Huang YW, Yang GP, Wang KK *et al.* Multi-scale deep cascade bi-forest for electrocardiogram biometric recognition. JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY 36(3): 617–632 May 2021. DOI 10.1007/s11390-021-1033-5

Multi-Scale Deep Cascade Bi-Forest for Electrocardiogram Biometric Recognition

Yu-Wen Huang[1,2], *Member*, *CCF*, Gong-Ping Yang[1,2,*], *Senior Member*, *CCF*, Kui-Kui Wang[1], Hai-Ying Liu[3] and Yi-Long Yin[1], *Senior Member*, *CCF*

[1] *School of Software, Shandong University, Jinan 250101, China*

[2] *School of Computer, Heze University, Heze 274015, China*

[3] *Department of Computer Engineering, Changji University, Changji 831100, China*

E-mail: ywhuang@mail.sdu.edu.cn; gpyang@sdu.edu.cn; 201820490@mail.sdu.edu.cn; liuhy@xju.edu.cn ylyin@sdu.edu.cn

Received September 29, 2020; accepted March 2, 2021.

Abstract Electrocardiogram (ECG) biometric recognition has emerged as a hot research topic in the past decade. Although some promising results have been reported, especially using sparse representation learning (SRL) and deep neural network, robust identification for small-scale data is still a challenge. To address this issue, we integrate SRL into a deep cascade model, and propose a multi-scale deep cascade bi-forest (MDCBF) model for ECG biometric recognition. We design the bi-forest based feature generator by fusing L1-norm sparsity and L2-norm collaborative representation to efficiently deal with noise. Then we propose a deep cascade framework, which includes multi-scale signal coding and deep cascade coding. In the former, we design an adaptive weighted pooling operation, which can fully explore the discriminative information of segments with low noise. In deep cascade coding, we propose level-wise class coding without backpropagation to mine more discriminative features. Extensive experiments are conducted on four small-scale ECG databases, and the results demonstrate that the proposed method performs competitively with state-of-the-art methods.

Keywords electrocardiogram (ECG) biometric recognition, small-scale data, deep cascade bi-forest, multi-scale division, sparse representation learning

1 Introduction

Many human traits, such as faces, fingerprints, irises, gaits, and voices, have been studied for the purpose of identity recognition. Over the past decade, electrocardiogram (ECG) biometric recognition has emerged as a popular research topic [1]. ECG biometric recognition has become popular, especially for continuous recognition systems, due to its several unique advantages [2]. 1) *Liveness Detection.* An ECG is recorded through sensors attached to the body; hence, it can only be captured from a living person. 2) *High Security.* An ECG is difficult to counterfeit or spoof, leading to the high security of ECG biometric recognition. 3) *Universality.* ECG signals can be acquired from all living individuals. 4) *Small Data Size.* An ECG is a one-dimensional, low-frequency signal that can be easily stored and processed.

ECG signals are acquired by the electrodes attached to the body surface, and they have inevitable contact noise. There are various types of noise from the signal-acquisition devices, such as from power-line interference, baseline wandering and electrode motion artifact. ECG signals are corrupted by different noise, which affects the recognition performance, and the signal preprocessing cannot eliminate all noise. Therefore, a chal-

Regular Paper

Special Section on Learning from Small Samples

This work was supported in part by the NSFC-Xinjiang Joint Fund under Grant No. U1903127 and in part by the Natural Science Foundation of Shandong Province of China under Grant No. ZR2020MF052.

*Corresponding Author

lenging problem for ECG biometric recognition is designing a robust and precise method.

Sparse representation learning (SRL) can efficiently handle noise, and many related methods have been proposed for ECG biometric recognition [3–6]. SRL-based coding methods belong to one-step models such that the latent discriminative information cannot be fully exploited [7].

The deep neural network (DNN) has shown good performance at ECG biometric recognition [8–12], and has seen increasing research. Although DNN has achieved better results in ECG biometric recognition, many factors limit its development. First, DNN has good recognition performance based on large-scale training data for massive weighted parameter tuning, but is not easy to train on small-scale data usually by means of data augmentation. Second, DNN requires powerful computational resources. Third, DNN has too many hyper-parameters that need intricate adjustment. Additionally, the interpretability and the theoretical analysis of DNN are still not completely clear.

The deep cascade model (DCM) has been proposed as an alternative to DNN [7, 13–15]. DCM can achieve competitive performance without backpropagation optimization, especially for small-scale training data. However, the existing DCM-based methods only focus on how to deal with higher-dimensional image data, without considering small-scale data with low intrinsic dimension.

To tackle the above-mentioned challenges and issues, we propose a multi-scale deep cascade bi-forest model for ECG biometric recognition, which can handle small-scale data with noise. We integrate the sparse representation coding and random forest into the deep-layered learning framework, and a specialized end-to-end improved DCM for one-dimensional data is developed. Specifically, the proposed framework includes multi-scale signal coding and deep cascade coding. For multi-scale signal coding, we design an adaptive weighted pooling operation, which can fully explore the discriminative information of segments with low noise. In deep cascade coding, we propose a level-wise class coding without backpropagation, so as to mine more discriminative features based on level-wise representation. To the best of our knowledge, this is the first work to utilize an improved DCM to deal with one-dimensional signals.

This paper makes the following contributions.

1) We integrate SRL into the deep cascade structure, and propose multi-scale deep cascade bi-forest for ECG biometric recognition, which can perform better even with only small-scale training data.

2) We design the bi-forest based feature generator (BFG), consisting of two forests with the respective inputs of sparse and collaborative representation. Sparse representation with the L1 norm can obtain high-level semantic information, collaborative representation with the L2 norm can capture the collaborative relationship, and the combined representation can efficiently deal with noise.

3) We propose an adaptive weighted pooling strategy to enhance the discriminability of local segments. This strategy can be regarded as an attention mechanism in a deep learning model.

The rest of this study is organized as follows. In Section 2, we briefly review related work. We introduce a multi-scale deep cascade model for ECG biometric recognition in Section 3. We report on experimental evaluation and analysis of our method in Section 4, followed by a brief conclusion and some suggestions for future work in Section 5.

2 Related Work

2.1 ECG Biometric Recognition

The main processes of ECG biometric recognition include preprocessing, feature extraction, dimensionality reduction, and decision. Current feature extraction methods include fiducial-based, non-fiducial based, and hybrid approaches. Fiducial-based approaches extract ECG fiducial markers between points, angles, areas, and amplitudes [16, 17]. Non-fiducial based approaches extract ECG features by the entire waveform morphology of the ECG signal or the isolated ECG heartbeat [18–21]. Hybrid approaches simultaneously fuse fiducial- and non-fiducial based features [22, 23].

After feature extraction, to obtain an appropriate low-dimensional subspace, dimensionality reduction methods are introduced to improve recognition performance. Wu *et al.* [24] proposed several ECG-based cancelable biometric schemes by signal subspace collapsing. He and Tan [25] presented a dimensionality reduction approach based on entropy principal component analysis (EPCA) for pattern recognition of ECG signals. Srivastva and Singh [26] proposed an ECG biometric feature extraction method utilizing a band-pass filter for quality checking and autocorrelation. Zheng *et al.* [27] proposed a feature extraction method based on an ECG superposition matrix of a single heartbeat ECG. Recently, kernel principal component analysis (KPCA)

for ECG biometric recognition has shown good performance in recognition rates and robustness [28]. Pal and Singh [29] presented an ECG biometric recognition system by KPCA, with recognition accuracy superior to that of principal component analysis (PCA). Hejazi *et al.* [30] presented a non-fiducial ECG authentication approach using kernel methods. Although ECG has long been deemed as high security by many researchers, some researches have recently showed that it is indeed possible to generate the adversarial ECG signals [31]. How to detect and reject fake ECG samples has becoming an important issue.

SRL can efficiently handle noise, and many SRL-based methods have been proposed for ECG biometric recognition. Wang *et al.* [3] proposed to extract compact and discriminative features from ECG signals for human identification based on sparse representation of local segments. Jaafar *et al.* [4] proposed a kernel sparse representation classifier to enhance system performance in a high-dimensional feature space for ECG biometric recognition. Li *et al.* [5] proposed a robust ECG biometric method based on graph regularized nonnegative matrix factorization and sparse representation. Goshvarpour and Goshvarpour [6] developed an identification system using a non-fiducial one-lead ECG feature set based on sparse representation.

Most recently, DNN-based methods have shown good performance at ECG biometric recognition, and have seen increasing research. Abdeldayem and Bourlai [8] investigated the spectral variation of the autocorrelation of the ECG segments to distinguish individuals, and two models of a 2D convolution neural network with different convolutional depths and fully-connected layer sizes were utilized as classifiers. Lobate *et al.* [9] presented a deep convolutional neural network for ECG biometric recognition. Luz *et al.* [10] used two convolutional neural network techniques to extract useful representation for ECG biometric recognition. Hammad *et al.* [11] developed two authentication systems with different levels of fusion algorithms using a convolution neural network. Zhang *et al.* [12] extracted distinctive features from an ECG segment without reference point detection via a deep convolutional neural network, thus avoiding the complicated signal fiducial characteristic point extraction process.

Although the existing ECG biometric recognition methods have reported some promising results, there are still some practical problems. 1) The amplitudes and intervals of fiducial points are sensitive to noise, and the recognition results of fiducial-based approaches are not always reliable. 2) The conventional feature representation methods belong to the one-step model, and the performance is degraded when ECG signals are contaminated by noise. 3) The existing ECG databases are small-scale data, while DNN-based methods for ECG biometric recognition are not easy to train on small-scale data. Therefore, a significant and challenging problem for ECG biometric recognition is how to design a better robust method that can handle small-scale data with noise. In this paper, we utilize the advantages of sparse representation and the deep cascade model, and propose a multi-scale deep cascade bi-forest for ECG biometric recognition, which can perform better even with only small-scale training data.

