

Received May 11, 2017, accepted May 19, 2017, date of publication May 24, 2017, date of current version July 17, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2707460

HeartID: A Multiresolution Convolutional Neural Network for ECG-Based Biometric Human Identification in Smart Health Applications

QINGXUE ZHANG¹, (Member, IEEE), DIAN ZHOU^{1,2}, (Senior Member, IEEE), AND XUAN ZENG², (Member, IEEE)

¹Department of Electrical Engineering, The University of Texas at Dallas, Richardson, TX 75080 USA

²Fudan University, Shanghai 200433, China

Corresponding author: Qingxue Zhang (qingxue.zhg@gmail.com)

This work was supported in part by the Recruitment Program of Global Experts (the Thousand Talents Plan) and in part by the National Natural Science Foundation of China under Project 61574044, Project 61376040, and Project 61574046.

ABSTRACT Body area networks, including smart sensors, are widely reshaping health applications in the new era of smart cities. To meet increasing security and privacy requirements, physiological signal-based biometric human identification is gaining tremendous attention. This paper focuses on two major impediments: the signal processing technique is usually both complicated and data-dependent and the feature engineering is time-consuming and can fit only specific datasets. To enable a data-independent and highly generalizable signal processing and feature learning process, a novel wavelet domain multiresolution convolutional neural network is proposed. Specifically, it allows for blindly selecting a physiological signal segment for identification purpose, avoiding the complicated signal fiducial characteristics extraction process. To enrich the data representation, the random chosen signal segment is then transformed to the wavelet domain, where multiresolution time-frequency representation is achieved. An auto-correlation operation is applied to the transformed data to remove the phase difference as the result of the blind segmentation operation. Afterward, a multiresolution 1-D-convolutional neural network (1-D-CNN) is introduced to automatically learn the intrinsic hierarchical features from the wavelet domain raw data without data-dependent and heavy feature engineering, and perform the user identification task. The effectiveness of the proposed algorithm is thoroughly evaluated on eight electrocardiogram datasets with diverse behaviors, such as with or without severe heart diseases, and with different sensor placement methods. Our evaluation is much more extensive than the state-of-the-art works, and an average identification rate of 93.5% is achieved. The proposed multiresolution 1-D-CNN algorithm can effectively identify human subjects, even from randomly selected signal segments and without heavy feature engineering. This paper is expected to demonstrate the feasibility and effectiveness of applying the blind signal processing and deep learning techniques to biometric human identification, to enable a low algorithm engineering effort and also a high generalization ability.

INDEX TERMS ECG, wavelet transformation, convolutional neural network, deep learning, machine learning, feature learning, blind signal processing, data representation.

I. INTRODUCTION

Body sensor networks are reshaping people's daily lives, especially in smart health applications [1], [2]. In a body sensor network, vital signs such as heart rate, temperature and activity can be continuously tracked by wearable computers and relayed by personal devices to the cloud infrastructure, for both real-time health management and long-term health statistics purposes. Physician instructions and medicine record [3] can also be accessed by or sent to individuals via the network. However, along with boosting

smart health sensors and applications, there are also increasing security and privacy requirements/challenges, to enable confidential biomedicine solutions, protect sensitive patient data, etc. Biometric-based human identification (ID) is a promising technology for automatic and accurate individual recognition, focusing on these challenges.

Leveraging the uniqueness and permanence to individuals, biometric characteristics are more reliable than traditional token-based and knowledge-based individual recognition methods, such as the identity card and the username/password

pair which may be lost or stolen [4]. Two major categories of biometrics have been widely studied, i.e., behavioral and physiological ones, typical examples of which include gait, voice, signature, fingerprint, retina and face. Among physiological biometrics, bio-potentials are emerging powerful modalities and playing a more and more important role in human identification, benefitting from a fast progress in both ubiquitous wearable devices and advanced signal processing/machine learning techniques. In this paper, we take special interest in the electrocardiogram (ECG) bio-potential, which is of many attractive characteristics for human identification applications, including universal, easily measured, unique and permanent [5]. Specifically, the sinus node in the heart modulated by both sympathetic and parasympathetic nerves, repeatedly produces electrical impulses and triggers the heart rhythm. Then the unique electrical waves, i.e., the ECG signal, are spread throughout the body and can be easily acquired with the ECG electrodes, either contacting or non-contacting ones [6]. These attractive signal characteristics and the unobtrusive measurement mechanism make the ECG biometric highly promising in terms of the human identification.

Many investigations have been reported for ECG biometric identification, usually based on two strategies, i.e., fiducial and non-fiducial methods. The former one extracts features based on the heartbeat characteristic points, such as the wave onset, peak amplitude, point-to-point interval, etc [7], [8]. These approaches usually heavily relies on robust heartbeat segmentation and fiducial point detection techniques, and also much manual feature engineering effort. Therefore, their generalization ability may be limited. For example, the heartbeat segmentation approaches and features extracted based on normal ECG datasets may not well fit the abnormal ones which usually have highly different heartbeat morphologies and variabilities, and vice versa [9]. Moreover, the features evaluated on the chest electrode placement may not perform well on other placements which may induce significant morphological changes [10]. On the other hand, to improve the generalization ability and to lower the feature engineering effort, some studies have introduced a non-fiducial strategy. Tantawi *et al.* [11] extracted features in the wavelet domain, however, their method still needs to firstly segment the ECG stream based on the detected R peaks (heartbeat locations) and thus suffer from some similar limitations, i.e., the R peak detection and ECG segmentation approaches which are effective for one dataset or sensor placement method may not be well generalized to other datasets or placements.

