Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

ECGID: a human identification method based on adaptive particle swarm optimization and the bidirectional LSTM model*

Yefei ZHANG, Zhidong ZHAO‡, Yanjun DENG, Xiaohong ZHANG, Yu ZHANG

School of Electronics and Information Engineering, *Hangzhou Dianzi University*, *Hangzhou 300318*, *China*

E-mail: zhangyf@hdu.edu.cn; zhaozd@hdu.edu.cn; yanjund@hdu.edu.cn; xhzhang@hdu.edu.cn; zy2009@hdu.edu.cn

Received Sept. 29, 2020; Revision accepted Jan. 31, 2021; Crosschecked June 21, 2021; Published online Sept. 2, 2021

Abstract: Physiological signal based biometric analysis has recently attracted attention as a means of meeting increasing privacy and security requirements. The real-time nature of an electrocardiogram (ECG) and the hidden nature of the information make it highly resistant to attacks. This paper focuses on three major bottlenecks of existing deep learning driven approaches: the lengthy time requirements for optimizing the hyperparameters, the slow and computationally intense identification process, and the unstable and complicated nature of ECG acquisition. We present a novel deep neural network framework for learning human identification feature representations directly from ECG time series. The proposed framework integrates deep bidirectional long short-term memory (BLSTM) and adaptive particle swarm optimization (APSO). The overall approach not only avoids the inefficient and experience-dependent search for hyperparameters, but also fully exploits the spatial information of ordinal local features and the memory characteristics of a recognition algorithm. The effectiveness of the proposed approach is thoroughly evaluated in two ECG datasets, using two protocols, simulating the influence of electrode placement and acquisition sessions in identification. Comparing four recurrent neural network structures and four classical machine learning and deep learning algorithms, we prove the superiority of the proposed algorithm in minimizing overfitting and self-learning of time series. The experimental results demonstrated an average identification rate of 97.71%, 99.41%, and 98.89% in training, validation, and test sets, respectively. Thus, this study proves that the application of APSO and LSTM techniques to biometric human identification can achieve a lower algorithm engineering effort and higher capacity for generalization.

Key words: ECG biometrics; Human identification; Long short-term memory (LSTM); Adaptive particle swarm optimization (APSO)

https://doi.org/10.1631/FITEE.2000511 **CLC number:** TN911.72

1 Introduction

Security and privacy of biometric information is considered one of the major concerns in this new era of artificial intelligence (AI). Traditional identification methods such as passwords and smart cards are now outdated because they can be lost, stolen, or shared. The Internet of Things (IoT) and AI technologies have promoted the maturation and widespread application of modern biometrics. Expert and intelligent biometric based identification technologies (e.g., fingerprints and voiceprints) are attracting more and more interest in many areas, ranging from national security to daily life (e.g., criminal investigations, healthcare applications, and smartphone access). However, IoT and AI technologies can also facilitate the development of pseudo-fingerprint and pseudo-face technologies, increasing the risk of forgery and hacking of personal information.

To summarize the existing biometric technologies, biometric information truly defines each one of us to a certain extent which is invariable and unique.

‡ Corresponding author

* Project supported by the Zhejiang Province Public Welfare Technology Application Research Project (No. LGG20F010008), the National Natural Science Foundation of China (No. 61571173), and the Welfare Project of the Science Technology Department of Zhejiang Province, China (No. LGG18F010012)

ORCID: Zhidong ZHAO, https://orcid.org/0000-0001-6659-3732

However, as most biometric measures in use are extrinsic and thus susceptible to replay attacks, once information is leaked or abused, it can cause incalculable losses. These limitations have led to the development of the following smart solutions: (1) improving voice recognition by sophisticated preprocessing, followed by heuristic selection of significant features in multiple domains; (2) integrating face- and body-related soft biometric traits; (3) simultaneous application of numeric passwords and fingerprint information; (4) finding new biometric technologies with built-in anti-counterfeiting features.

Since the beginning of the new millennium, the electrocardiogram (ECG) has been studied as a biometric trait for security systems and other applications. It has already been indicated that an ECG is unique to each individual and its real-time, hidden nature makes it hard to capture and input into a system for fraudulent purposes (Biel et al., 2001; Salloum and Kuo, 2017). It is considered to be the most promising approach by many scholars worldwide (Bassiouni et al., 2018). Recently, advancements in wireless sensor network technology have enabled the proliferation of miniaturized sensors, capable of biomedical signal acquisition. Furthermore, as a powerful recognition/classification tool, deep neural networks (DNNs) have been extensively used in computer vision, natural language processing, driverless vehicles, and other important fields. Hence, research on deep learning (DL) driven ECG recognition approaches is applicable to a broad spectrum of technologies in line with the current key research and breakthrough direction of international ECG studies. However, many technological challenges still need to be faced to allow for the application of this technology at scale.

1. Difficulties in optimizing hyperparameters in the network model's training phase. A common problem in training pattern recognition or classification models with DNNs is that a large number of parameters need to be manually set during the training phase. These hyperparameters directly control and affect the network's topology, thus affecting both speed of model fitting and classification accuracy. However, parameter selection mostly relies on a trial-and-error process and the experience of researchers. Hence, the entire training consumes quite a bit of time, labor, and computational resources, and is not necessarily effective or accurate.

2. Low identification speed. To some extent, the more network layers there are, the lower the recognition speed will be. Also, many existing convolutional neural network (CNN) based identification algorithms use two-dimensional (2D) high-pixel images as input data, which improves the accuracy but reduces the recognition rate. This speed limitation has created great obstacles for the application of fast and efficient identification (Liu et al., 2018).

3. Influence of different acquisition statuses on the ECG signal's waveform characteristics. As an ECG has a non-stationary and time-varying sequence, many factors (electrode placement and acquisition sessions) affect waveform changes during the acquisition process. For example, we may have the same individual presenting slight changes at different collection times (Ahmadi et al., 2015). It remains unknown whether the intraclass differences caused by the above two acquisition statuses could affect the results.

Aiming to tackle the aforementioned difficulties, we introduce a DL-driven ECG identification system that integrates deep bidirectional long short-term memory (BLSTM) and adaptive particle swarm optimization (APSO) (Fig. 1). We design a deep BLSTM-APSO network model supplemented by a denoised one-dimensional (1D) ECG signal as the input vector. Recurrent neural networks (RNNs) have introduced the concept of time series into the network structure design, fully considering the original signal changes in the time dimension and the memory characteristics of matching recognition algorithms to key features. LSTM further alleviates the main problem of RNN training, which is gradient disappearance/diffusion, by designing and adding three weight gates based on an RNN. Specifically, BLSTM consists of forward and backward LSTM to allow long-period signals to be trained synchronously in both forward and backward directions. The PSO algorithm for the adaptive learning strategy is used to match ECG signal features with the deep BLSTM network topology to optimize the recognition performance. The APSO algorithm can avoid local optimal solutions and improve the accuracy of the parameter optimization process through the adaptive partition of the particle swarm and updating of the particle position according to the distribution of the