2.2 Deep Cascade Model

DCM has been proposed as an alternative to DNN, and many related methods have been reported. Zhou and Feng [13] first proposed a multi-grained cascade forest (gcForest), which generates a deep forest ensemble approach, with a cascade structure for representation learning. Zhang *et al.* [7] proposed an end-to-end deep cascade model based on sparse representation learning and nuclear-norm matrix regression with hierarchical learning, nonlinear transformation, and a multi-layer structure for corrupted face recognition. Liu *et al.* [14] designed a deep forest for spectral-based hyperspectral image classification. Wen *et al.* [15] proposed multi-level deep cascade trees, which can leverage deep cascade structures by stacking gradient boosting decision trees to effectively learn feature representation.

The deep cascade forest is the most widely-used deep cascade model; it is a deep learning method ensemble of decision trees [13]. Although the deep cascade forest has not been exploited in ECG biometric recognition, it has been used in other fields, where it has shown competitive performance. Liu *et al.* [14] designed a deep forest for spectral-based hyperspectral image classification, and then proposed an improved deep forest algorithm for spatial-based hyperspectral image classification. Utkin and Ryabinin [32] proposed a discriminative deep forest method by assigning weights to decision trees in a random forest. Su *et al.* [33] proposed a deep cascade forest model to classify anti-cancer drug response. Pang *et al.* [34] proposed a simple, effective approach to improve the efficiency of deep forest, which can pass instances with high confidence directly to the final stage rather than through all levels.

We use the deep cascade forest for ECG biometric recognition and present a model including a bi-forest

feature generator, an adaptive weighted pooling strategy, and two-stage coding, which is quite different from the above work with a deep cascade model.

3 Proposed Methodology

We first introduce the framework of our proposed multi-scale deep cascade bi-forest (MDCBF). The MDCBF framework has two parts: multi-scale signal coding (Subsection 3.1) and deep cascade coding (Subsection 3.2). We then discuss the overall recognition process (Subsection 3.3). A simple illustration of our framework is shown in Fig.1.

3.1 Multi-Scale Signal Coding

Multi-scale signal coding includes three parts: the multi-scale division, BFG and adaptive weighted pooling. First, we can obtain different local segments of a heartbeat using the multi-scale division. Then, we use BFG to generate the class feature codings of different local segments. At last, we obtain one global feature coding of different segments by the adaptive weighted pooling.

3.1.1 Multi-Scale Division

After signal preprocessing, ECG signals are isolated into heartbeats. Different local segments of a heartbeat waveform are influenced by the noise more or less. The local segments with low noise levels can provide more discrimination information for ECG biometric recognition. To explore the local discrimination information of heartbeat wave, we propose a multi-scale dividing operation to obtain different local segments, which can reduce the influence of some local segments with more noise. For example, in three-scale division, a heartbeat is divided into one, two and three local segments, respectively, as shown in Fig.2.

In Fig.2, the 1/1 scale is the original heartbeat. In the 1/2 and 1/3 scales, the original heartbeat is equally divided into two and three segments, respectively. To reduce the influence of some local segments with more noise, we separately generate the features of different local segments. We next design BFG, which consists of two forests with inputs of sparse and collaborative representations.

3.1.2 Bi-Forest Based Feature Generator

Let us assume that $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_C) \in \mathbb{R}^{m \times n}$ represents training samples, each class $\boldsymbol{X}_i$ contains n training samples, C is the total number of classes, m is the number of dimensions, $\boldsymbol{X}_i = (\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, ..., \boldsymbol{x}_{in_i}) \in \mathbb{R}^{m \times n_i}$, and $n = \sum_{i=1}^{C} n_i$. A test sample $\boldsymbol{y}$ can be reconstructed by a sparse linear combination of training samples $\boldsymbol{X}$ as:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{w}_p,$$

where $\boldsymbol{w}_p = (0, ..., 0, \boldsymbol{w}_{i1}, \boldsymbol{w}_{i2}, ..., \boldsymbol{w}_{in_i}, 0, ..., 0)^{\mathrm{T}} \in \mathbb{R}^n$ is a sparse coefficient vector with only nonzero elements associated with the i-th class. It should be noted that the advantages of representing the test sample as a linear combination of training samples have been explored in [35–37].

According to sparse representation learning [37], we can obtain the sparse representation coefficient of sample $\boldsymbol{y}$ by solving the following optimization problem:

$$\min_{\boldsymbol{w}_p} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}_p\|_2^2 + \lambda_1 \|\boldsymbol{w}_p\|_1, \tag{1}$$

where λ_1 is a regularization coefficient, and $\|\bullet\|_1$ is the L_1 norm.

After solving for the sparse coefficient $\boldsymbol{w}_p$, the sparse residual $\boldsymbol{r}_i^p$ of each class can be computed as follows:

$$\boldsymbol{r}_i^p = \|\boldsymbol{y} - \boldsymbol{X}\delta_i(\boldsymbol{w}_p)\|_2^2,$$

where $\delta_i(\bullet)$ is a vector operator, which only sets nonzero elements corresponding to the i-th class, $1 \leqslant i \leqslant C$.

Then, the sparse residual-based class feature coding $\boldsymbol{s}^p$ is obtained as:

$$\boldsymbol{s}^p = \left(\frac{\mathrm{e}^{-\boldsymbol{r}_1^p}}{\sum_{i=1}^{C} \mathrm{e}^{-\boldsymbol{r}_i^p}}, \frac{\mathrm{e}^{-\boldsymbol{r}_2^p}}{\sum_{i=1}^{C} \mathrm{e}^{-\boldsymbol{r}_i^p}}, ..., \frac{\mathrm{e}^{-\boldsymbol{r}_C^p}}{\sum_{i=1}^{C} \mathrm{e}^{-\boldsymbol{r}_i^p}} \right). \tag{2}$$

If the testing sample $\boldsymbol{y}$ belongs to class i and $\boldsymbol{s}_i^p$ is obviously bigger than the other elements in class feature vector $\boldsymbol{s}^p$, then this shows class discrimination.

Collaborative representation learning replaces the L_1 norm with the L_2 norm in (1), and we can obtain the collaborative representation coefficients of test sample $\boldsymbol{y}$ by solving the following optimization problem:

$$\min_{\boldsymbol{w}_q} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}_q\|_2^2 + \lambda_2 \|\boldsymbol{w}_q\|_2. \tag{3}$$

Here $\mathbf{I}$ is the identity matrix.

The collaborative representation coefficient $\boldsymbol{w}_q$ is obtained by the direct derivative of (3) as:

$$\boldsymbol{w}_q = \left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} + \lambda_2 \boldsymbol{I}\right)^{-1} \boldsymbol{X}^{\mathrm{T}}\boldsymbol{y},$$

where $\boldsymbol{I}$ is the identity matrix.

The collaborative residuals $\boldsymbol{r}_i^q$ of each class can be computed as follows:

$$\boldsymbol{r}_i^q = \|\boldsymbol{y} - \boldsymbol{X}\delta_i(\boldsymbol{w}_q)\|_2^2.$$

