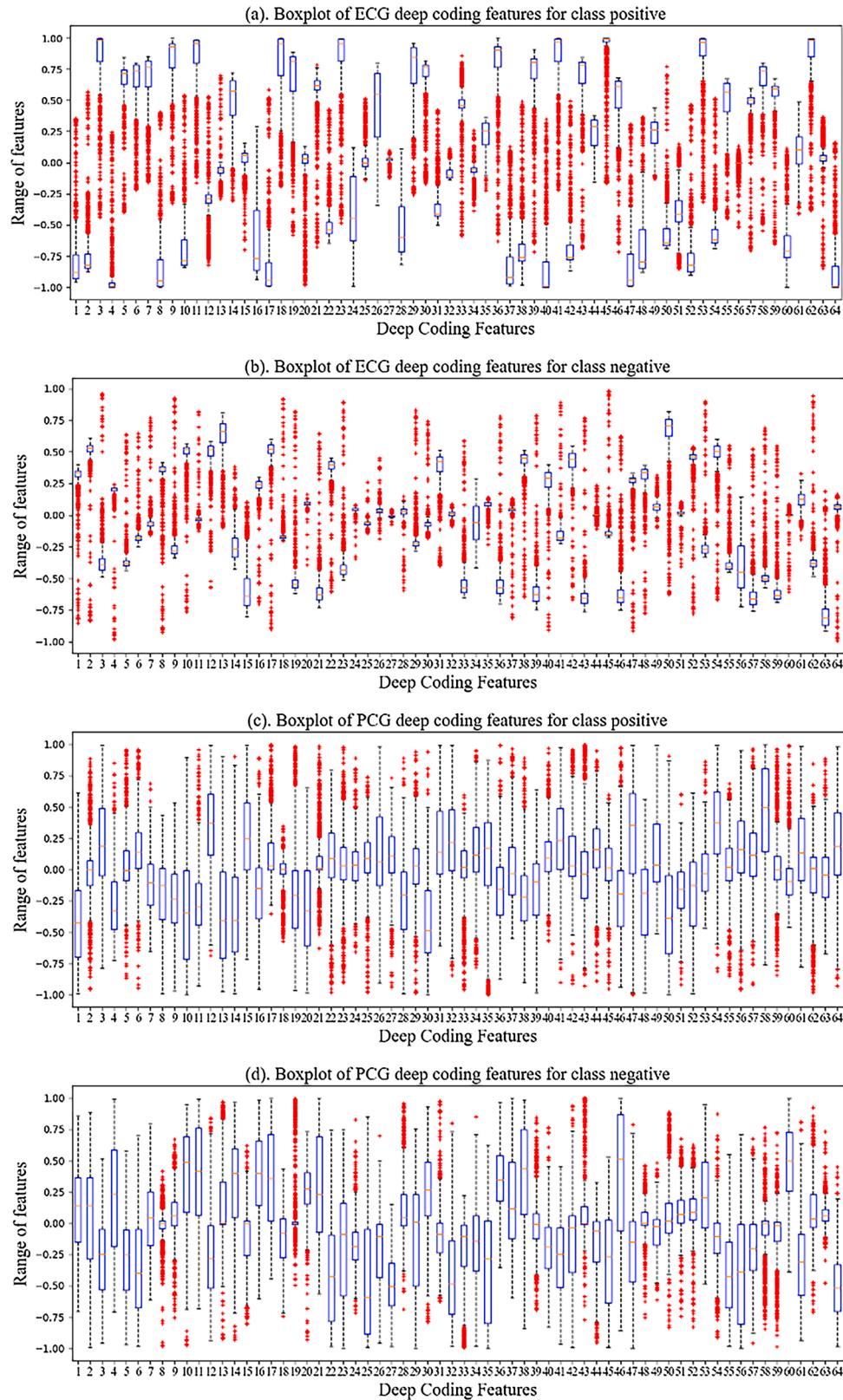
coding features from CL-ECG-Net and PCG ones from CL-PCG-Net respectively. Genetic algorithm is used to screen the fused features iteratively to obtain the optimal feature subset for classification. We employed a binary SVM classifier, which was trained with the optimal feature subset, to make the prediction. The classification performance demonstrates the effectiveness of our proposed strategy. The results also provide evidence that a multi-modal classification in predicting CVDs outperforms alternative single-modal methods.