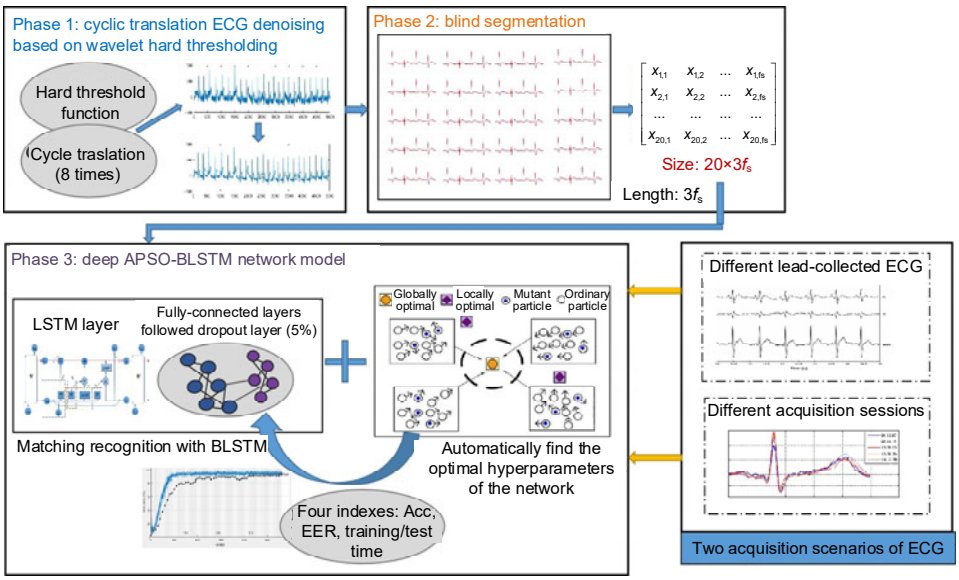

Fig. 1 Architecture of the deep learning driven ECG identification system

Highlighting the denoising section: considering real-life scenarios and applications, ECG signals are highly susceptible to noise. Thus, the cyclic translation ECG denoising based on wavelet hard thresholding is first used to remove noise. Then the denoised ECG is blindly segmented into signal segments with an equal length of $3f_s$ (f_s: sampling frequency), without leveraging any heartbeat location information. For this part, the detailed instructions are referred to Zhao et al. (2018)

population of solutions. Compared with other DL algorithms, BLSTM-APSO presents stronger adaptability in time-series data analysis.

To address waveform changes, we investigate the impact of different electrode placement and acquisition sessions for ECG biometrics, using two protocols. We also introduce four options of deep RNN structures (LSTM+Dropout B, BLSTM, BLSTM+Dropout A, BLSTM+Dropout B), demonstrating that the proposed deep BLSTM network structure does have advantages in preventing overfitting and increasing recognition accuracy. In parallel, four classical machine learning and deep learning algorithms (the proposed deep BLSTM, simple CNN, deep CNN, and support vector machine (SVM)) are compared to verify the validity of our conclusions. With all these experimental comparisons, we prove that the proposed method outperforms many alternative methods for ECG-based biometrics tasks on several public data sets.

2 Related works

Many proposed ECG identification algorithms are based on two strategies:

1. Fiducial or non-fiducial methods. Nowadays, very few studies use only one of linear, nonlinear, wavelet transform, and other fiducial or non-fiducial methods for feature extraction and classification. The majority of studies combine machine learning or deep learning algorithms. In the past 20 years, Biel et al. (2001) and many other scholars have studied the characteristics of ECG signals, extracted multiple features, and deployed various classification algorithms to achieve ECG identification. Among them, Tantawi et al. (2015) used discrete biorthogonal wavelet filters for feature extraction, achieving a high recognition accuracy of 98.8% with a radial basis function (RBF) network. Yu et al. (2017) extracted several time-domain features through R-wave detection, achieving good identification performance. Choi et al. (2019) recently introduced the spectrogram feature extraction method, demonstrating an accuracy of 93% with 100 ECG samples using only a simple linear classifier. In addition, many scholars have improved and optimized the classical fiducial and non-fiducial algorithms, enriching ECG features and improving their accuracy. Some striking examples were given in Palaniappan and Krishnan (2004) and Agrafioti and Hatzinakos (2008), in which the classical Pan-Thompkins algorithm (Pan and Tompkins,

1985) has been improved, which was initially developed for QRS detection. However, irrespective of categorization as a fiducial or non-fiducial algorithm, the above algorithm relies heavily on the waveform characteristics of ECG signals. Although it is very effective for small samples with satisfactory quality, the relationships among local features are treated independently. Moreover, the ECG signal is highly susceptible to interference by noise, the subject's health status, and many other factors in the actual acquisition process, and thus its practical application still faces enormous challenges.

2. Feature self-learning. With the vigorous development of deep learning, many attempts have been made to study DL-driven and ECG-based recognition technologies. In particular, during the period 2010–2019, scholars worldwide have conducted considerable research on DNN. Zhang et al. (2017) transformed the ECG signal segment to the wavelet domain to enrich the data representation and then used a 1D CNN for recognition. Labati et al. (2019) suggested direct use of the original ECG signal as the input vector, integrated the Hamming distance into a CNN, and achieved 100% accuracy with a dataset of 144 recordings. In addition, many studies have combined the time-frequency domain features of ECG signals with DNNs. Some striking examples are da Silva Luz et al. (2018), who used CNN networks for identity recognition, and Oh et al. (2018) and Yildirim (2018), who deployed LSTM for ECG classification.

Such algorithms greatly simplify the feature extraction process, but their high precision and fast recognition still rely on powerful computers and abundant large data sets.

3 A deep APSO-BLSTM-based biometric algorithm

The proposed deep bidirectional LSTM network model is based on APSO. Conventional LSTM networks have had great success in avoiding extreme gradient values, but a large number of hyperparameters in the network training process need to be manually set, such as the batch size, initial learning rate, and number of hidden layer units. These hyperparameters directly control the topology of the network model and affect the recognition results. In this section, for completeness, we initially present the basic LSTM, BLSTM, and APSO methods, and then introduce the new architecture developed for our ECG biometrics system.

3.1 Deep bidirectional LSTM

RNNs introduce the concept of time series into network structure design. Unlike traditional neural network structures, RNNs achieve memorization and learning from past moment features through the interconnection of the input, hidden, and output layers, exploiting the connection of input data in the time dimension. Unlike classical feed-forward neural networks, RNN neuron outputs are applied recursively to their own inputs through self-connection of hidden layers, capturing sequential information. In summary, RNNs are characterized by retroactive usage of the original data. Although the original data available in practical applications is limited, it is enough to learn and capture sufficient sequential features. Accordingly, RNNs show stronger adaptability in time series data analysis and have been successfully applied to natural language processing, image description generation, and more.

Due to the iterative nature of standard RNN, the biggest problems in model training are: (1) gradient disappearance and explosion, especially in deep network structures; (2) inappropriate initial weight settings. Hochreiter and Schmidhuber (1997) proposed the LSTM model to improve the RNN network structure by changing from the traditional cell structure to a structure with three weight gates. Moreover, this alleviates the problems of gradient disappearance and explosion, and insufficient long-term memory, so that RNN can be effectively applied to long-distance time series (Oh et al., 2018; Yildirim, 2018).