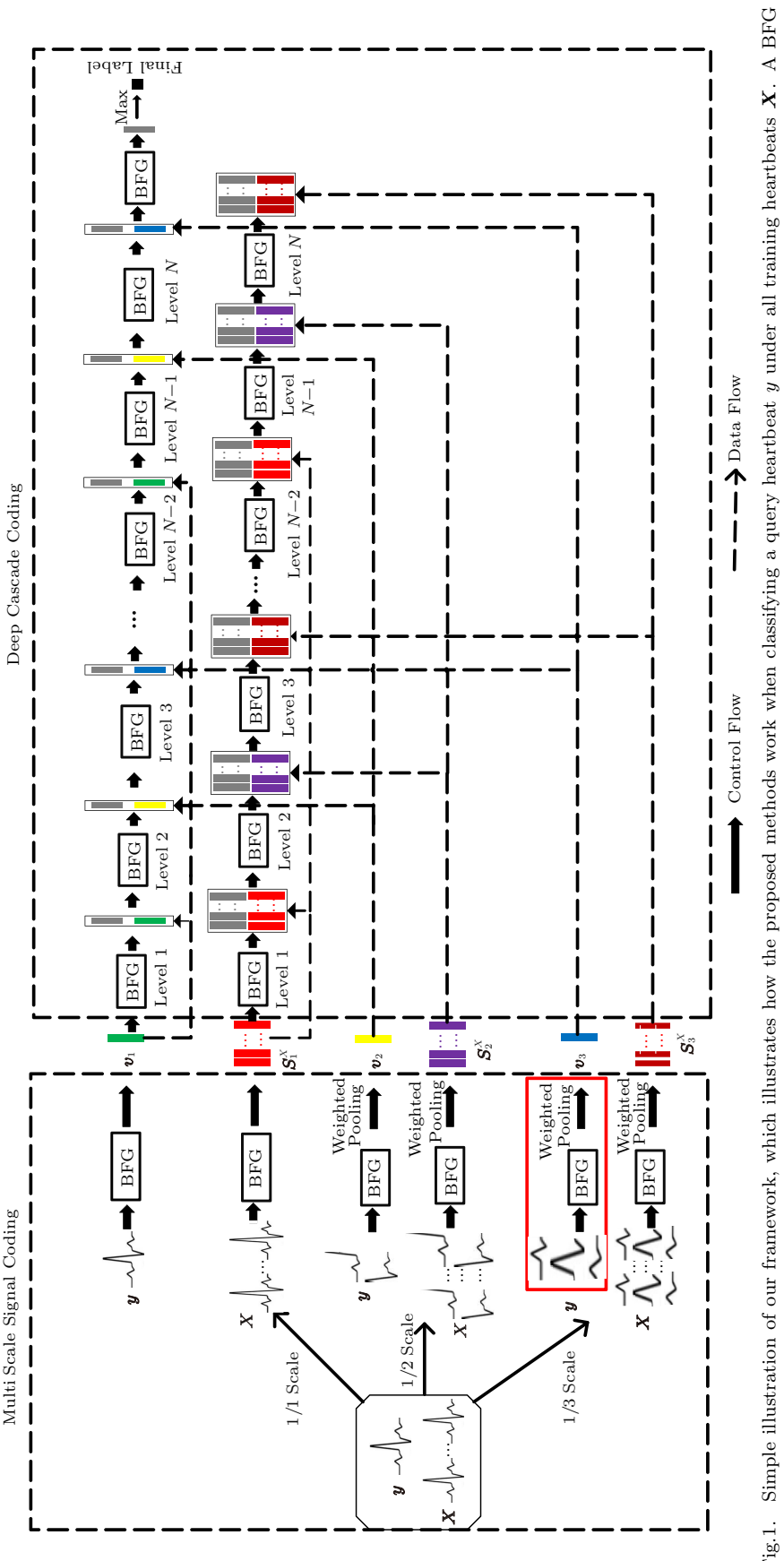


Fig.1. Simple illustration of our framework, which illustrates how the proposed methods work when classifying a query heartbeat y under all training heartbeats $\boldsymbol{X}$. A BFG computes the class vector. The detailed operation in the red-dashed box is shown in Fig.4.

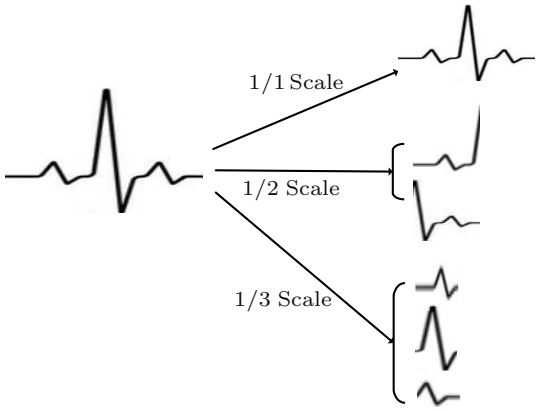


Fig.2. One heartbeat in three-scale division.

Then, the collaborative residual-based class feature vector $\boldsymbol{s}^q$ is obtained as follows:

$$\boldsymbol{s}^q = \left(\frac{\mathrm{e}^{-\boldsymbol{r}_1^q}}{\sum_{i=1}^{C} \mathrm{e}^{-\boldsymbol{r}_i^q}}, \frac{\mathrm{e}^{-\boldsymbol{r}_2^q}}{\sum_{i=1}^{C} \mathrm{e}^{-\boldsymbol{r}_i^q}}, ..., \frac{\mathrm{e}^{-\boldsymbol{r}_C^q}}{\sum_{i=1}^{C} \mathrm{e}^{-\boldsymbol{r}_i^q}} \right). \quad (4)$$

Since the decision forest can significantly improve the accuracy and generalization ability of classification, we design a bi-forest to generate class features, which takes the above sparse and collaborative residuals as the inputs of two forests. BFG is illustrated in Fig.3 and described in Algorithm 1.

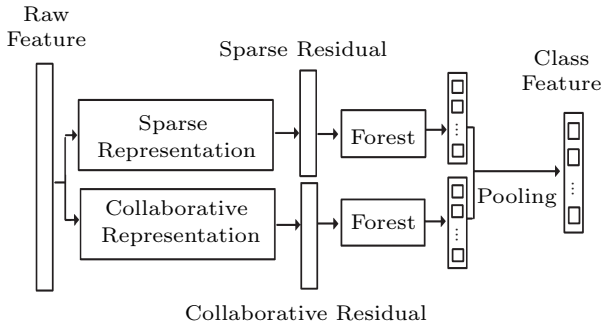


Fig.3. Illustration of BFG.

For convenience, we define the procedure of generating feature coding $\boldsymbol{v}$ using training samples $\boldsymbol{X}$ for a given query sample $\boldsymbol{y}$ as follows:

$$\boldsymbol{v} = BFG_{\boldsymbol{X}}(\boldsymbol{y}). \quad (5)$$

To encourage the diversity that is crucial for ensemble construction, the two forests of BFG are a random forest and a complete-random forest.

Algorithm 1. Bi-Forest Based Feature Generator (BFG)

Input: training samples $\boldsymbol{X} \in \mathbb{R}^{m \times n}$ and testing sample $\boldsymbol{y} \in \mathbb{R}^m$, class number C, regularization coefficients λ_1 and λ_2

Output: the class feature coding of testing sample $\boldsymbol{y}$

1: Obtain the sparse residual matrix $\boldsymbol{S}^p = (\boldsymbol{s}_1^p, \boldsymbol{s}_2^p, ..., \boldsymbol{s}_n^p) \in \mathbb{R}^{C \times n}$ of $\boldsymbol{X}$ using (2);
2: Obtain the collaborative residual matrix $\boldsymbol{S}^q = (\boldsymbol{s}_1^q, \boldsymbol{s}_2^q, ..., \boldsymbol{s}_n^q) \in \mathbb{R}^{C \times n}$ of $\boldsymbol{X}$ using (4);
3: Separately train two decision forests by $\boldsymbol{S}^p$ and $\boldsymbol{S}^q$;
4: Obtain the sparse residual vector $\boldsymbol{v}^p$ and collaborative residual vector $\boldsymbol{v}^q$ of testing sample $\boldsymbol{y}$ using (2) and (4), respectively;
5: $\boldsymbol{v}^p$ and $\boldsymbol{v}^q$ are classified by training two decision forests to generate class feature coding vectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$;
6: Transform $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ to $\boldsymbol{v}$ using the average pooling operator, and $\boldsymbol{v}$ is the class feature coding of testing sample $\boldsymbol{y}$.

3.1.3 *Adaptive Weighted Pooling Strategy*

After obtaining multiple local feature codes by BFG with coding for local segments, we need a pooling operation such that one original heartbeat can only be represented by one global feature code. For example, in Fig.4, for the 1/3 scale, the original heartbeat is equally divided into three local segments, and three local feature coding vectors ($\boldsymbol{s}_{31} \in \mathbb{R}^{C \times 1}$, $\boldsymbol{s}_{32} \in \mathbb{R}^{C \times 1}$, and $\boldsymbol{s}_{33} \in \mathbb{R}^{C \times 1}$) are generated by BFG. Here, we design an adaptive weighted pooling strategy to obtain the global feature coding vector $\boldsymbol{v}_3 \in \mathbb{R}^{C \times 1}$.

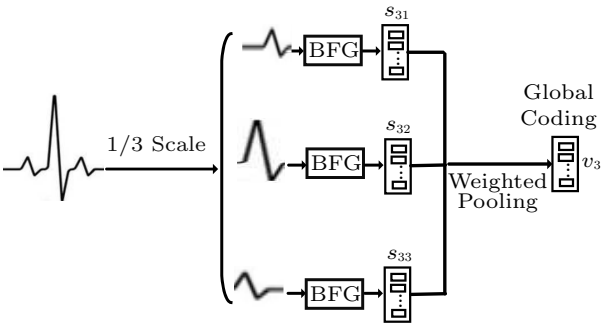


Fig.4. Generating global feature coding in the 1/3 scale.

Most pooling strategies only use the same local weight to build the global feature code, which cannot capture the discriminative information when some segments are easily affected by noise. For example, a heartbeat consists of a QRS segment and the other segments, among which the QRS segment has a larger amplitude

and is less affected by noise.

Therefore, when the global feature code is generated by the multiple coding of local segments, the segment with larger amplitude should have a larger weight, and segments with small amplitudes should have smaller weights. We define the weight w_{ij} of the j-th local segment in the $1/i$ scale as follows:

$$w_{ij} = \frac{A_{ij}}{\sum_{j=1}^{i} A_{ij}},$$

where A_{ij} is the average amplitude value of the sampling point, $1 \leqslant j \leqslant i$.

In the $1/i$ scale, the final feature code $\boldsymbol{v}_i \in \mathbb{R}^{C\times 1}$ of a sample is computed as follows:

$$\boldsymbol{v}_i = \sum_{j=1}^{i} w_{ij}\boldsymbol{s}_{ij},$$

where $\boldsymbol{s}_{ij} \in \mathbb{R}^{C\times 1}$ is the feature coding vector of the j-th local segment in the $1/i$ scale. Since the weight w_{ij} is computed automatically without training, $\boldsymbol{v}_i$ can be gained by an adaptive weighted pooling strategy.