Convolutional neural network (CNN), as one of the major deep learning algorithms, is now gaining tremendous attentions leveraging its powerfullness in automatically learning the intrinsic patterns from the data, which can both prevent time-consuming manual feature engineering and capture hidden intrinsic patterns more effectively [12]. Inspired by biological process of the visual cortex, CNN consists of multiple layers, each of which owns a small neuron collection to process portions of the input image. These collections are

tiled to introduce region overlap, and the process is repeated layer by layer to achieve a high level abstraction of the original image. Inspired by the obserstation that the ECG stream can be seen as a 1D-image, we thus explore how to effectively apply the 1D-CNN to ECG biometric identification, to avoid heavy feature engineering efforts, and also let the CNN capture more hidden patterns from data and learn a high level abstraction.

In this paper, we propose a novel wavelet domain multiresolution convolutional neural network approach (MCNN) for ECG biometric identification as shown in Fig. 1, which avoids data-dependent complicated heartbeat detection/segmentation techniques and heavy manual feature engineering that are both time-consuming and of a limited generalization ability. Specifically, it allows for blind segmentation of both normal and abnormal ECG streams (i.e., we can randomly select an ECG segment for user identification purpose), provides a multiresolution data representation in the wavelet domain to achieve richer temporal and spectral characteristics, and leverages the self-learning ability of CNN to automatically adapt its internal parameters (i.e., features encoded in network parameters) to wavelet-domain raw data. For algorithm evalutation, a one-lead ECG configuration is chosen, considering it is more convenient than the multi-lead ECG configuration in daily applications, and of course, it also poses more challenges to the identification algorithm. Moreover, to demonstrate the generalization ability of the proposed framework composed of blind segmentation, data representation enrichment, phase difference removal, parallel multiresolution feature self-learning and classification, eight diverse datasets are considered which include not only different electrode placement methods (chest and wrist) but also various heart health conditions (with and without cardiac abnormalities), which are much more challenging than other works [13]–[17].

II. METHODS

A. SYSTEM OVERVIEW

The system diagram of the proposed approach is shown in Fig. 1, including pre-processing, wavelet transform, auto-correlation, component selection and parallel 1D-CNN. This section gives detailed description of our approach according to the signal processing flow.

Specifically, the ECG stream is firstly blindly split up into signal segments with an equal length of two seconds without leveraging any heartbeat location information, which is not only immune to diverse morphological/beat-to-beat interval variabilities, but also tolerant to signal artifacts that are usually major challenges in non-blind segmentation approaches [9], [18]. Afterwards, the ECG segments are transformed to the wavelet domain which is expected to reveal more detailed time and frequency characteristics in multiple resolutions than the original time domain [19]. Then the auto-correlation operation is performed to each wavelet component to remove the blind-segmentation-induced phase