Fig. 2 presents a classical LSTM memory block and its internal structure. Each cell consists of four main elements: input gate $\boldsymbol{i}$, forget gate $\boldsymbol{f}$, output gate $\boldsymbol{o}$, and cell state $\boldsymbol{c}$. The forget, input, and output gates control the memory retention of the cell's internal state, their receptivity to the cell's input, and the degree of recognition of cell output, respectively. All three gates are determined by the current input signal and the output of the hidden layer at the previous moment. The structure stores information for a long time by updating the internal state. The key to LSTM is the cell state, named c, which maintains the

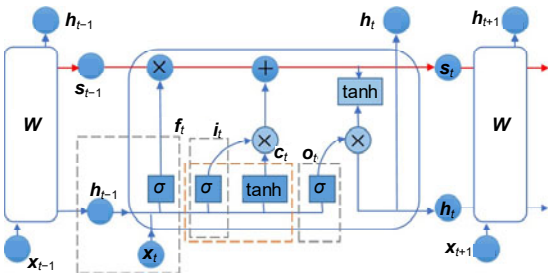

Fig. 2 Structure of the LSTM cell including four interactive layers

memory of the cell state at T-moment and is adjusted by inputting and forgetting:

Forget gate:

$$\boldsymbol{f}_t = \text{sigmod}\left(\boldsymbol{b}_\text{f} + \boldsymbol{U}_\text{f}\boldsymbol{x}_t + \boldsymbol{W}_\text{f}\boldsymbol{h}_{t-1}\right). \tag{1}$$

Input gate:

$$\boldsymbol{i}_t = \text{sigmod}\left(\boldsymbol{b}_\text{in} + \boldsymbol{U}_\text{in}\boldsymbol{x}_t + \boldsymbol{W}_\text{in}\boldsymbol{h}_{t-1}\right). \tag{2}$$

Output gate:

$$\boldsymbol{o}_t = \text{sigmod}\left(\boldsymbol{b}_\text{o} + \boldsymbol{U}_\text{o}\boldsymbol{x}_t + \boldsymbol{W}_\text{o}\boldsymbol{h}_{t-1}\right). \tag{3}$$

Candidate memory (new contribution):

$$\boldsymbol{c}_t = \tanh\left(\boldsymbol{b} + \boldsymbol{U}\boldsymbol{x}_t + \boldsymbol{W}\boldsymbol{h}_{t-1}\right). \tag{4}$$

New memory generated by the cell:

$$\boldsymbol{s}_t = \boldsymbol{f}_t \otimes \boldsymbol{s}_{t-1} + \boldsymbol{i}_t \otimes \boldsymbol{c}_t. \tag{5}$$

Output of the cell:

$$\boldsymbol{h}_t = \boldsymbol{o}_t \otimes \tanh\left(\boldsymbol{s}_t\right). \tag{6}$$

Output of the network:

$$\boldsymbol{z}_t = \text{softmax}\left(\boldsymbol{V}\boldsymbol{h}_t + \boldsymbol{c}\right), \tag{7}$$

where $\boldsymbol{x}$ and $\boldsymbol{h}$ represent the current input and hidden layer vectors of LSTM respectively, $\boldsymbol{b}$ is the offset vector, and $\boldsymbol{U}$, $\boldsymbol{V}$, $\boldsymbol{W}$ are the input weight matrix, output weight matrix, and regression weight matrix, respectively.

LSTM is very suitable for time series data modeling because of its inherent design characteristics. BLSTM combines forward LSTM and backward LSTM to better capture the forward and backward changes in the time dimension and the memory aspects of matching the recognition algorithm to key features. The basic structure of the BLSTM network model is shown in Fig. 3. For the present study, we design a deep bidirectional LSTM network for identification. In this method, deep BLSTM networks are generated by stacking multiple recurrent BLSTM layers. The network is deep in the sense that each BLSTM layer's output works as an input to the next BLSTM layer. Each recurrent layer may be unfolded in time to an equivalent feed-forward network whose layers share the same hyperparameters. Stacking recurrent layers allows for processing of the sequence at different timescales and yields a richer set of temporal features.

3.2 Adaptive particle swarm optimization

As explained above, a large number of hyperparameters in the network training process need to be manually set. The present paper proposes a parameter optimization approach based on an APSO algorithm to alleviate this problem. In the subgroup, the local optimum particle guides the search direction of the whole subgroup. However, all particles in the classical PSO optimization algorithm use a consistent learning strategy to update the particle position (Rodriguez and Laio, 2014). Once the local optimum particle deviates from the search direction of the optimal solution, the whole subgroup will fall into the local optimum. Therefore, the local optimal particle needs to break through the control of subgroup and obtain information from other subgroups. A fast search and clustering method with high performance is proposed to perform the rapid division of subgroups. For each subgroup, we propose different learning strategies to update the positions of different types of particles in order to improve population diversity.

1. Optimization 1, partition of particle swarm

We follow the two basic principles of a cluster center: (1) It should be surrounded by the points with lower local density; (2) It should be far away from the points with higher local density. The population $S = \{x_i\}_{i=1}^{H}$ is composed of H particles, the local density of each particle is defined as ρ_i, and the distance to the higher local density particle δ_i is as shown in the following Eqs. (8) and (9), where d_{ij} is the Euclidean distance between particles x_i and x_j, and d_c is the cut-off distance:

$$\rho_i = \sum_{j \neq i} \exp\left(-\left(\frac{d_{ij}}{d_\text{c}}\right)^2\right), \tag{8}$$

$$\delta_i = \min_{j:\rho_j>\rho_i}\left(d_{ij}\right). \tag{9}$$

Analyzing Eq. (9), if the density of the particle x_i is the maximum local density, δ_i is far greater than the distance δ of its nearest particle. Therefore, the center of the subgroup is usually the particles with abnormally large δ, and the density ρ of these particles is relatively high. Thus, the particles with larger δ and ρ are selected as the clustering center. The other particles x_j are classified into subgroups of samples whose density is larger than x_j and whose distance is closest to x_j. Based on the results of subgroup partition, the particles in each population are divided into two categories: ordinary particles and local optimum ones. The objectives of this classification are to ensure that the cluster's center of dataset samples can be automatically obtained, and to achieve efficient clustering of an arbitrary shape with dataset samples.

2. Optimization 2, particle update

Ordinary particle: extend local search ability under the guidance of the optimal particle in each subgroup:

$$x_i^d = \omega x_i^d + c_1\alpha_1^d\left(\text{pbest}_i^d - x_i^d\right) + c_2\beta_2^d\left(\text{cbest}_c^d - x_i^d\right), \tag{10}$$

where ω is the inertia weight, c_1, c_2 are the learning factors, α, β are random numbers (α, $\beta\in[0, 1]$), pbest is the global optimal particle, and cbest is the optimal particle in the c^{th} subgroup.