Although our proposed method can naturally integrate ECG and PCG, the performances of the two networks individually are not satisfied, especially for CL-PCG-Net. Nevertheless, the networks adopted in this paper have been adjusted optimally for the limited datasets. In

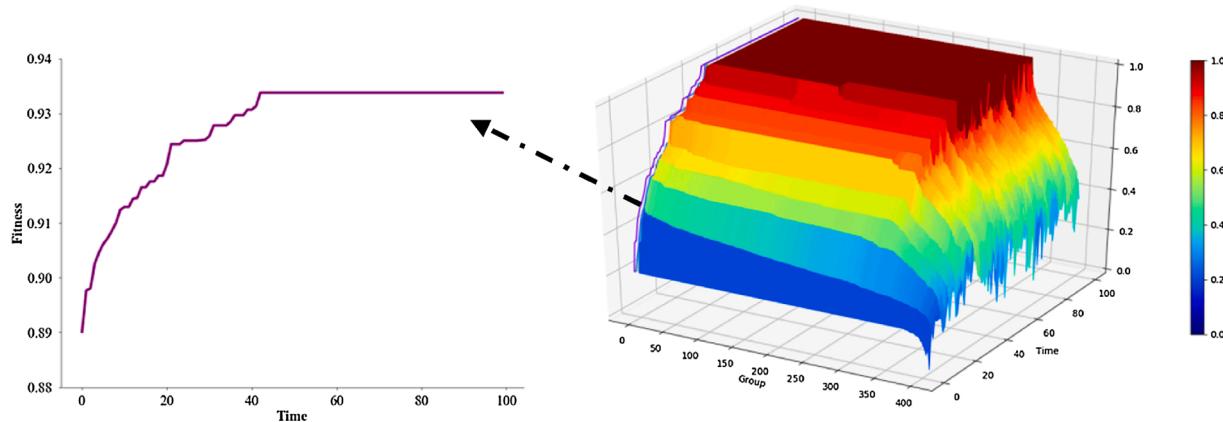
clinics, there exists rich and abundant CADs data. Our proposed method can provide a general framework to address the multi-modal problem. In future work, we will cooperate with clinicians to collect more valuable data and further optimize our model.

#### Code availability

The ECG and PCG originated from Physionet/CinC challenge 2016 can be downloaded at <http://doi.org/10.5281/zenodo.4263528>. All the source code of the proposed method is available at [https://github.com/Aolibaba/ecg\\_pcg\\_ga](https://github.com/Aolibaba/ecg_pcg_ga).



**Fig. 8.** Boxplots of deep-coding features obtained from ECG and PCG. (a) and (b) describe the features of CL-ECG-Net. (c) and (d) describe the features of CL-PCG-Net.



**Fig. 9.** Population evolution diagram during genetic algorithm iterations.

#### CRediT authorship contribution statement

**Pengpai Li:** Conceptualization, Methodology, Software, Data curation, Writing - original draft, Visualization, Investigation, Validation. **Yongmei Hu:** Supervision. **Zhi-Ping Liu:** Conceptualization, Methodology, Supervision, Writing - review & editing.

#### Acknowledgements

We thank the members of our lab at Shandong University for their assistance. This work was partially supported by National Natural Science Foundation of China (NSFC) (61973190, 61572287); Shandong Provincial Key Research and Development Program (Major Scientific and Technological Innovation Project), 2019JZZY010423; Natural Science Foundation of Shandong Province of China (ZR2020ZD25); the Innovation Method Fund of China (Ministry of Science and Technology of China), 2018IM020200; the Program of Qilu Young Scholars of Shandong University.

#### Declaration of Competing Interest

The authors report no declarations of interest.