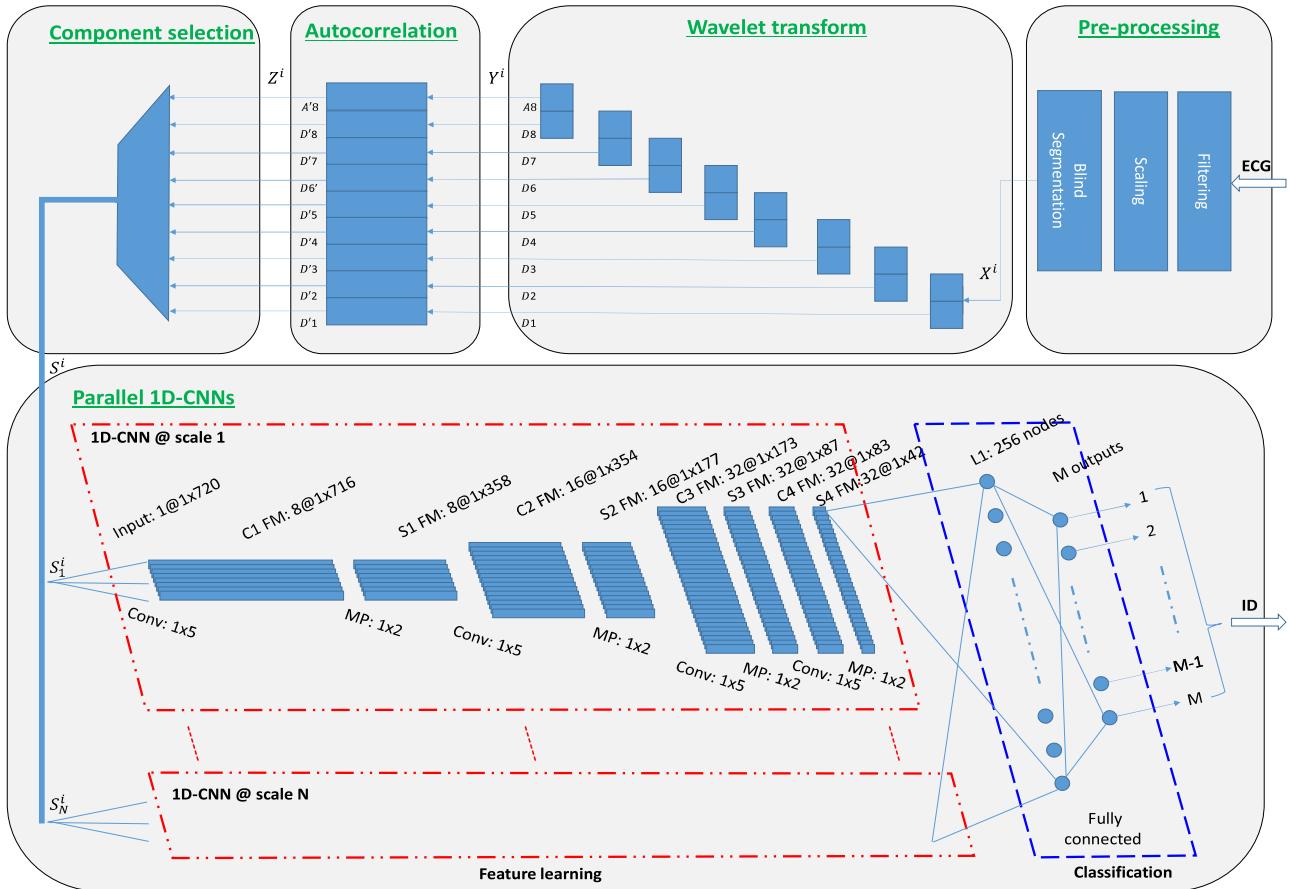


FIGURE 1. The system diagram of the proposed multiresolution convolutional neural network for human identification with blindly segmented ECG signal. Notes. CNN: convolutional neural network; FM: feature map; Conv: convolution; MP: max pooling; C1 FM: convolutional layer 1 feature map; S1 FM: stage 1 final feature map; ID: identification; definitions of variables are given in the ‘methods’ section.

difference. Finally, based on the enriched data representation, a 1D-CNN is applied to each wavelet component to learn the intrinsic patterns automatically, which allows for parallel feature self-learning in various wavelet scales, avoiding time-consuming manual feature engineering. The learned features reflected by the CNN internal parameters are then used to identify users on the unseen random ECG segments.

B. DATASETS

Eight datasets with diverse ECG behaviors have been considered including CEBSDB [20], [21], WECG [9], [18], FANTASIA [22], NSRDB [23], STDB [24], MITDB [25], AFDB [26], VFDB [27] as shown in Table 1 and Fig. 2. These datasets may be acquired by different lead configurations. Moreover, the first four datasets were collected from healthy or quasi-healthy participants, and last four include severe heart diseases such as ST depression/elevation, arrhythmia, atrial fibrillation and malignant ventricular ectopy. Considering these datasets were not acquired with the same sampling rate, all the ECG recordings were resampled to 360 Hz to fairly illustrate the performance.

C. PRE-PROCESSING

The pre-processing operation includes three steps, i.e., filtering, scaling and blind-segmentation, as shown in Fig. 1. Firstly, a 6-order Butterworth bandpass (2-50 Hz) filter is applied to each ECG recording to remove the baseline wander and the powerline interference. Then all the recordings are scaled to be between 0 and 1 and subtracted by their mean to balance their contribution in the algorithm training phase, as $(1-2)$ where ECG and ECG^s are the original and the new ECG stream, respectively. Afterwards, the filtered ECG recording is blindly segmented to ECG windows X^i with an equal length where i is the window index. The window length is chosen as 2-second (720 samples) to include as least one heartbeat, since the typical range of heart rate is from 40 to 208 beats per minute [28]. For each recording, 500 random windows are chosen, half of which are used to train the CNN (also the component selection step in Fig. 1) and another half for testing. An example of randomly chosen ECG windows are shown in Fig. 2, which usually include different number of heartbeats and highly different signal morphologies (either normal or abnormal). It is clear that the blind segmentation strategy can effectively avoid data-specific complicated

TABLE 1. Eight ECG datasets.

Abbreviations	# of subjects	Electrode placement	Type of abnormalities
CEBSDB	20	Lead I	Health
WECG	22	Lead I	Health
FANTASIA	40	Not specified	Health
NSRDB	18	Not specified	Quasi-health (no significant arrhythmias)
STDB	28	Not specified	ST depression/elevation
MITDB	47	MLII (modified limb lead II)	Arrhythmia (along with 18 kinds of other diseases)
AFDB	23	Not specified	Atrial fibrillation
VFDB	22	Not specified	Malignant ventricular ectopy

Notes. CEBSDB: Combined measurement of ECG, breathing and seismocardiogram; WECG: Wrist-ECG measurement; NSRDB: MIT-BIH Normal Sinus Rhythm Database; STDB: MIT-BIH ST Database; MITDB: MIT-BIH Arrhythmia Database; AFDB: MIT-BIH Atrial Fibrillation Database; VFDB: MIT-BIH Malignant Ventricular Arrhythmia Database.