For a query sample $\boldsymbol{y}$, for the 1/1 scale, we can only obtain one feature coding vector $\boldsymbol{v}_1^y \in \mathbb{R}^{C\times 1}$. For the 1/2 scale, we can obtain two feature coding vectors: $\boldsymbol{s}_{21} \in \mathbb{R}^{C\times 1}$ and $\boldsymbol{s}_{22} \in \mathbb{R}^{C\times 1}$, which can be transformed to one final feature coding vector $\boldsymbol{v}_2^y \in \mathbb{R}^{C\times 1}$:

$$\boldsymbol{v}_2^y = w_{21}\boldsymbol{s}_{21} + w_{22}\boldsymbol{s}_{22}.$$

Similarly, for the 1/3 scale, we can obtain three feature coding vectors, $\boldsymbol{s}_{31} \in \mathbb{R}^{C\times 1}$, $\boldsymbol{s}_{32} \in \mathbb{R}^{C\times 1}$ and $\boldsymbol{s}_{33} \in \mathbb{R}^{C\times 1}$, which can be transformed to one final feature coding vector $\boldsymbol{v}_3 \in \mathbb{R}^{C\times 1}$:

$$\boldsymbol{v}_3^y = w_{31}\boldsymbol{s}_{31} + w_{32}\boldsymbol{s}_{32} + w_{33}\boldsymbol{s}_{33}.$$

Fig.5 demonstrates the final feature coding vectors of a query heartbeat in three scales.

For each sample $\boldsymbol{x}_i$ of training set $\boldsymbol{X} \in \mathbb{R}^{m\times n}$, we can also obtain its feature coding vector $\boldsymbol{S}_{1i}^X \in \mathbb{R}^{C\times 1}$ in the 1/1 scale, $\boldsymbol{S}_{2i}^X \in \mathbb{R}^{C\times 1}$ in the 1/2 scale, and $\boldsymbol{S}_{3i}^X \in \mathbb{R}^{C\times 1}$ in the 1/3 scale, respectively. By putting each feature coding vector of each training heartbeat together, we will obtain three groups of feature coding vector sets $\boldsymbol{S}_1^X \in \mathbb{R}^{C\times n}$, $\boldsymbol{S}_2^X \in \mathbb{R}^{C\times n}$, and $\boldsymbol{S}_3^X \in \mathbb{R}^{C\times n}$. Note that for difficult tasks, we can try more multi-scale division if computational resources allow.

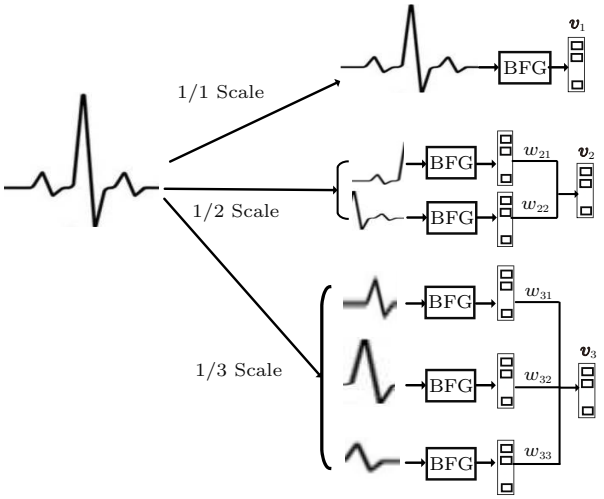


Fig.5. Generating feature coding vectors from different scales.

3.2 Deep Cascade Coding

Deep cascade coding is a level-by-level procedure, similar to a layer-by-layer DNN procedure. Each coding level consists of one BFG unit, as proposed in Subsection 3.1.2, which receives feature information containing two kinds of class vectors, one generated by its previous level and the other by multi-scale division. Each level of deep cascade learning outputs the result to the next level, as shown in Fig.6.

After the multi-scale division coding described above, three class coding vectors ($\boldsymbol{v}_1^y$, $\boldsymbol{v}_2^y$, and $\boldsymbol{v}_3^y$) of a query sample y and three class coding matrices ($\boldsymbol{S}_1^X$, $\boldsymbol{S}_2^X$ and $\boldsymbol{S}_3^X$) of the training set $\boldsymbol{X}$ are obtained. Then, the query coding vector $\boldsymbol{v}_1^y$ and the training coding matrix $\boldsymbol{S}_1^X$ are recognized as the input of the first level of the deep cascade structure.

First, $\boldsymbol{v}_1^y$ is fed into the BFG unit to generate the class vector $\boldsymbol{v}_1$, which is concatenated with $\boldsymbol{v}_1^y$ to construct the input query sample $\boldsymbol{d}_1$ of level 2, $\boldsymbol{v}_1 = BFG_{\boldsymbol{S}_1^X}(\boldsymbol{v}_1^y)$, $\boldsymbol{d}_1 = \left(\boldsymbol{v}_1^{\mathrm{T}}, \boldsymbol{v}_1^{y^{\mathrm{T}}}\right)$. Similarly, each column in $\boldsymbol{S}_1^X$ is fed into the BFG unit to generate the coding vector set $\boldsymbol{S}_1$, which is concatenated with $\boldsymbol{S}_1^X$ to construct the input training coding matrix $\boldsymbol{M}_1$ of level 2, $\boldsymbol{S}_1 = BFG_{\boldsymbol{S}_1^X}(\boldsymbol{S}_1^X)$, $\boldsymbol{M}_1 = \left(\boldsymbol{S}_1^{\mathrm{T}}, \boldsymbol{S}_1^{X^{\mathrm{T}}}\right)$.

Second, in the same way, the features $\boldsymbol{v}_2^y$ and $\boldsymbol{v}_3^y$ of query sample $\boldsymbol{y}$, augmented by the coding vectors generated by the previous level, are used as the query sample of the third- and fourth-level cascade structures, respectively. $\boldsymbol{S}_2^X$ and $\boldsymbol{S}_3^X$, augmented by the training coding matrices generated by the previous level, are used as the respective training sets to train the third- and fourth-level cascade structures, respectively. This

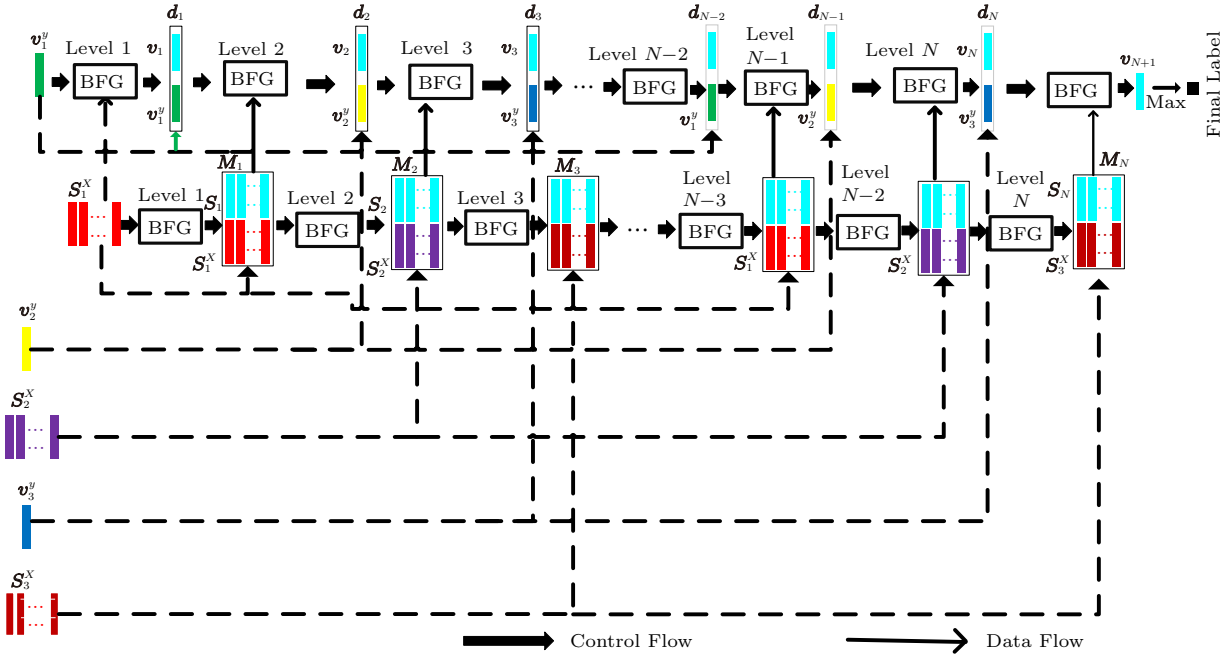


Fig.6. Deep cascade coding.

procedure is repeated, going forward through the cascade structure until level N where we obtain the final coding vector $\boldsymbol{d}_N \in \mathbb{R}^{2C\times 1}$ of query $\boldsymbol{y}$ and the final training coding matrix $\boldsymbol{M}_N \in \mathbb{R}^{2C\times n}$.

Finally, $\boldsymbol{d}_N \in \mathbb{R}^{2C\times 1}$ and $\boldsymbol{M}_N \in \mathbb{R}^{2C\times n}$ generated by level N are fed into the BFG to generate the final prediction coding vector $\boldsymbol{v}_{(N+1)} \in \mathbb{R}^{C\times 1}$, $\boldsymbol{v}_{(N+1)} = BFG_{\boldsymbol{M}_N}(\boldsymbol{d}_N)$.

3.3 Recognition

In the recognition of a query heartbeat $\boldsymbol{y}$, we can obtain its class label by the final prediction coding vector $\boldsymbol{v}_{(N+1)} \in \mathbb{R}^{C\times 1}$ as follows:

$$\text{label}(\boldsymbol{y}) = \arg\max_i \boldsymbol{v}_{(N+1)}. \tag{6}$$

Taking three-scale signal division as an example, the procedure of MDCBF for ECG biometric recognition is summarized in Algorithm 2.

4 Experiments

4.1 Databases

To comprehensively validate the effectiveness of our method, we conducted extensive experiments on four databases: MIT-BIH Arrhythmia Database (MITDB)[38], PTB Diagnostic ECG Database (PTBDB)[39], Check Your Biosignals Here initiative Database (CYBHiDB)[40], and University of Toronto Database (UofTDB)[41].