#### References

- [1] World Health Organization, Cardiovascular Diseases (CVDs), 2017. <https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-cvds>.
- [2] P.M. Rautaharju, Eyewitness to history: landmarks in the development of computerized electrocardiography, *J. Electrocardiol.* 49 (2016) 1–6, <https://doi.org/10.1016/j.jelectrocard.2015.11.002>.
- [3] A. Leatham, *Auscultation of the Heart and Phonocardiography*, Churchill Livingstone, London, 1975.
- [4] Y. Özbay, G. Tezel, A new method for classification of ECG arrhythmias using neural network with adaptive activation function, *Digit. Signal Process.* 20 (2010) 1040–1049, <https://doi.org/10.1016/j.dsp.2009.10.016>.
- [5] U.B. Baloglu, M. Talo, O. Yildirim, R.S. Tan, U.R. Acharya, Classification of myocardial infarction with multi-lead ECG signals and deep CNN, *Pattern Recognit. Lett.* 122 (2019) 23–30, <https://doi.org/10.1016/j.patrec.2019.02.016>.
- [6] A.Y. Hannun, P. Rajpurkar, M. Haghpanahi, G.H. Tison, C. Bourn, M.P. Turakhia, A.Y. Ng, Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network, *Nat. Med.* 25 (2019) 65–69, <https://doi.org/10.1038/s41591-018-0268-3>.
- [7] Q. Yao, R. Wang, X. Fan, Multi-class arrhythmia detection from 12-lead varied-length ECG using Attention-based Time-Incremental Convolutional Neural Network, *Inf. Fusion* 53 (2020) 174–182, <https://doi.org/10.1016/j.inffus.2019.06.024>.
- [8] A. Ghaffari, N. Madani, Atrial fibrillation identification based on a deep transfer learning approach, *Biomed. Phys. Eng. Express* 5 (2019) 035015, <https://doi.org/10.1088/2057-1976/ab1104>.
- [9] G.D. Clifford, C. Liu, B. Moody, J. Millet, S. Schmidt, Q. Li, I. Silva, R.G. Mark, Recent advances in heart sound analysis, *Physiol. Meas.* 38 (2017) E10–E25, <https://doi.org/10.1088/1361-6579/aa7ec8>.
- [10] H. Tang, Z. Dai, Y. Jiang, T. Li, C. Liu, PCG classification using multidomain features and SVM classifier, *Biomed. Res. Int.* 2018 (2018) 1–14, <https://doi.org/10.1155/2018/4205027>.
- [11] F. Li, M. Liu, Y. Zhao, L. Kong, L. Dong, X. Liu, M. Hui, Feature extraction and classification of heart sound using 1D convolutional neural networks, *EURASIP J. Adv. Signal Process.* 2019 (2019) 59, <https://doi.org/10.1186/s13634-019-0651-3>.
- [12] C. Scholzel, A. Dominik, Can electrocardiogram classification be applied to phonocardiogram data? – An analysis using recurrent neural networks, *2016 Computing in Cardiology Conference* (2016), <https://doi.org/10.22489/CinC.2016.167-215>.
- [13] M. Zarrabi, H. Parsaei, R. Boostani, A. Zare, Z. Dorfeshan, K. Zarrabi, J. Kojuri, A system for accurately predicting the risk of myocardial infarction using PCG, ECG and clinical features, *Biomed. Eng.* 29 (2017) 1750023, <https://doi.org/10.4015/S1016237217500235>.
- [14] H. Li, X. Wang, C. Liu, Y. Wang, P. Li, H. Tang, L. Yao, H. Zhang, Dual-input neural network integrating feature extraction and deep learning for coronary artery disease detection using electrocardiogram and phonocardiogram, *IEEE Access* 7 (2019) 146457–146469, <https://doi.org/10.1109/ACCESS.2019.2943197>.
- [15] T. Tuncer, S. Dogan, F. Ertam, A. Subasi, A novel ensemble local graph structure based feature extraction network for EEG signal analysis, *Biomed. Signal Process. Control* 61 (2020) 102006, <https://doi.org/10.1016/j.bspc.2020.102006>.
- [16] T. Tuncer, S. Dogan, Novel dynamic center based binary and ternary pattern network using M4 pooling for real world voice recognition, *Appl. Acoust.* 156 (2019) 176–185, <https://doi.org/10.1016/j.apacoust.2019.06.029>.
- [17] C. Liu, D. Springer, Q. Li, B. Moody, R.A. Juan, F.J. Chorro, F. Castells, J.M. Roig, I. Silva, A.E.W. Johnson, Z. Syed, S.E. Schmidt, C.D. Papadaniil, L. Hadjileontiadis, H. Naseri, A. Moukadem, A. Dieterlen, C. Brandt, H. Tang, M. Samieeinasaab, M. R. Samieeinasaab, R. Sameni, R.G. Mark, G.D. Clifford, An open access database for the evaluation of heart sound algorithms, *Physiol. Meas.* 37 (2016) 2181–2213, <https://doi.org/10.1088/0967-3334/37/12/2181>.
- [18] Syed Hassan, MIT Automated Auscultation System, 2005.
- [19] Z. Syed, D. Leeds, D. Curtis, F. Nesta, R.A. Levine, J. Guttag, A framework for the analysis of acoustical cardiac signals, *IEEE Trans. Biomed. Eng.* 54 (2007) 651–662, <https://doi.org/10.1109/TBME.2006.889189>.
- [20] C. Liu, D. Springer, G.D. Clifford, Performance of an open-source heart sound segmentation algorithm on eight independent databases, *Physiol. Meas.* 38 (2017) 1730–1745, <https://doi.org/10.1088/1361-6579/aa6e9f>.
- [21] S. Ioffe, C. Szegedy, Batch normalization: accelerating deep network training by reducing internal covariate shift, *ArXiv* 1502 (03167) (2015) [Cs], <http://arxiv.org/abs/1502.03167>.
- [22] V. Nair, G.E. Hinton, Rectified linear units improve restricted boltzmann machines, *Proceedings of the 27th International Conference on Machine Learning (ICML-10)* (2010) 807–814.
- [23] B. Bozkurt, I. Germanakis, Y. Stylianou, A study of time-frequency features for CNN-based automatic heart sound classification for pathology detection, *Comput. Biol. Med.* 100 (2018) 132–143, <https://doi.org/10.1016/j.combiom.2018.06.026>.
- [24] J. Rubin, R. Abreu, A. Ganguli, S. Nelaturi, I. Matei, K. Sricharan, Classifying heart sound recordings using deep convolutional neural networks and mel-frequency cepstral coefficients, *2016 Computing in Cardiology Conference* (2016), <https://doi.org/10.22489/CinC.2016.236-175>.
- [25] C. Potes, S. Parvaneh, A. Rahman, B. Conroy, Ensemble of feature-based and deep learning-based classifiers for detection of abnormal heartsounds, *2016 Computing in Cardiology Conference* (2016), <https://doi.org/10.22489/CinC.2016.182-399>.
- [26] A. Oriols-Puig, J. Casillas, E. Bernadó-Mansilla, Genetic-based machine learning systems are competitive for pattern recognition, *Evol. Intel.* 1 (2008) 209–232, <https://doi.org/10.1007/s12065-008-0013-9>.
- [27] J. Yang, H. Xu, P. Jia, Effective search for genetic-based machine learning systems via estimation of distribution algorithms and embedded feature reduction techniques, *Neurocomputing* 113 (2013) 105–121, <https://doi.org/10.1016/j.neucom.2013.01.014>.
- [28] Yong Xia, L. Wen, S. Eberl, M. Fulham, D. Feng, Genetic algorithm-based PCA eigenvector selection and weighting for automated identification of dementia using FDG-PET imaging, in: *2008 30th Annual International Conference of the IEEE*