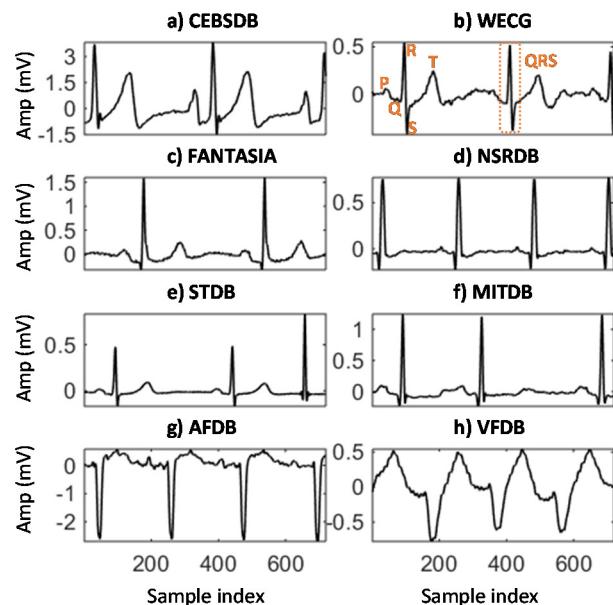


FIGURE 2. Blinely chosen ECG segments with diverse behaviors from eight normal and abnormal datasets. Amp: amplitude; P/Q/R/S/T: characteristic points of a normal heartbeat; QRS: central part of a heartbeat; definition of abbreviations is given in Table 1.

heartbeat identification and segmentation techniques, but at the same time, also introduces a high variability to the ECG windows (number of heartbeats, onset of the segment, etc.) and poses a big challenge to the following data representation and machine learning algorithms.

$$ECG^s = \frac{(ECG - \min(ECG))}{(\max(ECG) - \min(ECG))} \quad (1)$$

$$ECG^s = ECG^s - \text{mean}(ECG^s) \quad (2)$$

D. WAVELET TRANSFORM

To enrich signal characteristics, the wavelet transform (WT) [19] approach is applied to each 2-second window to obtain a multiresolution time-frequency representation for the raw ECG signal. Compared with traditional fast Fourier transform (FFT) and short-term Fourier transform (STFT) approaches, WT introduces a new strategy to reveal the spectral character of signals [29]. It is known that FFT can only give the frequency spectrum of the whole time series, which is enough for stationary signals but does not adequately represent the time-varying modes of non-stationary signals such as ECG. To support non-stationary signals, STFT introduces the windowing tools to analyze only a small section of the signal at a time. However, due to the unchanged window, it is limited by both the dilemma of resolution and the Heisenberg uncertainty principle [30]. The former one means that it suffers from a poor frequency resolution with a narrow window or a poor temporal resolution with a wide window, and the latter one indicates that it cannot give what frequency exists as what time interval.

An alternative approach to STFT to overcome the above-mentioned limitations is WT, which analyzes the signal at different frequencies with different resolutions [19]. Specifically, it can overcome the dilemma of resolution after introducing different window sizes for different frequency components, and thus provide a good temporal resolution at high frequencies and a good frequency resolution at low frequencies. Another contribution in overcoming Heisenberg uncertainty principle is that it can detect the sharp changes in spectral character, which is often the cases in ECG signals, such as the sharp R peaks and other shape changes due to the heart diseases, as shown in Fig. 2.

A computation-effective algorithm of the WT is discrete wavelet transform (DWT), which has a recursive structure as shown in Fig. 1. Here, we choose an 8-level DWT using the Daubechies6 wavelet as the mother wavelet [19]. The corresponding basis functions or baby wavelets are defined as (3), where j and k are the scale factor and translation factor, respectively, a_0 is chosen to be 2, and $\Psi(t)$ represents the mother wavelet. For the first level, the 2-second ECG window is decomposed to two parts, i.e., the low frequency part (approximation) and the high frequency part (detail). Afterwards, the approximation of each level is further decomposed to the approximation and the detail parts until the last level. Therefore, the whole process of DWT in Fig. 1 is decomposing the time series to a set of wavelet components Y^i defined in (4), where i is the index of the blindly segmented ECG time window, Y_1^i to Y_{C-1}^i correspond to eight details $D1$ to $D8$, Y_C^i represents the approximation $A8$, and C is equal to 9 here. Y^i now can reveal more time-frequency patterns in multi-frequency bands with different resolutions which are expected to be learned by the CNN.