Locally optimal particle: update by synthesizing the information of each subgroup:

$$\begin{aligned} x_i^d = {} & \omega x_i^d + c_1\alpha_1^d\left(\text{pbest}_i^d - x_i^d\right) \\ & + c_2\beta_2^d\left(\frac{1}{C}\sum_{c=1}^{C}\text{cbest}_c^d - x_i^d\right). \end{aligned} \tag{11}$$

The local optimal particle not only guides the learning of ordinary particles but acts as a medium for information exchange between subgroups. Considering that the local optimal particle is most likely to find the optimal solution in the subgroup, in Eq. (11) the average information of the local optimal particle in each subgroup is used to guide the particle's update process. By sharing information among the subgroups, we can facilitate information transfer between them, further improving the diversity of the population and avoiding falling into local optimum solutions.

The PSO variation which deploys an adaptive learning strategy is used to match ECG signal features with deep BLSTM network topology in order to obtain the optimal hyperparameters for best recognition performance.

3.3 Deep APSO-BLSTM biometric model

The proposed deep BLSTM network model is based on the APSO algorithm, which uses the original ECG signals as input vectors (Fig. 3).

Step 1: simple construction of input vectors. The denoised 1D ECG signal is blindly segmented into equal-length segments using a fixed window. The window's length is regulated as $3f_s$ to ensure that there is at least one complete heartbeat. Thereafter, the divided ECG signal is directly used as the input vector of the model. It is noteworthy that for each subject, 20 blind segmentation operations are performed, and then 20 samples are obtained as the input vector of the current subject. That is to say, the input vector size of each individual is $20\times3f_s$. In contrast to the CNN-based DL recognition/classification algorithm that requires complex preprocessing to transform a 1D signal into a 2D image, the proposed method makes full use of the advantages of DNNs in feature self-learning. This is achieved without complicated preprocessing steps using the BLSTM network to make full use of the characteristics of time series in the time dimension before and after changes of amplitude, skewness, kurtosis, etc.

Step 2: preliminary design of the deep BLSTM network structure. A DNN structure is constructed with several BLSTM structures. Overfitting is a considerable obstacle for this type of network. In this study, we introduce dropout at different layers to prevent overfitting (Srivastava et al., 2014). Randomly dropping out part of the network during the training process at the specified rates prevents the neurons from adapting overly well to the training data. As the neurons are dropped out, the connecting weights are excluded from the update process. This attribute forces the network to learn from imperfect patterns and thus improves the generalization properties of the model.

Step 3: random initialization of APSO. Three hyperparameters (batch size, initial learning rate, and number of hidden layer units) are optimized by APSO,

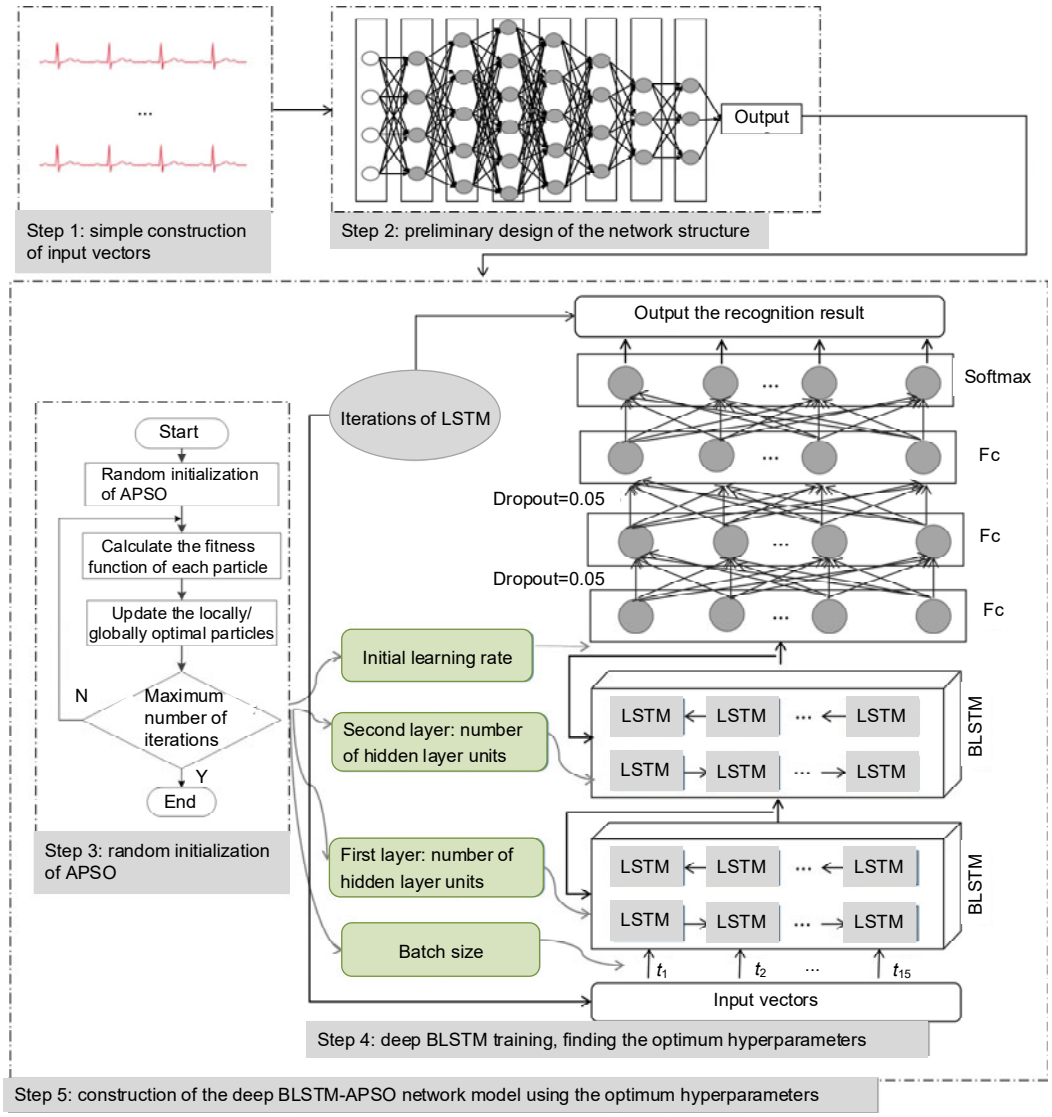

Fig. 3 Detailed structure of the deep APSO-BLSTM network used for identification

and the position information of each particle is randomly initialized according to the range of the hyperparameters. Through Eqs. (8) and (9), the local density ρ_i and the distance from ρ_i to the higher local density particles which is labeled σ_i are calculated, achieving adaptive partition of the particle swarm.

Step 4: deep BLSTM training, finding the optimum hyperparameters. The deep BLSTM network model is established according to the value of the hyperparameter corresponding to the particle position. The recognition results under the current hyperparameters are used as the fitness values of each particle in the training and validation sets. The global optimal particle position pbest and the local optimal particle position gbest are determined based on the particle fitness value and results of particle swarm partition. Thereafter, particle position is updated according to Eqs. (10) and (11). The optimal hyperparameter value is returned until the termination condition (maximum number of iterations) is met.

Step 5: construction of the deep BLSTM-APSO network model using the optimal hyperparameters. Through ECG data training, the trained model is saved for identification.