Algorithm 2. Recognition Procedure

Input: the ECG training heartbeat sets $\boldsymbol{X} \in \mathbb{R}^{m\times n}$, query heartbeat $\boldsymbol{y} \in \mathbb{R}^m$, class number C, regularization coefficients λ_1 and λ_2, and level number N

Output: the predicted label of the query sample $\boldsymbol{y}$.

1: Initialize: $i = 1$;
2: Obtain feature coding vectors $\boldsymbol{v}_1^y$, $\boldsymbol{v}_2^y$ and $\boldsymbol{v}_3^y$ of a query sample $\boldsymbol{y}$ and feature coding matrices $\boldsymbol{S}_1^X$, $\boldsymbol{S}_2^X$ and $\boldsymbol{S}_3^X$ by multi-scale signal coding;
3: $\boldsymbol{d}_0 = \boldsymbol{v}_1^y$, $\boldsymbol{M}_0 = \boldsymbol{S}_1^X$;
4: **while** $(i \leqslant N)$ **do**
5: Obtain $\boldsymbol{v}_i = BFG_{\boldsymbol{M}_{i-1}}(\boldsymbol{d}_{i-1})$ using (5);
6: Obtain $\boldsymbol{S}_i = (BFG_{\boldsymbol{M}_{i-1}}(\boldsymbol{M}_{i-1}^1), BFG_{\boldsymbol{M}_{i-1}}(\boldsymbol{M}_{i-1}^2), ..., BFG_{\boldsymbol{M}_{i-1}}(\boldsymbol{M}_{i-1}^n))$ using (5);
7: Update $\boldsymbol{d}_i = \left(\boldsymbol{v}_i^{\mathrm{T}}, \left(\boldsymbol{v}_{(i\%3)}^y\right)^{\mathrm{T}}\right)$, $\boldsymbol{M}_i = \left((\boldsymbol{S}_i)^{\mathrm{T}}, \left(\boldsymbol{S}_{(i\%3)}^X\right)^{\mathrm{T}}\right)$;
8: $i = i + 1$;
9: **end while**
10: Obtain feature coding vector $\boldsymbol{v}_{(N+1)} = BFG_{\boldsymbol{M}_N}(\boldsymbol{d}_N)$;
11: Find the index of maximum value in $\boldsymbol{v}_{(N+1)}$ using (6), which shows the label of $\boldsymbol{y}$.

MITDB is an arrhythmia database collected at

the chest, whose recordings have wide intra-class variety. The database contains 48 two-channel recordings from 47 subjects, and each recording is a 30-minute long excerpt from two-channel ambulatory ECG recordings [38]. We chose the 47 individuals with one recording for each subject.

PTBDB is usually used for medical diagnosis, and it includes 549 recordings from 290 subjects. This database has 1–5 recordings per subject, recorded from a 12-lead standard and three Frank leads, ranging between 38.4 seconds and 104.2 seconds [39]. We chose 248 subjects with ranges longer than 100 seconds, and every subject had one recording.

CYBHiDB was captured from palms and fingertips, and it contained short- and long-term databases. The short-term database was collected from 65 healthy subjects at intervals of two days. The long-term database was collected from 63 healthy subjects, each with two distinct sessions at a three-month interval [40]. The long-term database has more inter-class variations in heartbeat signals over time, and thus we chose it as the experimental data. The first and the second sessions in the long-term database are called $T1$ and $T2$, respectively.

UofTDB was collected from the thumbs of both hands of 1 020 subjects, and was specifically created for biometric recognition. The database includes up to six sessions over a period of six months in five postures: sitting, standing, exercising, supine and tripod. The first session includes 1 012 subjects and only 100 subjects were selected to participate in follow-up sessions over a period of six months [41]. We considered 46 of the 100 subjects who participated in all five sessions ($S1$, $S2$, $S3$, $S4$ and $S6$), in a sitting posture.

4.2 Preprocessing

ECG signal preprocessing included denoising, R peaks detection, segmentation and normalization. First, we used a fourth-order bandpass Butterworth filter to remove the baseline drift and higher frequency noise. Second, the R peaks were detected by the Pan-Tompkins algorithm [42]. Third, we used the R peaks to obtain heartbeat segmentation, and centered at R peaks with a certain length from each side of the peaks. Finally, we normalized all heartbeats to have a minimum value of 0.0 and a maximum value of 1.0.

4.3 Experimental Settings

For MITDB and PTBDB, the data were selected from one session, and we used 60% of the data for training, 30% for validation and 10% for testing. For CYBHiDB, we considered the first session as training and validation data, and the second as testing data. For UofTDB, we considered the $S1$ session as the training and validation data, and $S2$, $S3$, $S4$ and $S6$ as the testing data. For CYBHiDB and UofTDB, the ratio of training to validation data is $7:3$ for each session.

In all databases, each heartbeat was acquired by detecting the R peaks. It was centered at the R peaks with a segment length of 880 sampling points, with 220 sampling points before the peak and 660 sampling points after. In all experiments, the testing data were never seen by the classifier system in the training phase, and the average results were reported. All experiments were performed on a PC with an Intel i7-4790 3.60 GHz CPU and 16 GB RAM, and the programming environment was MATLAB 2016b.

4.4 Performance Metrics

To evaluate the performance of the proposed method, we conducted experiments on two modes of identification and authentication. In the identification mode, we used the heartbeat recognition rate as an evaluation criterion. This is the ratio of correctly identified testing heartbeats,

$$Heartbeat\ recognition\ rate = \frac{N\_correct\_beat}{N\_test\_beat},$$

where N_test_beat is the total number of test heartbeats, and $N\_correct\_beat$ is the number of correctly identified test heartbeats.

Like most of the work in the literature [1, 43, 44], we used the subject recognition rate as another evaluation criterion, which was computed by several continuous heartbeats of voters. For this criterion, each sample was represented by several continuous heartbeats, and the subject recognition rate was defined as follows:

$$Subject\ recognition\ rate = \frac{N\_correct\_sample}{N\_test\_sample},$$

where N_test_sample is the total number of test samples, and $N\_correct\_sample$ is the number of correctly identified testing samples.

In the authentication mode, we calculated a similarity measure with one heartbeat and all the other heartbeats from the same database. The equal error

rate (EER) is a system operating point where the false acceptance rate (FAR) and false rejection rate (FRR) are approximately equal by a given threshold.

$$FAR = \frac{NFA}{NIRA} \times 100\%,$$
$$FRR = \frac{NFR}{NGRA} \times 100\%, \quad (7)$$

where NFA is the number of false acceptances, $NIRA$ is the number of impostor recognition attempts, NFR is the number of false rejections, and $NGRA$ is the number of genuine recognition attempts.

4.5 Parameter Evaluation

We conducted experiments to evaluate the parameter sensitivity of our MDCBF model, including the number of cascade levels, the number of BFG units, and the number of decision trees in each forest. To obtain the optimal solutions of different parameters, we trained different MDCBF models by changing one parameter while fixing the others.

First, we evaluated the influence of the number of cascade levels. The heartbeat recognition rates with different numbers of cascade levels on four databases are shown in Fig.7.

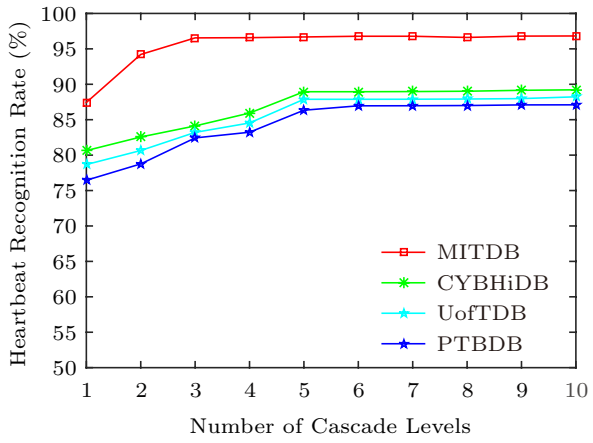


Fig.7. Heartbeat recognition rate with different numbers of cascade levels.

In Fig.7, we find that the heartbeat recognition rate increases consistently with the growth of the cascade levels. On MITDB, after the number of cascade levels reached 3, the heartbeat recognition rate increased slightly; on CYBHiDB and UofTDB, the heartbeat recognition rate increased slightly after the number of cascade levels reached 5; on PTBDB, the heartbeat recognition rate was stabilized when the number of cascade levels reached 6. Considering the above results, we set the number of cascade levels to 6 for all databases.

Then, we evaluated the influence of the number of decision trees of each forest from the basic classifier BFG, with results as shown in Fig.8.

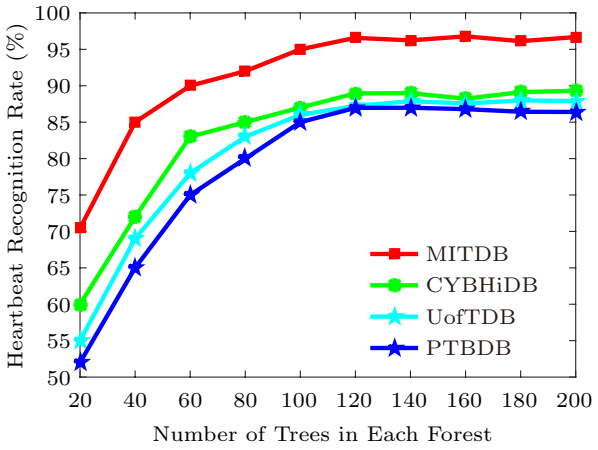


Fig.8. Heartbeat recognition rate with different numbers of decision trees.