$$\Psi_{j,k}(t) = a_0^{j/2} \Psi(a_0^j t - k) \quad (3)$$

$$Y^i = \{Y_1^i, \dots, Y_{C-1}^i, Y_C^i\} = \{D1, \dots, D8, A8\} \quad (4)$$

E. AUTO-CORRELATION

The auto-correlation operation is applied to the ECG windows in the wavelet domain, to remove the phase difference induced by blind segmentation and thus provide a shift invariant multiresolution data representation, as shown in Fig. 1 and defined as (5), where $Z_j^i[t]$ is the $t - th$ sample in the $j - th$ wavelet component of the $i - th$ ECG window after auto-correlation, $Y_j^i[m + t]$ corresponds to $Y_j^i[t]$ with a time lag of m , m is chosen from 0 to $T - t - 1$, and T , W and C correspond to the number of samples in an ECG window ($=720$), the number of ECG windows ($=500$) and the number of wavelet components ($=9$), respectively.

$$Z_j^i[t] = \sum_{m=0}^{T-t-1} Y_j^i[m]Y_j^i[m+t], \quad \forall t \in [0, T-1], \quad \forall i \in [1, W], \quad \forall j \in [1, C] \quad (5)$$

F. COMPONENT SELECTION

Another thing worth noting is that, although DWT is expected to reveal more time-frequency patterns from the ECG signals, the contribution of each decomposed component may be different. For example, for QRS complexes as shown in Fig. 2, most of the energy is concentrated in the 3rd/4th/5th detail components [19]. Therefore, to select out the optimal combination of wavelet components after auto-correlation, we have evaluated thirteen wavelet options, named as wv1-13, including Z_{1-2} , Z_{1-4} , Z_{1-6} , Z_{1-8} , Z_{7-8} , Z_{5-8} , Z_{3-8} , Z_{7-9} , Z_{5-9} , Z_{3-9} , Z_{3-6} , Z_{4-5} and Z_{3-5} . The superscript in Z_j^i is omitted here for a good readability. The set of selected components is denoted as S^i given in (6), where N is the number of selected components. The component selection is performed on the CEBSDB dataset by maximizing the human identification rate on the testing data. The selected components are then applied to other seven datasets to evaluate the generalization ability.

$$S^i = \{S_n^i | n = 1, \dots, N\} \quad (6)$$

G. CNN

CNN is a powerful deep neural network inspired by visual neuroscience, which was firstly used in the area of computer vision and has already produced extremely promising progress [12]. Given a 2-D input image, CNN can effectively learn the hierarchical features to construct a final feature set of a high level abstraction, which can then be fed into a simpler classifier such as a fully connected neural network (NN) or a support vector machine (SVM) for category identification purpose. The patterns in the images are automatically learned by CNN and stored in the parameters of network connections, so CNN requires very little manual engineering. Moreover, CNN is more capable of discovering intricate patterns in high-dimensional data compared with manual feature engineering.

There are several key considerations behind CNN, such as local connections, shared weights, pooling operations and dropout techniques [31]. As shown in Fig. 1, each CNN stage

is composed of two types of layers, i.e., the convolutional and pooling layers. The convolutional layer includes many feature maps to extract a higher level representation from the previous layer. Each feature map has many units, each of which is connected to local patches of units in the feature maps in the previous layer. The connection between a unit and a local patch in the previous layer is called a filter bank, which performs a discrete convolution operation. So the value of each unit is actually a local weighted sum of the previous patch, which is then fed into a non-linear activation function to determine whether this unit (neuron) fires or not. A same filter bank is shared by all units within a feature map, not only to form distinguishable local motifs from locally correlated values, but also to invariantly detect a same pattern even appearing in different locations.

A pooling layer is further introduced in each stage to merge similar features [32], which can reduce the feature dimension and also deal with some motif variations due to small signal variability (shifts and distortions). A widely used max pooling operation is chosen here, which captures the maximum of the corresponding local patch in the convolutional layer of the same stage. To further regularize the large number of parameters, the dropout technique is introduced which randomly ignores some neurons during training [31]. This operation can suppress the specialization of neighboring neurons which may result in a fragile model overfitted to the training data (too smart in learning data), by forcing other neurons to step in and handle some more by their own (so they are also less dependent on the nodes they are connected to). In such a way, the network is more insensitive to the specific parameters of neurons (prevent neurons from co-adapting too much) and owns a better generalization ability to the unseen fresh data.

To train the multi-layer CNN, the backpropagation approach is usually used [33], which computes the gradient of a predefined objective function with respect to all the neuron parameters by applying the chain rule for derivatives. The gradients can be propagated backwards from the output layer to the input layer, to adjust the parameters such that the network can converge to a state to be able to encode the training patterns.

Leveraging key ideas in CNN architecture establishment and training techniques mentioned above, the neuron units are well organized in hierarchical features maps and can provide gradually increasing level of abstraction to enable the final identification task.