Although the classical PSO algorithm is simple in design and fast in convergence, it is easy for it to fall into a local optimal solution. As described in detail below, the APSO algorithm enables avoidance

of local optima and improves the accuracy of parameter optimization, through adaptive partition of the particle swarm and particle updating according to the distribution of the population itself. The adaptive characteristics of APSO enable the deep BLSTM to quickly and accurately determine the optimal hyperparameters based on ECG waveform features, achieving an effective combination of model network structure and ECG time-series features.

4 Experiments and results

4.1 Data description

We carried out a large number of experiments using the PhysioNet/Cinc Challenge 2011 dataset and the ECG-ID dataset from the PhysioNet database. Table 1 presents the characteristics and role of ECG signals, using a ratio of 4:1 between training and validation sets. That is, during the model's training process, 4/5 of the 90 subjects were used as the training set and 1/5 as the validation set. The validation set was used for parameter tuning and identification of the optimal unit numbers of the designed models. An independent test set was used to evaluate the performance of the trained models.

The challenge data is standard multi-lead/single-lead ECG recordings, based on chest collection. Among them, D1 is standard 12-lead ECG recordings (leads I, II, II, aVR, aVL, aVF, V1, V2, V3, V4, V5, and V6) with multiple sets of recordings. Although D2 was obtained from only 90 volunteers (44 men and 46 women), everyone had multiple sets of recordings, ranging from one day to periodic collection over six months. Intraclass differences in ECG signals are prone to cause recognition errors in most biometric systems. This study uses two protocols, as discussed below, to analyze the influence of intraclass difference in multi-lead collected ECG and multi-session collected ECG on ECG-based biometrics:

1. Protocol 1. In this protocol, we used only ECG recordings from lead I in D1, and the same session in D2, for model training. That is, 20 groups of short period signals from lead I-collected ECG/the same session-collected ECG were intercepted, forming the input vector of the subject. Then a lead II-collected ECG dataset in D1, and another session-collected ECG dataset in D2 were used for test. Therefore, the obtained APSO-BLSTM model was trained with the same session and the same lead collected dataset.

2. Protocol 2. In this protocol, we used ECG recordings from two different leads (I and II) in D1, and two sessions from different days in D2, for model training. That is, as for D1, 10 groups of short-period signals were intercepted respectively from lead I- and lead II-collected ECG signals, forming the input vector of the subject. Similarly, as for D2, 10 groups of short-period signals were intercepted respectively from different two days. Then a lead II-collected ECG dataset in D1 and another session-collected ECG dataset were used for test. Therefore, the obtained model was trained with multi-session and multi-lead collected datasets.

4.2 Setup of the case environment, parameters, and comparison

1. Case environment. Neural networks are inherently parallel algorithms. Graphical processing units (GPUs) are frequently adopted as the execution environment to take advantage of the parallel nature of RNNs and to expedite the classification process. In this study, the algorithm was implemented using the MATLAB environment, deploying a single CUDA-enabled NVIDIA Quadro P4000 GPU.

2. Parameters. The following parameters were used for the APSO algorithm: the particle swarm

Table 1 Characteristics of ECGs obtained from the Physionet ECG database

Protocol	Dataset	Data source	Characteristics and role	Acquisition method	Number of subjects
1	D1	PhysioNet/Cinc Challenge 2011	Lead I-collected	Standard 12-lead	Training+validation: 90 Test: 90
	D2	ECG-ID Database	Collected from the same session	Lead I	
2	D1	PhysioNet/Cinc Challenge 2011	Different leads (leads I, II)	Standard 12-lead	
	D2	ECG-ID Database	Collected from different sessions, ranging from 1 d to 2 months	Lead I	

number was set to 30, the maximum number of iterations to 200, the inertia weight to ω=0.96, and the acceleration factor to α=0.5, β=0.7. The range of batch size, the initial learning rate, and the number of hidden layer units were set to [1, 120], [10^{-4}, 10^{-3}], and [10, 400], respectively. The training times were directly controlled by the fitness value of the APSO-BLSTM model, which was the recognition rate with the current hyperparameters.

3. Comparison. Four different network approaches were tested for network structure optimization and to analyze the need for dropout and BLSTM layers. These were labeled as schemes A to D and a detailed description of each is included in Table 2. Moreover, the proposed deep APSO-BLSTM approach was benchmarked against three other methods. The methods used for comparison were the simple CNN model proposed in Zhao et al. (2018), the deep CNN model (original GoogLeNet), and the machine-learning classifier SVM.

The above ECG recognition algorithms were evaluated by four accepted formulas and measures: (1) training time (average computational time consumed by each iteration during the model training process); (2) testing time (authentication time for each subject starting from the input of ECG signals to returning the recognition result); (3) equal error rate (EER), which is the value of the false rejection rate (FRR) when FRR is equal to the false acceptance rate (FAR); (4) acc, as shown in the following Eqs. (12) and (13).

$$\text{Training / Validation acc} = T/(T+F), \tag{12}$$

$$\text{Test acc} = \frac{N_{\text{true}}}{N_{\text{total}}}, \tag{13}$$

where T and F represent the numbers of recurrence plots with true acceptance and false acceptance, respectively, N_{true} means the number of true recognitions (the logical definition of true recognition is as given in Fig. 4), and N_{total} represents the total number of testees in the test set.

4.3 Optimization of the deep APSO-BLSTM model

We introduce a DL-driven ECG biometric system which takes the original denoised 1D ECG signals as input vectors and BLSTM as an important network structure unit. The suggested method combines dropout layers to reduce overfitting and deploys APSO to optimize the hyperparameters of DNN, so as to construct an ECG identification system with high recognition accuracy, fast recognition, and multi-scene application.

In dropout method A, 5% of the input connections of the BLSTM layer were dropped out. In dropout method B, 5% of the twice fully connected layers were dropped out. This is the most classical setting for dropout layers (Rodriguez and Laio, 2014). In addition, sufficient samples for each subject ensured the existence of topical diversity, prevented overfitting, and increased the thoroughness of

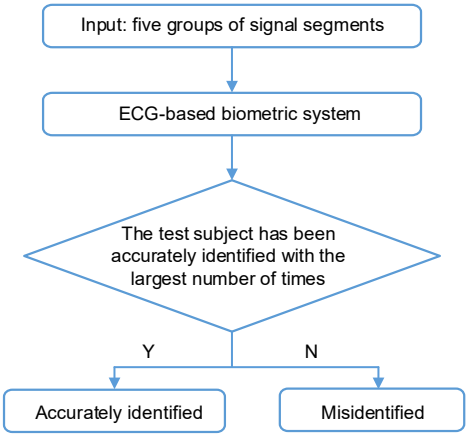

Fig. 4 Logical definition of the identification

Table 2 Overview of the four schemes proposed

Scheme	Number of layers	Type/Action	Note	Role
A	10	In-LSTM-LSTM-Fc-Dropout, 5%-Fc-Dropout, 5%-Fc-Sm-O	LSTM, dropout B	Performance of the BLSTM and LSTM layers
B	7	In-BLSTM-BLSTM-Fc-Fc-Sm-O	BLSTM, without dropout	Role of the dropout layer
C	9	In-BLSTM-Dropout, 5%-BLSTM-Dropout, 5%-Fc-Fc-Sm-O	BLSTM, dropout A	Influence of the dropout layer position
D	10	In-BLSTM-BLSTM-Fc-Dropout, 5%-Fc-Dropout, 5%-Fc-Sm-O	BLSTM, dropout B	–

Scheme D was selected as the best network structure based on our experiments. In: input; Fc: fully-connected; Sm: softmax; O: output

performance evaluation. Hence, we defined the data enhancement operation as follows: use the fixed window to randomly intercept 20 signal segments with the length of $3f_s$ and use them as training samples for each subject. It must be noted that this operation applies only for the training and optimization processes, and blind segmentation needs to be performed only five times to obtain five sets of vectors in practical applications.