As we could see in Fig.8, on all databases, the heartbeat recognition rates increased with the number of decision trees when it was under 120. The heartbeat recognition rates changed little when the number of decision trees was from 120 to 200, thereby we set the number of decision trees to 120 on the four databases.

Next, we evaluated the performance influence of parameters of sparse and collaborative representations. To generate new class vectors using the basic classifier BFG, we needed to set the regularization coefficient and number of iterations for sparse and collaborative representations. As suggested in [45], the regularization parameter λ_1 of sparse representation was set to $\frac{1.2}{\sqrt{m}}$, where m was the dimension of the data feature. The regularization parameter λ_2 of collaborative representation varied from 0.01 to 2 with an interval of 0.01. When $\lambda_2 = 0.06$, the heartbeat recognition rates were all better on the four databases. The heartbeat recognition rate with different numbers of iterations of sparse representations is shown in Fig.9.

In Fig.9, we could observe that the performance begin to achieve the best when the number of iterations reaches 60.

Finally, we evaluated the subject recognition rate under different numbers of heartbeats on four databases, as shown in Table 1.

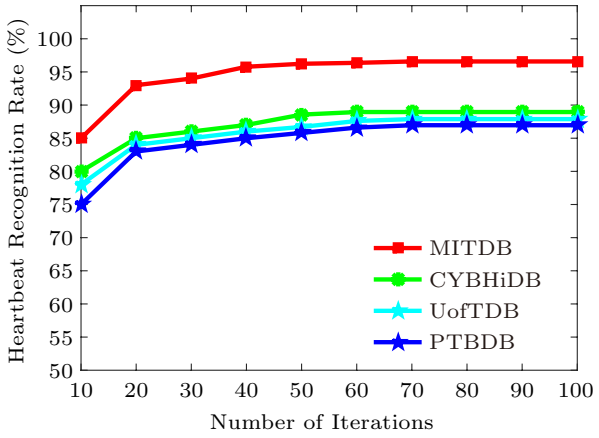


Fig.9. Heartbeat recognition rate with different numbers of iterations.

Table 1. Subject Recognition Rate Under Different Numbers of Heartbeats (n)

Database	$n=1$ (%)	$n=3$ (%)	$n=5$ (%)	$n=7$ (%)	$n=9$ (%)
MITDB	96.58	99.98	99.99	100.00	100.00
PTBDB	86.96	98.26	98.35	99.45	99.93
CYBHiDB	88.95	99.14	99.28	100.00	100.00
UofTDB	87.88	98.85	98.89	99.69	99.97

Table 1 shows that the subject recognition rate increases consistently with the number of heartbeats per testing sample on the four databases. When the number of heartbeats was 7 or more, our method achieved a subject recognition rate of 100% on MITDB and CYBHiDB. When the number of heartbeats varied from 1 to 3, the subject recognition rates increased quickly on all four databases. For example, the subject recognition rate increased by 10.19% on CYBHiDB. After the number of heartbeats reached 3, the subject recognition rate increased slowly and tended to be stable on four databases, and thus we used three heartbeats as a sample to perform subject recognition. Generally, three continuous heartbeats were acquired in about 2–4 seconds, which was acceptable for practical application.

4.6 Performance of Multi-Scale Signal Coding

We first verified the effectiveness of multi-scale signal coding. We designed a variant of MDCBF called DCBF, which contains only deep cascade bi-forest without multi-scale signal coding. We used six cascade levels, and each forest included 120 decision trees in both methods. The results on the four databases are shown in Table 2.

Table 2. Heartbeat and Subject Recognition Rates of MDCBF and DCBF

Database	Heartbeat Recognition Rate (%)		Subject Recognition Rate (%)	
	DCBF	MDCBF	DCBF	MDCBF
MITDB	90.23	96.58	98.34	99.98
PTBDB	80.23	86.96	97.32	98.26
CYBHiDB	82.46	88.95	98.08	99.14
UofTDB	81.37	87.88	97.94	98.85

In Table 2, we could see that the heartbeat and subject recognition rates of MDCBF are both higher than those of DCBF on the four databases, and it is evident that multi-scale signal coding could significantly improve the recognition performance.

To verify the effectiveness of adaptive weight pooling, we designed two variants of MDCBF: MDCBF-AP and MDCBF-MP, which use the average and max pooling operation, respectively. The heartbeat recognition rates are shown in Table 3.

Table 3. Heartbeat Recognition Rates of Different Weighted Pooling Operations

Method	MITDB (%)	PTBDB (%)	CYBHiDB (%)	UofTDB (%)
MDCBF-MP	96.25	86.12	88.23	87.32
MDCBF-AP	95.16	86.37	88.27	87.74
MDCBF	96.58	86.96	88.95	87.88

Table 3 shows that MDCBF using adaptive weighted pooling performs better than MDCBF-AP and MDCBF-MP on all four databases. From a deep learning perspective, our adaptive weighted pooling operation could be regarded as an attention mechanism usually used in convolutional neural networks and other deep models.

4.7 Performance with Different Basic Classifiers

BFG has two forests whose inputs are sparse and collaborative residuals, respectively. To verify the influence of different basic classifiers, we designed two variants of BFG: single decision forest using sparse residual (SDFS), and single decision forest using collaborative residual (SDFC). We also considered decision forest (DF) and sparse representation classifier (SRC) [37] as basic classifiers. In the following experiments, except for the basic coding units, the settings were all the same. These methods were fully tuned to achieve the best results by choosing the optimal parameters, and the results are shown in Table 4.

Table 4. Heartbeat Recognition Rate of MDCBF with Different Basic Classifiers

Basic Classifier	MITDB (%)	PTBDB (%)	CYBHiDB (%)	UofTDB (%)
SRC	85.61	77.04	78.65	77.48
DF	91.25	81.46	83.62	82.15
SDFC	95.17	85.47	86.78	85.83
SDFS	95.79	85.72	87.12	86.09
BFG	96.58	86.96	88.95	87.88

Table 4 demonstrates that BFG could improve the recognition performance. Typically, on MITDB, the recognition rate of our method using BFG is almost 10% higher than that of the method using SRC as the basic classifier. In addition, the method with the SRC base classifier achieves a worse result than that with the DF classifier.

4.8 Robustness to Noise

To validate the robustness of the proposed method, we added Gaussian noise to four databases. We compared the performance of the proposed method and the baseline methods of decision forest and sparse representation [37]. The performance with different noise levels is shown in Fig.10.

In Fig.10, it could be seen that the performance of our method is more stable with different noise levels on four databases. MDCBF could mine more discriminative features based on level-wise representation, and has more robust performance than the other methods.

4.9 Comparison with State-of-the-Art Methods

We compared the proposed method with state-of-the-art methods for ECG biometric recognition on the four databases, including non-deep-learning methods in [8, 29, 46–49], and deep-learning methods in [10–12, 50]. The subject recognition rate and EER of comparison with state-of-the-art methods are shown in Table 5.

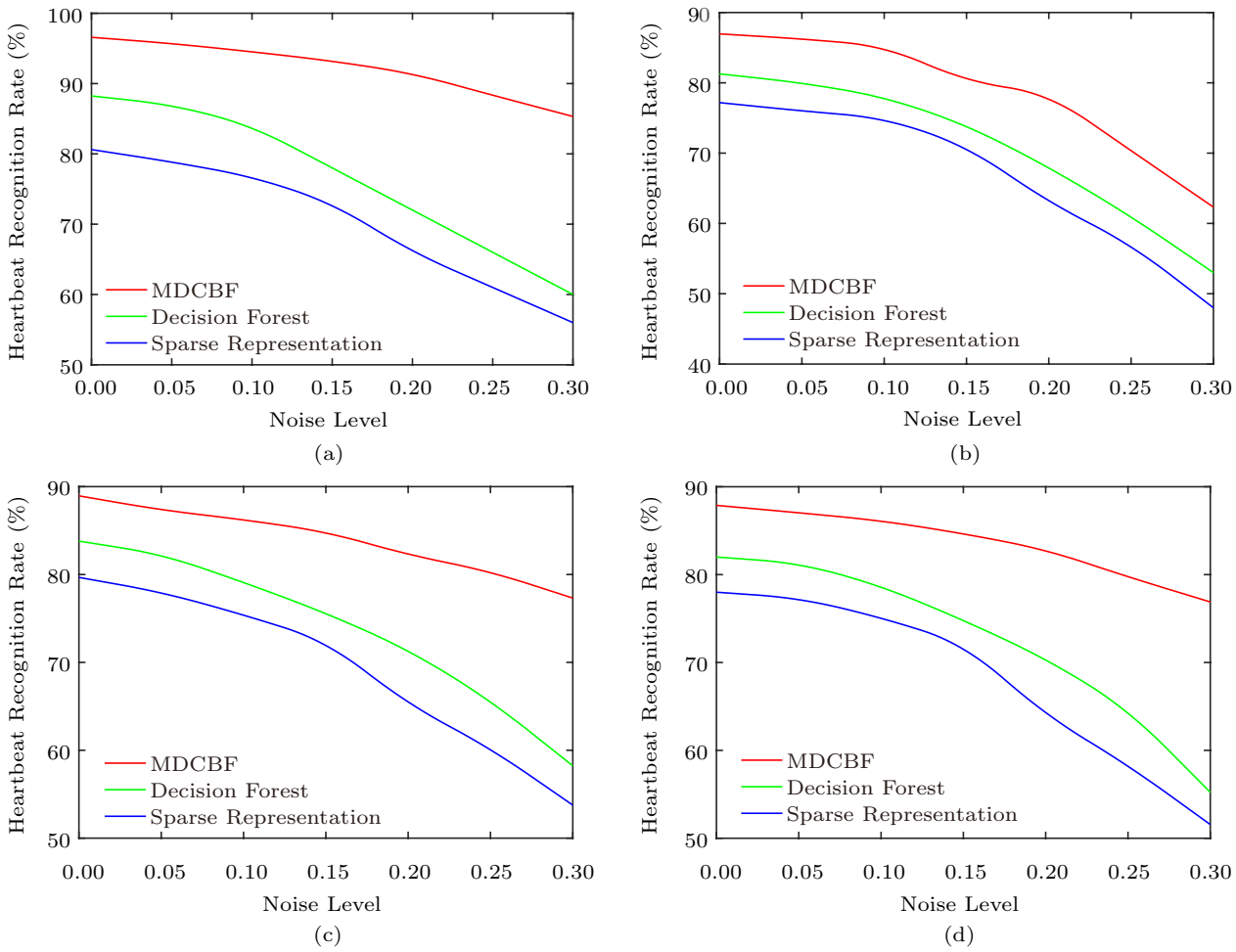


Fig.10. Heartbeat recognition rates with different noise levels on four databases. (a) MITDB. (b) PTBDB. (c) CYBHiDB. (d) UofTDB.