H. MULTISCALE 1D-CNN

The data representation in the wavelet domain is mainly composed of signal components at different frequency resolutions, meaning that most signal characteristics are reflected by intra-component patterns, not inter-component behaviors. Therefore, we apply 1D-CNN to each wavelet component after auto-correlation processing. In such a case, each signal sequence S_n^i can be seen as a 1D-image (single-row image), and 1D-CNN can still effectively learn underlying patterns by

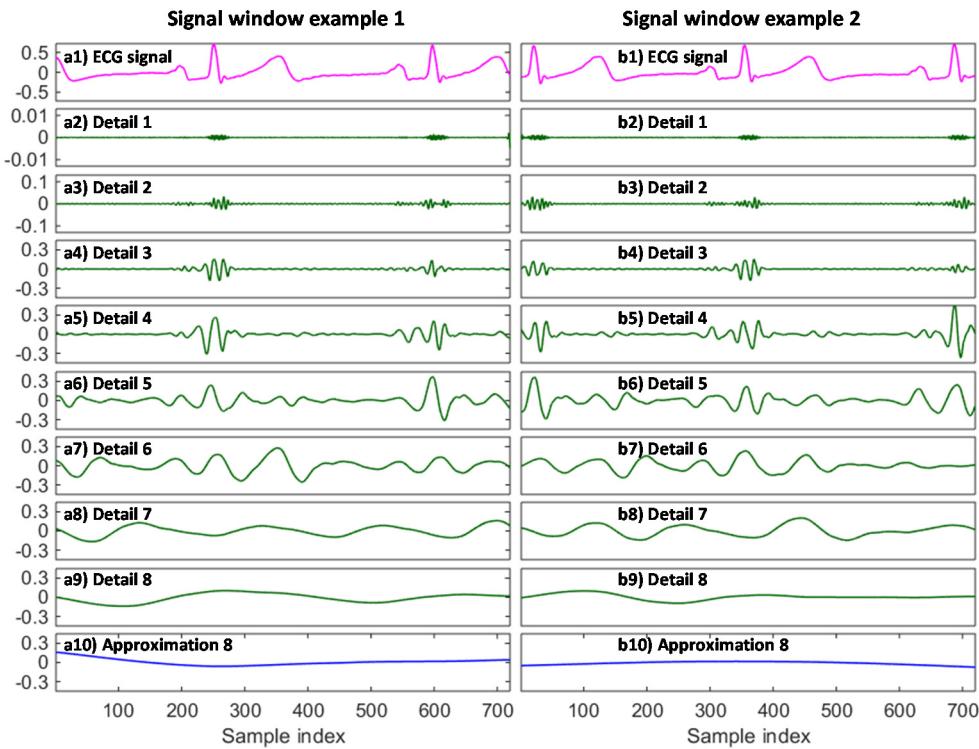


FIGURE 3. Multiresolution representation of two signal segments in the wavelet domain.

performing convolutional and pooling operations similar to 2D-CNN. We name this topology as parallel 1D-CNN or multiscale 1D-CNN, which includes several 1D-CNN in parallel to learn data presentation at different scales, as shown in Fig. 1.

The CNN topology is usually hard to be pre-selected [34]. Therefore, to explore effective topologies of parallel 1D-CNN, we have taken into account eight possibilities, named as *CNN1-8*, the former four corresponding to CNN with one to four stages without dropout regularization, and the last four with dropout operation (25% dropout). Specifically, the filter bank used to generate the convolutional layer is chosen as 1×5 , and the max pooling matrix is chosen as 1×2 . The firstly stage includes a convolution layer of eight 1×716 feature maps, and a max pooling layer of eight 1×358 feature maps. Other stages are also given in Fig. 1 with all details. The learned feature maps are finally fed to a fully connected neural network with an input layer of 256 nodes and an output layer of M nodes (M human subjects). The activation function is chosen as ‘tanh’, except the output layer which uses ‘softmax’ function to represent a categorical probability distribution. CNN topology selection is also performed on the CEBSDB dataset by maximizing the human identification rate on the testing data. The selected topology is then applied to other seven datasets to evaluate the generalization ability.

To speed up the training procedure which is usually a bottleneck when using a deep learner with multiple layers, we have implemented the algorithm using Keras deep

learning library written in Python and running on top of Theano [35], which can be executed on Graphics Processing Unit (GPU), to handle vector and matrix operations in parallel. GPU usually brings at least 5x to 10x speedups compared with Central Processing Unit (CPU) and thus can substantially accelerate the training process, and we have chosen GeForce GTX 960M GPU, empowered by a parallel computing platform and programming model ‘CUDA’ invented by NVIDIA [36], [37]. The CUDA Deep Neural Network library (CuDNN 5005) is chosen here.

The trained parallel 1D-CNN is then applied to the unseen testing data for human identification purpose, leveraging the self-learned multiresolution hierarchical features in the wavelet domain. The performance is reported in terms of the confusion matrix and identification rate.

III. RESULTS AND DISCUSSION

In this section, detailed experimental results and discussion are given according to the signal processing flow as shown in Fig. 1.

A. WAVELET TRANSFORM

The wavelet transformation is expected to provide a richer data representation in the wavelet domain. Two examples of the transformed signals are given in Fig. 3. On the left part, the ECG signal is decomposed to eight details and one approximation. The detail 1 is extracted using the baby wavelet of the highest frequency, and the detail 8 is generated by the baby wavelet of the lowest frequency.