- Engineering in Medicine and Biology Society, IEEE, Vancouver, BC, 2008, pp. 4812–4815, <https://doi.org/10.1109/IEMBB.2008.4650290>.
- [29] C. Cortes, V. Vapnik, Support-vector networks, *Mach. Learn.* 20 (1995) 273–297, <https://doi.org/10.1007/BF00994018>.
- [30] Jin Huang, C.X. Ling, Using AUC and accuracy in evaluating learning algorithms, *IEEE Trans. Knowl. Data Eng.* 17 (2005) 299–310, <https://doi.org/10.1109/TKDE.2005.50>.
- [31] K.H. Zou, A.J. O’Malley, L. Mauri, Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models, *Circulation* 115 (2007) 654–657, <https://doi.org/10.1161/CIRCULATIONAHA.105.594929>.
- [32] K. He, X. Zhang, S. Ren, J. Sun, Delving deep into rectifiers: surpassing human-level performance on ImageNet classification, *ArXiv* 1502 (01852) (2015) [Cs], <http://arxiv.org/abs/1502.01852>.
- [33] X. Glorot, Y. Bengio, Understanding the difficulty of training deep feedforward neural networks, *J. Mach. Learn. Res.* 9 (2010) 249–256.
- [34] D.P. Kingma, J. Ba, Adam: a method for stochastic optimization, International Conference on Learning Representations (2014). <http://arxiv.org/abs/1412.6980>.
- [35] T. Tuncer, S. Dogan, A. Subasi, Surface EMG signal classification using ternary pattern and discrete wavelet transform based feature extraction for hand movement recognition, *Biomed. Signal Process. Control* 58 (2020) 101872, <https://doi.org/10.1016/j.bspc.2020.101872>.