Table 5. Performance Comparison with the State of the Art

Dataset	Method	Subject Recognition Rate (%)	EER (%)
MITDB	[46]	94.68	2.73
	[47]	95.99	4.74
	[8]	96.50 [8]	–
	[50]	98.40	2.27
	Ours	99.98	0.34
PTBDB	[8]	94.90 [8]	–
	[29]	97.10	–
	[46]	98.19	2.55
	[12]	99.54	–
	[11]	–	1.63 [11]
	Ours	98.26	1.78
CYBHiDB	[11]	–	4.40 [11]
	[10]	98.74	2.45
	[49]	98.93	1.67
	Ours	99.14	1.23
UofTDB	[48]	–	3.70 [48]
	[10]	98.52	2.68
	[49]	98.77	2.43
	Ours	98.85	2.23

As shown in Table 5, our method outperformed the non-deep learning methods on all four databases. As for the deep-learning methods, our method performed much better than these methods on MITDB, CYBHiDB and UofTDB, and slightly worse than them on PTBDB. In addition, our method significantly improved the recognition performance over state-of-the-art methods on MITDB with only 47 subjects, and the experimental results showed that our method was more robust than the other methods when handling small-scale ECG data.

4.10 Analysis of Computation Time

As a multi-scale, layer-by-layer learning model, MD-CBF consumes more time in the training phase. For better insight on the computational complexity, we reported the training time of the proposed method and other methods on MITDB, as shown in Table 6.

Table 6. Training Time of Different Methods on MITDB

Method	Time (s)
[51]	3.54
[46]	3.87
[47]	29 640.00
[10]	27 860.00
Ours	767.00

As shown in Table 6, compared with non-deep-learning methods in [46, 51], our method took far more training time, but it had a clear advantage over methods based on a deep convolutional neural network in [10, 47]. As we know, training a deep model is time-consuming. Our model should be more efficient due to the one-step and level-to-level cascade learning without gradient backpropagation and many iterations, which are usually seen in deep neural networks.

5 Conclusions

The field of ECG biometric recognition has received considerable attention in recent years. Although some promising methods have been proposed, it is challenging to design a robust and precise method for only small-scale training ECG data. We presented a deep cascade framework for ECG biometric recognition, which has three key merits. First, we designed a BFG with L_1-norm sparsity and L_2-norm collaborative representation, which can efficiently deal with noise in ECG biometric recognition. Second, we proposed an adaptive weighted pooling strategy that can fully explore the discriminative information of segments with low noise. This strategy can be regarded as an attention mechanism in a deep learning model. Third, we presented a level-wise class coding without backpropagation, aiming to mine more discriminative features based on level-wise representation. Our work contributes to the improvement of robustness for ECG biometric recognition even with only small scale. In addition, our recognition approach is not only suitable for ECG signal but for the other biosignals, such as electromyogram (EMG) and electroencephalogram (EEG). In future work, we aim to improve the deep cascade structure to obtain more discriminative information for ECG biometric recognition.

References

[1] Odinaka I, Lai P H, Kaplan A D, Osullivan J A, Sirevaag E J, Rohrbaugh J W. ECG biometric recognition: A comparative analysis. *IEEE Trans. Information Forensics and Security*, 2012, 7(6): 1812-1824. DOI: 10.1109/TIFS.2012.2215324.

[2] Kim M, Pan S B. Deep learning based on 1-D ensemble networks using ECG for real time user recognition. *IEEE Trans. Industrial Informatics*, 2019, 15(10): 5656-5663. DOI: 10.1109/TII.2019.2909730.

[3] Wang J, She M, Nahavandi S, Kouzani A Z. Human identification from ECG signals via sparse representation of local segments. *IEEE Signal Processing Letters*, 2013, 20(10): 937-940. DOI: 10.1109/LSP.2013.2267593.

[4] Jaafar H, Ramli N, Nasir A S A. Implementation of kernel sparse representation classifier for ECG biometric system. *Journal of Telecommunication, Electronic and Computer Engineering*, 2018, 10(1-13): 89-94

[5] Li R, Yang G, Wang K, Huang Y, Yuan F, Yin Y. Robust ECG biometrics using GNMF and sparse representation. *Pattern Recognition Letters*, 2019, 129: 70-76. DOI: 10.1016/j.patrec.2019.11.005.

[6] Goshvarpour A, Goshvarpour A. Human identification using a new matching pursuit-based feature set of ECG. *Computer Methods and Programs in Biomedicine*, 2019, 172: 87-94. DOI: 10.1016/j.cmpb.2019.02.009.

[7] Zhang L, Liu J, Zhang B, Zhang D, Zhu C. Deep cascade model-based face recognition: When deep-layered learning meets small data. *IEEE Trans. Image Process*, 2019, 29: 1016-1029. DOI: 10.1109/TIP.2019.2938307.

[8] Abdeldayem S S, Bourlai T. A novel approach for ECG-based human identification using spectral correlation and deep learning. *IEEE Trans. Biometrics, Behavior, and Identity Science*, 2020, 2(1): 1-14. DOI: 10.1109/TBIOM.2019.2947434.

[9] Labati R D, Munoz E, Piuri V, Sassi R, Scotti F. Deep-ECG: Convolutional neural networks for ECG biometric recognition. *Pattern Recognition Letter*, 2019, 126: 78-85. DOI: 10.1016/j.patrec.2018.03.028.

[10] da Silva Luz E, Moreira G, Oliveira L S, Schwartz W R, Menotti D. Learning deep off-the-person heart biometrics representations. *IEEE Trans. Information Forensics and Security*, 2017, 13(5): 1258-1270. DOI: 10.1109/TIFS.2017.2784362.

[11] Hammad M, Liu Y, Wang K. Multimodal biometric authentication systems using convolution neural network based on different level fusion of ECG and fingerprint. *IEEE Access*, 2019, 7: 26527-26542. DOI: 10.1109/ACCESS.2018.2886573.

[12] Zhang Y, Xiao Z, Guo Z, Wang Z. ECG-based personal recognition using a convolutional neural network. *Pattern Recognition Letter*, 2019, 125: 668-676. DOI: 10.1016/j.patrec.2019.07.009.

[13] Zhou Z, Feng J. Deep forest: Towards an alternative to deep neural networks. In *Proc. the 26th Int. Artificial Intelligence*, August 2017, pp.3553-3559. DOI: 10.24963/ijcai.2017/497.

[14] Liu X B, Wang R, Cai Z, Cai Y, Yin X. Deep multigrained cascade forest for hyperspectral image classification. *IEEE Trans. Geoscience and Remote Sensing*, 2019, 57(10): 8169-8183. DOI: 0.1109/TGRS.2019.2918587.

[15] Wen H, Zhang J, Lin Q, Yang K, Huang P. Multi-level deep cascade trees for conversion rate prediction in recommendation system. In *Proc. the 23rd Int. Artificial Intelligence*, May 2019, pp.338-345. DOI: 10.1609/aaai.v33i01.3301338.

[16] Lim C, Woo W L, Dlay S, Gao B. Heartrate-dependent heartwave biometric identification with thresholding-based GMM-HMM methodology. *IEEE Trans. Industrial Informatics*, 2019, 15(1): 45-53. DOI: 10.1109/TII.2018.2874462.

[17] Meltzer D, Luengo D. Fiducial ECG-based biometry: Comparison of classifiers and dimensionality reduction methods. In *Proc. the 42nd Int. Telecommunications and Signal Processing*, Jul. 2019, pp.552-556. DOI: 10.1109/TSP.2019.8768891.

[18] Diab M O, Seif A, El-Abed M, Sabbah M. Individual identification using ECG signals. *Journal of Computer and Communications*, 2018, 6: 74-80. DOI: 10.4236/jcc.2018.61008.

[19] Dong X, Si W, Huang W. ECG-based identity recognition via deterministic learning. *Biotechnology & Biotechnological Equipment*, 2018, 32(3): 769-777. DOI: 10.1080/13102818.2018.1428500.

[20] Rahman S A E. Biometric human recognition system based on ECG. *Multimedia Tools and Applications*, 2019, 78(3): 17555-17572. DOI: 10.1007/s11042-019-7152-0.

[21] Srivastva R, Singh Y N. Human recognition using discrete cosine transform and discriminant analysis of ECG. In *Proc. the 4th Int. Image Information Processing*, Dec. 2017, pp.368-372. DOI: 10.1109/ICIIP.2017.8313742.