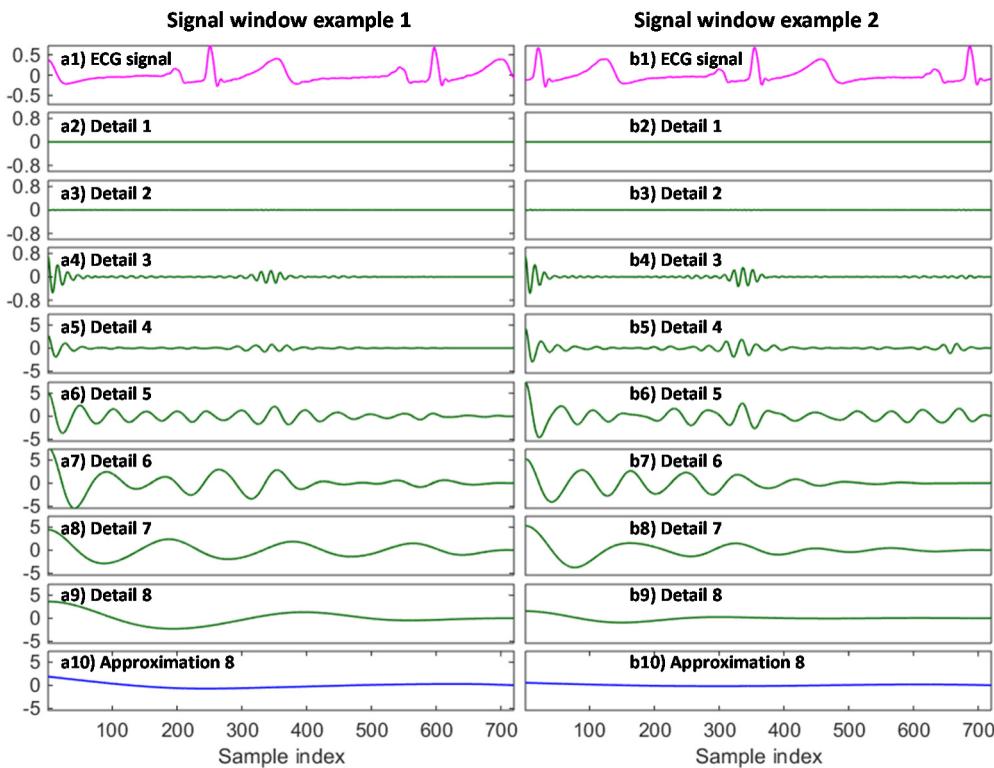


FIGURE 4. Phase different removal by auto-correlation to enable blind segmentation of ECG windows.

As mentioned above, the wavelet transformation owns two advantages compared with STFT. Firstly, it overcomes the dilemma of resolution faced by STFT, i.e., a fixed small window results in a poor frequency resolution and a fixed large window causes a low temporal resolution. DWT, instead, applies windows of different sizes at different frequency levels. Specifically, a small window is chosen at a high frequency level to achieve a high temporal resolution such as detail 1, which is based on the consideration that high frequency component usually makes the signal quickly fluctuate and thus requires a high temporal resolution to track the signal dynamics. Moreover, a gradually extended window is applied to extract signal characteristics at a gradually decreasing frequency level. Take the detail 8 as an instance, the low frequency component makes the signal change slowly and thus a large window is applied to get a big picture of the signal in order to guarantee a high frequency resolution. In such a way, a time-frequency data representation of richer signal characteristics is obtained in the wavelet domain, taking into account different resolution requirements at different frequency levels.

Secondly, DWT can overcome Heisenberg uncertainty principle and thus can effectively capture the sharp signal changes in spectral manner. As shown in example 1, details 3 to 5 in the wavelet domain own distinguishable signal fluctuations at different frequency levels, each corresponding to part of the information belonging to the QRS complexes in the original time domain. One thing worth noting is that

other details and approximation may also own more or less useful information unique to individuals, which is the reason we have considered 13 wavelet topology options to select out the best data representation method.

Another thing worth noting is that we have randomly selected the ECG segments for user identification purpose, in order to avoid heavy engineering and limited generalization ability of data-dependent heartbeat identification and segmentation methods. However, this blind segmentation also brings more variability to the ECG segments, as shown in example 1 and 2, which have different numbers and occurrence time of heartbeats. This variability will be further processed by the next auto-correlation step.

B. AUTO-CORRELATION

The autocorrelation operation is introduced to remove the phase difference due to blind segmentation. Fig. 4 shows similar outputs when applying auto-correlation to two wavelet domain signal segments given in Fig. 3. The auto-correlation calculates the correlation of a series with its delayed copy, i.e., the similarity between series as a function of the time lag between them. Therefore, it can effectively discover repeating patterns in the quasi-periodic ECG signals even with different numbers and occurrence time of heartbeats. After removing the phase difference, the multiresolution data can now be fed to the parallel 1D-CNN for automatic feature learning and user identification purpose.