Optimization and selection of networks: the network structure was optimized and screened using dataset D2 in protocol 2. Each subject's ECG signals successively underwent the two preprocessing steps of denoising and blind segmentation to obtain a feature vector with size $20\times3f_s$. Then schemes A to D, as presented in Table 2, were used for model training.

The training and validation accuracy curves for each scheme are presented in Figs. 5a–5d. It can be observed from Fig. 5b that the scheme without dropout layers was clearly overfitting beyond epoch 90, as the training accuracy started to plateau around 100% while the validation accuracy continued to rise slowly, with a gap distance of 10%. In contrast, the accuracy metric achieved by the dropout networks (Figs. 5c and 5d) was fairly stable. Both the validation and training curves increased at a steady rate and eventually almost met. The performance is summarized in Table 3. The deep network structure of scheme D was the best performing among the four training schemes, which proves the utility of the dropout layer and BLSTM layer. Thus, it was selected as the network model for our ECG biometric system. Through the hyperparameter optimization of APSO, the optimal values for batch size, initial learning rate, and number of hidden layer units were found to be 100, 0.0007, and 220, respectively. The detailed architecture is shown in the bottom right of Fig. 3.

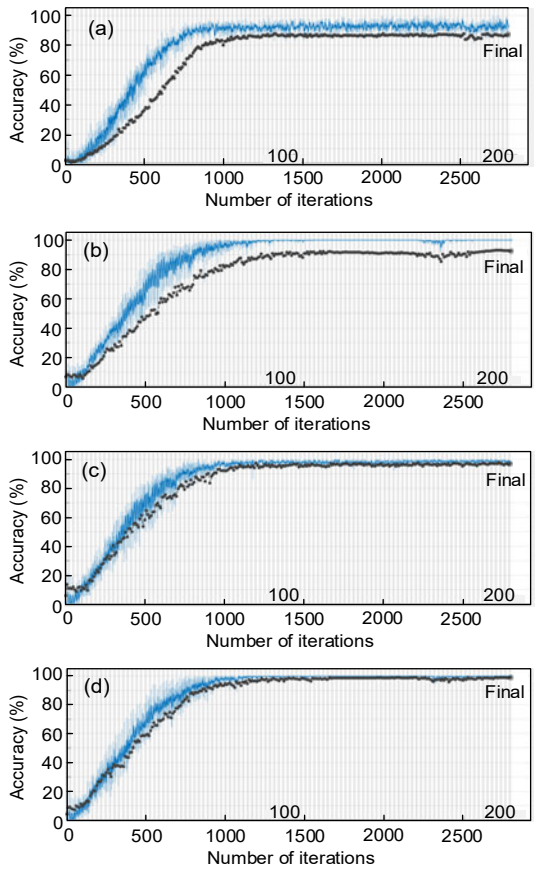

Fig. 5 Accuracy plots for the four schemes during model training (database D2 in protocol 2): (a) scheme A; (b) scheme B; (c) scheme C; (d) scheme D

5 Discussion

5.1 Performance of the proposed ECG-based biometric algorithm with different protocols

The proposed algorithm was thoroughly trained and evaluated on the two datasets with two protocols, simulating the influence of electrode placement and acquisition sessions in identification. The training and test results from these protocols are shown in Table 4.

1. Influence of different electrode placement. Analyzing the performance of D1 in two different protocols, it becomes evident that they have the same training and validation accuracy, but protocol 2 performed a little better than protocol 1 on the test set. This was mainly because the test set for protocol 1 was from lead II-collected ECG recordings, but the model had been trained only by lead I-collected ECG recordings.

The ECG recordings collected by different leads on the same individual are quite different. Fig. 6 shows the ECG signals obtained by three different collection methods, with visible differences in waveform amplitude and changing trend in the time domain. To a certain extent, these are such differences that directly lead to the different levels of intraclass similarity within different samples of the same individual. This results in a large recognition error for dataset D1, especially in protocol 1. Therefore, multi-

lead acquisition methods do affect the waveform amplitude of ECG signals and changes in time dimension to a certain extent.

In addition, it can be seen that after the training of a multi-lead collected ECG dataset, protocol 2 performed a little better. This can be explained as follows: ECG is an internal signal generated within the human body, and reflects the characteristic internal information and working state of the heart at different levels, and importantly, is unique. Even the ECG signals of twins are different. Therefore, the effect of multi-lead acquisition methods on their key characteristics is limited.

Overall, it remains to be seen whether the impact of ECG acquisition with different leads on their key characteristics is so serious as to directly mislead the performance of human identification, resulting in greater intraclass difference than the interclass similarity. This needs more experimental data, collection time, and theoretical basis to verify.

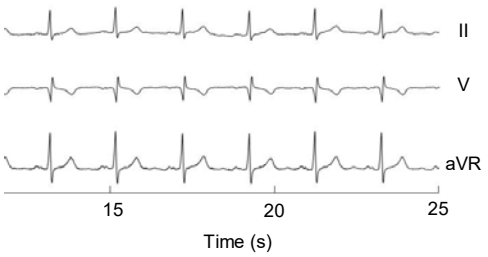

Fig. 6 ECG recordings collected by different leads of the same individual: lead II-collected ECG (upper); lead V-collected ECG (middle); lead aVR-collected ECG (bottom)

2. Influence of different acquisition sessions. Analyzing the performance of D2 in two different protocols, one can see that the accuracy from protocol 2 was indeed much higher than that from protocol 1, whether in training or test. In particular, the gap distance of accuracy between the two protocols on the test set is almost twice that of the training and validation set.