[22] Bassiouni M M, El-Dahshan E A, Khalefa W, Salem A M. Intelligent hybrid approaches for human ECG signals identification. *Signal, Image and Video Processing*, 2018, 12(5): 941-949. DOI: 10.1007/s11760-018-1237-5.

[23] Ur Rehman U, Kamal K, Iqbal J, Sheikh M F. Biometric identification through ECG signal using a hybridized approach. In *Proc. the 5th Int. Computing and Artificial Intelligence*, Apr. 2019, pp.226-230. DOI: 10.1145/3330482.3330496.

[24] Wu S C, Chen P T, Swindlehurst A L, Hung P L. Cancelable biometric recognition with ECGs: Subspace-based approaches. *IEEE Trans. Information Forensics and Security*, 2019, 14(5): 1323-1336. DOI: 10.1109/TIFS.2018.2876838.

[25] He H, Tan Y. Automatic pattern recognition of ECG signals using entropy-based adaptive dimensionality reduction and clustering. *Applied Soft Computing*, 2017, 55: 238-252. DOI: 10.1016/j.asoc.2017.02.001.

[26] Srivastva R, Singh Y. ECG biometric analysis using Walsh-Hadamard transform. In *Advances in Data and Information Sciences*, Kolhe M L, Trivedi M C, Tiwari S, Singh V (eds.), Springer, 2018, pp.201-210. DOI: 10.1007/978-981-10-8360-0 19.

[27] Zheng G, Wang Y, Sun X, Sun Y, Ji S. Study on Matthew effect based feature extraction for ECG biometric. In *Proc. the 8th Int. Conf. Intelligent Science and Big Data Engineering*, Aug. 2018, pp.623-634. DOI: 10.1007/978-3-030-02698-1-54.

[28] Pinto J R, Cardoso J S, Lourenço A. Evolution, current challenges, and future possibilities in ECG biometrics. *IEEE Access*, 2018, 6: 34746-34776. DOI: 10.1109/ACCESS.2018.2849870.

[29] Pal A, Singh Y. ECG biometric recognition. In *Proc. the 4th Int. Conf. Mathematics and Computing*, Jan. 2018, pp.61-73. DOI: 10.1007/978-981-13-0023-37.

[30] Hejazi M, Al-Haddad S A R, Singh Y P, Hashim S J, Aziz A F A. ECG biometric authentication based on non-fiducial approach using kernel methods. *Digital Signal Processing*, 2016, 52: 72-86. DOI: 10.1016/j.dsp.2016.02.008.

[31] Chen H, Huang C, Huang Q, Zhang Q, Wang W. ECGadv: Generating adversarial electrocardiogram to misguide arrhythmia classification system. In *Proc. the 34th AAAI Conference on Artificial Intelligence*, Feb. 2020, pp.3446-3453. DOI: 10.1609/aaai.v34i04.5748.

[32] Utkin L V, Ryabinin M A. Discriminative metric learning with deep forest. *International Journal of Artificial Intelligence Tools*, 2017, 28(2): 195-204. DOI: 10.1142/S0218213019500076.

[33] Su R, Liu X, Wei L, Zou Q. Deep-Resp-forest: A deep forest model to predict anti-cancer drug response. *Methods*, 2019, 166: 91-102. DOI: 10.1016/j.ymeth.2019.02.009.

[34] Pang M, Ting K M, Zhao P, Zhou Z. Improving deep forest by confidence screening. In *Proc. the 20th Int. Data Mining*, Nov. 2018, pp.1194-1199. DOI: 10.1109/ICDM.2018.00158.

[35] Lan X, Zhang S, Yuen P C, Chellappa R. Learning common and feature-specific patterns: A novel multiple-sparse-representation-based tracker. *IEEE Trans. Image Processing*, 2017, 27(4): 2022-2037. DOI: 10.1109/TIP.2017.2777183.

[36] Panagakis Y, Kotropoulos C, Arce G R. Music genre classification via joint sparse low-rank representation of audio features. *IEEE/ACM Trans. Audio, Speech, and Language Processing*, 2014, 22(12): 1905-1917. DOI: 10.1109/TASLP.2014.2355774.

[37] Wright J, Yang A Y, Ganesh A, Sastry S, Ma Y. Robust face recognition via sparse representation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2009, 31(2): 210-227. DOI: 10.1109/TPAMI.2008.79.

[38] Moody G B, Mark R G. The impact of the MIT-BIH arrhythmia database. *IEEE Engineering in Medicine and Biology Magazine*, 2001, 20(3): 45-50. DOI: 10.1109/51.932724.

[39] Bousseljot R, Kreiseler D, Schnabel A. Nutzung der EKG-signaldatenbank CARDIODAT der PTB uber das internet. *Biomedizinische Technik/Biomedical Engineering*, 2009, 40(s1): 317-318. DOI: 10.1515/bmte.1995.40.s1.317.

[40] Da Silva H P, Lourenço A, Fred A, Raposo N, Aires-de-Sousa M. Check your biosignals here: A new dataset for off-the-person ECG biometrics. *Computer Methods and Programs in Biomedicine*, 2014, 113(2): 503-514. DOI: 10.1016/j.cmpb.2013.11.017.

[41] Pouryayevali S, Wahabi S, Hari S, Hatzinakos D. On establishing evaluation standards for ECG biometrics. In *Proc. the 2014 Int. Conf. Acoustics, Speech and Signal Processing*, May 2014, pp.3774-3778. DOI: 10.1109/ICASSP.2014.6854307.

[42] Pan J, Tompkins W J. A real-time QRS detection algorithm. *IEEE Trans. Biomedical Engineering*, 1985, BME-32(3): 230-236. DOI: 10.1109/TBME.1985.325532.

[43] Biel L, Pettersson O, Philipson L, Wide P. ECG analysis: A new approach in human identification. *IEEE Trans. Instrumentation and Measurement*, 2001, 50(3): 808-812. DOI: 10.1109/19.930458.

[44] Chan A A, Hamdy M M, Badre A, Badee V. Wavelet distance measure for person identification using electrocardiograms. *IEEE Trans. Instrumentation & Measurement*, 2008, 57(2): 248-253. DOI: 10.1109/TIM.2007.909996.

[45] Yang J, Yu K, Gong Y, Huang T S. Linear spatial pyramid matching using sparse coding for image classification. In *Proc. the 2009 Computer Vision and Pattern Recognition*, Jun. 2009, pp.1794-1801. DOI: 10.1109/CVPR.2009.5206757.

[46] Wang K, Yang G, Huang Y, Yin Y. Multi-scale differential feature for ECG biometrics with collective matrix factorization. *Pattern Recognition*, 2020, 102: Article No. 107211. DOI: 10.1016/j.patcog.2020.107211.

[47] Chu Y, Shen H, Huang K. ECG authentication method based on parallel multi-scale one-dimensional residual network with center and margin loss. *IEEE Access*, 2019, 7: 51598-51607. DOI: 10.1109/ACCESS.2019.2912519.

[48] Komeili M, Louis W, Armanfard N, Hatzinakos D. Feature selection for nonstationary data: Application to human recognition using medical biometrics. *IEEE Trans. Cybern.*, 2018, 48(5): 1446-1459. DOI: 10.1109/TCYB.2017.2702059.

[49] Huang Y W, Yang G P, Wang K K, Liu H Y, Yin Y L. Learning joint and specific patterns: A unified sparse representation for off-the-person ECG biometric recognition. *IEEE Trans. Information Forensics & Security*, 2021, 16(1): 147-160. DOI: 10.1109/TIFS.2020.3006384.

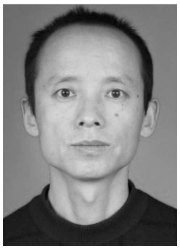
[50] Lynn H M, Pan S B, Kim P. A deep bidirectional GRU network model for biometric electrocardiogram classification based on recurrent neural networks. *IEEE Access*, 2019, 7: 145395-145405. DOI: 10.1109/ACCESS.2019.2939947.

[51] Louis W, Komeili M, Hatzinakos D. Continuous authentication using one-dimensional multi-resolution local binary patterns (1DMRLBP) in ECG biometrics. *IEEE Trans. Information Forensics & Security*, 2016, 11(12): 2818-2832. DOI: 10.1109/TIFS.2016.2599270.

Yu-Wen Huang received his Master's degree in computer science from Guangxi Normal University, Guilin, in 2009. He is an associate professor in the School of Computer, Heze University, Heze, and pursuing his Ph.D. degree at Shandong University, Jinan. His research interests include ECG recognition, biometrics and machine learning.

Gong-Ping Yang received his Ph.D. degree in computer software and theory from Shandong University, Jinan, in 2007. Now he is a professor in the School of Software, Shandong University, Jinan, and an adjunct professor in the School of Computer, Heze University, Heze. His research interests are pattern recognition, image processing, biometrics, and so forth.

Kui-Kui Wang received her Master's degree in computer science from Shandong University, Jinan, in 2017. Currently, she is pursuing her Ph.D. degree at Shandong University, Jinan. Her main research interests are biometrics and machine learning.



Hai-Ying Liu received his Ph.D. degree in 2019 from Shandong University, Jinan. Now, he is a lecturer in the Department of Computer Engineering, Changji University, Changji. His main research interests are biometrics and machine learning.

Yi-Long Yin received his Ph.D. degree in mechanical engineering from Jilin University, Changchun, in 2000. From 2000 to 2002, he was a postdoctoral fellow with the Department of Electronic Science and Engineering, Nanjing University, Nanjing. He is currently the director of the Machine Learning and Applications Group and a professor with Shandong University, Jinan. His research interests include machine learning, data mining, and biometrics.