The recognition rate of 94.44% in protocol 1 is mainly due to the failures of individuals 1, 2, 59, 64, and 89. These subjects have a certain interclass similarity in their ECG signals (just like the faces of different subjects have certain similarities sometimes). In addition, if they have similar personal collection status in protocol 1 (including heart rate, exercise, and mood status), this can cause a certain degree of interclass similarity, resulting in misidentification. From protocol 1 to protocol 2, after multi-session acquisitions, this kind of accidental interclass similarity decreases, and the intraclass difference increases. That is, the interclass difference between these five subjects and others, and the intraclass similarity of themselves become obvious, allowing them to be correctly identified. Therefore, it is necessary to train the algorithm with multi-session databases. More importantly, in this study we use APSO to support the training of the DNN model. Through the learning and training of 20 groups of each individual with a total of 20 000 points, a BLSTM-based human

Table 3 Average classification performance for different schemes (database D2 in protocol 2)

Scheme	Training		Validation	Test	
	Accuracy (%)	EER (%)	Accuracy (%)	Accuracy (%)	Testing time (s)
A	92.36	6.94	88.67	94.44	2.8344
B	100.00	6.25	91.89	90.00	3.5124
C	98.13	6.02	97.00	97.78	2.8332
D	98.19	4.40	98.83	100.00	2.9620

Table 4 Recognition results from multiple databases in different protocols

Dataset	Protocol	Target	Number of subjects			Accuracy (%)		
			Training	Validation	Test	Training	Validation	Test
D1	1	Electrode placement	72	18	90	97.22	100.00	95.56
	2					**97.22**	**100.00**	**97.78**
D2	1	Acquisition sessions				95.56	95.83	94.44
	2					**98.19**	**98.83**	**100.00**

Each subject had 20 sets of signal segment samples in the training and validation processes. Protocol 2 was used in the experiments in this study

identification model is constructed, which can fully learn the change law in the time dimension and the memory characteristics of the classification algorithm on key features.

One thing we must admit, of course, is that a high test acc of 100% in protocol 2 does not mean that the proposed algorithm can achieve such high-precision recognition in any case of ECG-based human identification with any ECG signal acquisition. To a certain extent, this is very dependent on high-quality signal acquisition.

5.2 Comparison with classical machine learning/deep learning algorithms

Table 5 summarizes the comparative results of the deep APSO-BLSTM approach and three other methods. The first three DL-driven approaches performed slightly better than the traditional classifier SVM in the accuracy metric. To a great extent, this exemplifies the superiority of deep learning in self-processing with ECG signals. After comparing training time, we found that our proposed approach was significantly faster than CNN. One factor was the fact that LSTM processes 1D time series, as opposed to CNN, which processes 2D image sequences. Thus, the length of the original ECG sequence is greatly reduced, in turn reducing computing burden and speeding up the training process. In addition, the APSO algorithm used in our study optimized the model parameters quickly, which reduced the overall time required.

It does not mean that, of course, CNN is weaker than BLSTM in dealing with classification problems, but the application scope is different. CNN does have a very mature application in computer vision. However, the difference of different individuals' ECG waveform is not obvious in the time domain, and its main characteristics reflect the change law in the time dimension. In summary, both CNN and LSTM showed better learning ability in feature self-learning. Considering that rapid identification is a priority in practical applications, the proposed approach may have superior applicability compared to the other three methods.

5.3 Comparison with state-of-the-art methods

Over the years, many studies on ECG-based biometrics have been conducted using PhysioNet and multiple other private ECG databases. Tables 6 and 7 summarize the research that has been particularly prominent or representative in the last few years, to allow a more extensive and accurate comparative performance evaluation.

In this study, we developed a new non-fiducial method for biometric identification, combining APSO and BLSTM to learn feature representations directly from ECG time series. The recognition rates on the ECG-ID database and PhysioNet/CinC Challenge 2011 database were 100.00% and 97.78%, respectively. Compared to other studies with the same number of subjects (Salloum and Kuo, 2017; Bassiouni et al., 2018; Zhao et al., 2018; Chu et al., 2019), our method was more accurate on the ECG-ID dataset. Comparing the highest performance of the proposed method with the performance of the other algorithms in Table 6, it can be observed that our method achieved better performance, which further confirms that it can serve as an effective non-fiducial algorithm for ECG-based human biometric identification.

Similarly, the first four DL-driven approaches performed slightly better than the traditional classifier SVM in Table 7. The accuracy obtained by simple CNN is weaker than those by BLSTM and deep CNN with the same protocol. On one hand, it shows that with the deepening of the network, the learning ability of the model increases. On the other hand, it reflects

Table 5 Recognition accuracy of different methods (database D1-D2 in protocol 2)

Method	Testing accuracy (%)	Testing time (s)	Training time (s)
Proposed	97.78, 100.00 (**98.89**)	2.8951, 2.9620 (**2.9286**)	202.8461, 198.5893 (200.72)
Simple CNN	90.28, 95.83 (93.06)	3.7850, 3.2108 (3.4979)	483.5204, 432.4406 (457.9805)
Deep CNN	95.56, 100.00 (97.78)	5.1186, 4.7860 (4.9523)	More than one day
Machine learning classifier SVM	91.67, 92.22 (91.95)	2.0744, 2.1509 (2.1127)	18.9383, 17.2834 (**18.1109**)

Two data points in each grid represent the performance of D1–D2, and the value in the brackets represents the average performance of 90 testing samples. Simple CNN proposed in Zhao et al. (2018) using generalized S-transform to perform the time-frequency domain conversion. Deep CNN: original GoogLeNet using generalized S-transform to perform the time-frequency domain conversion. Machine learning classifier SVM: using the original 1D denoised ECG signal as input vectors

Table 6 Comparison with state-of-the-art algorithms (ECG-ID)

Reference	Algorithm	Dataset (N_{total})	Accuracy (%)
Salloum and Kuo (2017)	RNN	ECG-ID (90)	97.00
Yu et al. (2017)	PCA	ECG-ID (89)	96.60
Wu et al. (2018)	1D CNN	ECG-ID (89)	97.54
Bassiouni et al. (2018)	SVM+KNN	ECG-ID (90)	99.00
Zhao et al. (2018)	2D CNN	ECG-ID (90)	96.63
Chu et al. (2019)	1D CNN	ECG-ID (90)	98.24
This paper	1D APSO-BLSTM	ECG-ID (90)	100.00

N_{total} represents the total number of testees in the test set

Table 7 Comparison with state-of-the-art algorithms (PhysioNet/CinC Challenge 2011 database)

Algorithm	Accuracy (%)	
	Protocol 1	Protocol 2
Proposed	95.56	**97.78**
Simple CNN+GST	88.89	90.00
Deep CNN (GoogLeNet)+GST	**95.56**	**95.56**
Deep CNN (DenseNet)+Spectrograms (Nuno et al., 2020)	92.22	95.56
Machine learning classifier SVM	87.78	91.11

the advantage of BLSTM in processing time series.

Furthermore, in most DL-driven classification and recognition algorithms, model training and parameter optimization are time-consuming, labor-intensive, and have a high computational burden. In our suggested method, APSO is used to find the optimal values for three key hyperparameters of BLSTM (batch size, initial learning rate, and number of hidden layer units), and thus reduces the complexity of model training and improves the accuracy of model recognition. Considering the superior training time and computational resource consumption of the model, the deep APSO-BLSTM approach proposed in this study has demonstrated clear practical application value.

6 Conclusions

The study and improvement of ECG-based biometric technology has very important practical value for the development of modern identification technology, and is an important research direction in the field of intelligent health applications. The major concerns for ECG identification applications appear to be long-term stability and recognition speed. Therefore, solving the hyperparameter optimization problem of the existing DNN structure in model training, reducing the high computational complexity of traditional classification algorithms, and increasing the applicability of the model for different acquisition statuses, are very important topics. In this paper, we briefly summarized the existing problems in the application of ECG-based biometric systems and then focused on the key technology of a multi-scenario, DL-driven method based on deep BLSTM and APSO; finally, we applied it to biometric identification.

This paper introduces a novel DNN framework that integrates deep BLSTM and APSO, supplemented by original denoised ECG time series. Most previous research has used ECG as a whole to extract features or average the features of different windows after partitioning windows, ignoring the changes of ECG signals in the time dimension and the memory characteristics of recognition algorithms to key features. BLSTM performs forward and backward synchronization training on long-period signals, effectively identifying key features in time series. In addition, the PSO algorithm of the adaptive learning strategy is used to match ECG signal features with deep BLSTM network topology, achieving fast and effective parameter optimization, and the best recognition performance yet under the existing conditions.

Experimental results provide reference values for practical applications of the DL-driven ECG biometrics system. We would like to note that the concept of APSO used in a DNN could also be used in other research such as diagnosis and classification based on a 1D signal or a 2D image. Of course, the main limitation of this paper is that in order to avoid the influence of a subject's sample number on feature self-learning, this study involved 20 blind segmentations in the model training to provide 20 groups of samples for each subject. This calculation was computationally intensive and slow, and an interesting future direction could be to study the impact of different individual sample numbers. Another important research direction is to determine the impact of multi-lead acquisition methods on ECG-based human identification.

Contributors

Yefei ZHANG and Zhidong ZHAO designed the research. Yanjun DENG and Xiaohong ZHANG processed the data. Yefei ZHANG and Zhidong ZHAO drafted the manuscript. Yu ZHANG helped organize the manuscript. Yefei ZHANG and Zhidong ZHAO revised and finalized the paper.

Compliance with ethics guidelines

Yefei ZHANG, Zhidong ZHAO, Yanjun DENG, Xiaohong ZHANG, and Yu ZHANG declare that they have no conflict of interest.

Data availability

The data used in this work is publicly available from https://archive.physionet.org/challenge/2011/ and https://physionet.org/content/ecgiddb/1.0.0/.

References

Agrafioti F, Hatzinakos D, 2008. ECG based recognition using second order statistics. Proc 6th Annual Communication Networks and Services Research Conf, p.82-87. https://doi.org/10.1109/CNSR.2008.38

Ahmadi A, Mitchell E, Richter C, et al., 2015. Toward automatic activity classification and movement assessment during a sports training session. *IEEE Int Things J*, 2(1):23-32. https://doi.org/10.1109/JIOT.2014.2377238

Bassiouni MM, El-Dahshan ESA, Khalefa W, et al., 2018. Intelligent hybrid approaches for human ECG signals identification. *Signal Image Video Process*, 12(5):941-949. https://doi.org/10.1007/s11760-018-1237-5

Biel L, Pettersson O, Philipson L, et al., 2001. ECG analysis: a new approach in human identification. *IEEE Trans Instrum Meas*, 50(3):808-812. https://doi.org/10.1109/19.930458

Choi GH, Bak ES, Pan SB, 2019. User identification system using 2D resized spectrogram features of ECG. *IEEE Access*, 7:34862-34873. https://doi.org/10.1109/ACCESS.2019.2902870

Chu YF, Shen HB, Huang KJ, 2019. ECG authentication method based on parallel multi-scale one-dimensional residual network with center and margin loss. *IEEE Access*, 7:51598-51607. https://doi.org/10.1109/ACCESS.2019.2912519

da Silva Luz EJ, Moreira GJP, Oliveira LS, et al., 2018. Learning deep off-the-person heart biometrics representations. *IEEE Trans Inform Forens Secur*, 13(5):1258-1270. https://doi.org/10.1109/TIFS.2017.2784362

Hochreiter S, Schmidhuber J, 1997. Long short-term memory. *Neur Comput*, 9(8):1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

Labati RD, Muñoz E, Piuri V, et al., 2019. Deep-ECG: convolutional neural networks for ECG biometric recognition. *Patt Recogn Lett*, 126:78-85. https://doi.org/10.1016/j.patrec.2018.03.028

Liu JK, Yin LY, He CG, et al., 2018. A multiscale autoregressive model-based electrocardiogram identification method. *IEEE Access*, 6:18251-18263. https://doi.org/10.1109/ACCESS.2018.2820684

Nuno B, Belo D, Gamboa H, 2020. ECG biometrics using spectrograms and deep neural networks. *Int J Mach Learn Comput*, 10(2):259-264. https://doi.org/10.18178/ijmlc.2020.10.2.929

Oh SL, Ng EYK, Tan RS, et al., 2018. Automated diagnosis of arrhythmia using combination of CNN and LSTM techniques with variable length heart beats. *Comput Biol Med*, 102:278-287. https://doi.org/10.1016/j.compbiomed.2018.06.002

Palaniappan R, Krishnan SM, 2004. Identifying individuals using ECG beats. Proc Int Conf on Signal Processing and Communications, p.569-572. https://doi.org/10.1109/SPCOM.2004.1458524

Pan JP, Tompkins WJ, 1985. A real-time QRS detection algorithm. *IEEE Trans Biomed Eng*, 32(3):230-236. https://doi.org/10.1109/TBME.1985.325532

Rodriguez A, Laio A, 2014. Clustering by fast search and find of density peaks. *Science*, 344(6191):1492-1496. https://doi.org/10.1126/science.1242072

Salloum R, Kuo CCJ, 2017. ECG-based biometrics using recurrent neural networks. Proc IEEE Int Conf on Acoustics, Speech and Signal Processing, p.2062-2066. https://doi.org/10.1109/ICASSP.2017.7952519

Srivastava N, Hinton G, Krizhevsky A, et al., 2014. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res*, 15(1):1929-1958.

Tantawi MM, Revett K, Salem AB, et al., 2015. A wavelet feature extraction method for electrocardiogram (ECG)-based biometric recognition. *Signal Image Video Process*, 9(6):1271-1280. https://doi.org/10.1007/s11760-013-0568-5

Wu B, Yang GP, Yang L, et al., 2018. Robust ECG biometrics using two-stage model. Proc 24th Int Conf on Pattern Recognition, p.1062-1067. https://doi.org/10.1109/ICPR.2018.8545285

Yildirim Ö, 2018. A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification. *Comput Biol Med*, 96:189-202. https://doi.org/10.1016/j.compbiomed.2018.03.016

Yu JR, Si YJ, Liu X, 2017. ECG identification based on PCA-RPROP. Proc 8th Int Conf on Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management, p.419-432. https://doi.org/10.1007/978-3-319-58466-9_37

Zhang QX, Zhou D, Zeng X, 2017. HeartID: a multiresolution convolutional neural network for ECG-based biometric human identification in smart health applications. *IEEE Access*, 5:11805-11816. https://doi.org/10.1109/ACCESS.2017.2707460

Zhao ZD, Zhang YF, Deng YJ, et al., 2018. ECG authentication system design incorporating a convolutional neural network and generalized S-transformation. *Comput Biol Med*, 102:168-179. https://doi.org/10.1016/j.compbiomed.2018.09.027