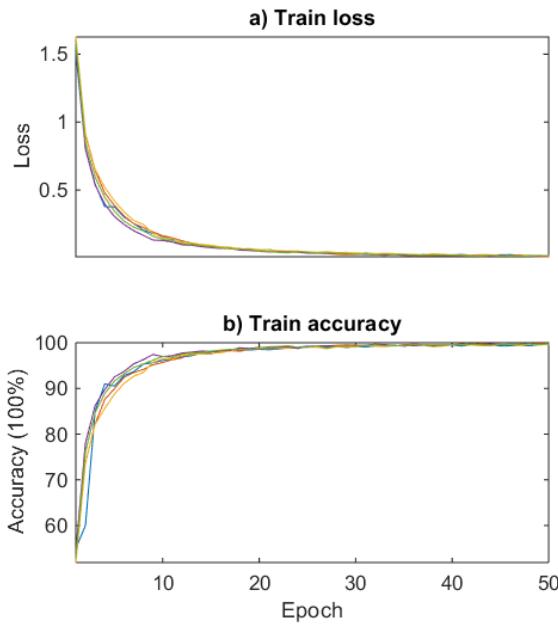


FIGURE 5. Training process of the parallel 1D-CNN network.

C. MULTISCALE 1D-CNN

The multiscale 1D-CNN is firstly trained on the data with ground truth identify labels and then tested on the unseen fresh data. In the training phase, the network self-learns hierarchical features by convolutional and pooling operations from pairs of data representation and user label. An example of the training process is given in Fig. 5, where the top part shows the gradually decreasing training loss and the bottom part corresponds to the increasing training accuracy. Besides GPG-based acceleration techniques, a mini-batch SGD training strategy is chosen to further accelerate the training process [31], which allows for passing a subset of training data to the neural network each time. The mini-batch size is chosen as 100 to trade off two considerations, i.e., a large size results in a short convergence time by reducing the variance of stochastic gradient updates, and a small size brings more power for SGD to jump out of shallow minima in the error function. The epoch size is set as 50 to balance underfitting and over-fitting considerations. Actually, the network can already effectively learn most of the underlying patterns of the wavelet domain data and basically converges around 25 epochs. The learned hierarchical features encoded in the neuron connection parameters are then used to predict the user label on the testing data. The testing performance will be given later.

It is worth noting that we have trained and tested the multiscale 1D-CNN model on each dataset both for five times to average the performance, consideration that the learned hierarchical features are of some randomness resulting from the stochastic gradient descent optimization approach [38]. Fig. 5 shows the training process for all five trials and interestingly we can find that they own a similar convergence speed. This is consistent with the theoretical study that poor local

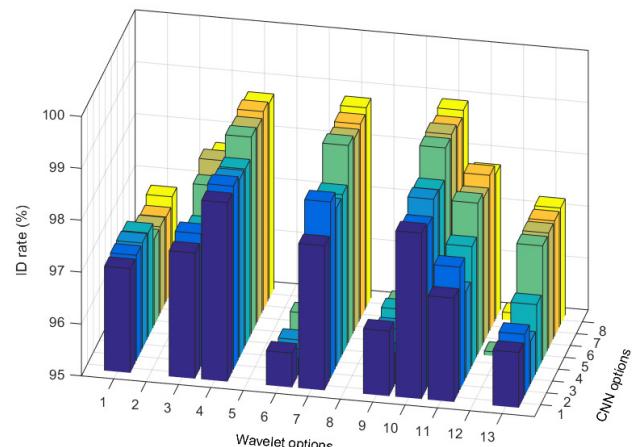


FIGURE 6. Wavelet components and CNN topology evaluation and selection. ID rate: identification rate.

minima are rarely a problem in deep neural networks with a large number of parameters. Instead, the landscape of the object function is packed with a large amount of valleys which seems to mostly have local minima with similar values [31]. Therefore, the randomness in SGD-based parameter tuning process actually often results in only small fluctuations to the convergence curve in the training process.

D. TOPOLOGY SELECTION

To determine appropriate wavelet and parallel 1D-CNN topologies, we have evaluated 13 wavelet options and 8 CNN options on the CEBSDB dataset, as shown in Fig. 6, which gives the identification rate over the testing data for 13x8 combination possibilities. Wavelet options 1 to 13 correspond to Z_{1-2} , Z_{1-4} , Z_{1-6} , Z_{1-8} , Z_{7-8} , Z_{5-8} , Z_{3-8} , Z_{7-9} , Z_{5-9} , Z_{3-9} , Z_{3-6} , Z_{4-5} and Z_{3-5} , respectively. CNN options 1 to 8 mean CNN with one to four stages without dropout regularization, and with dropout operations, respectively. There are several interesting observations here. Firstly, different wavelet topologies own different time-frequency representation ability, and thus yield different performance. Wavelet options 4, 7 and 10 correspond to relatively higher performance compared with other options, benefitting from more information fed to the parallel 1D-CNN. This indicates that the parallel 1D-CNN can effectively learn hierarchical features from a richer data representation automatically by many layers of convolution and pooling operations. One thing worth noting is that the option 13 aims to only extract the information from 3rd/4th/5th detail components where most of the ECG energy concentrates in [19], but owns worse performance compared with options 4, 7 and 10, due to losing part of the information in other signal components.

The second interesting observation is that applying different number of neural network layers and the introduction of the dropout regularization operation both have significant impact on the performance. When we gradually increase the number of neural network layers, we can notice that the performance basically gradually improves (CNN options 1 to 4, or options 5 to 8). If we further compare option group 1-4 with

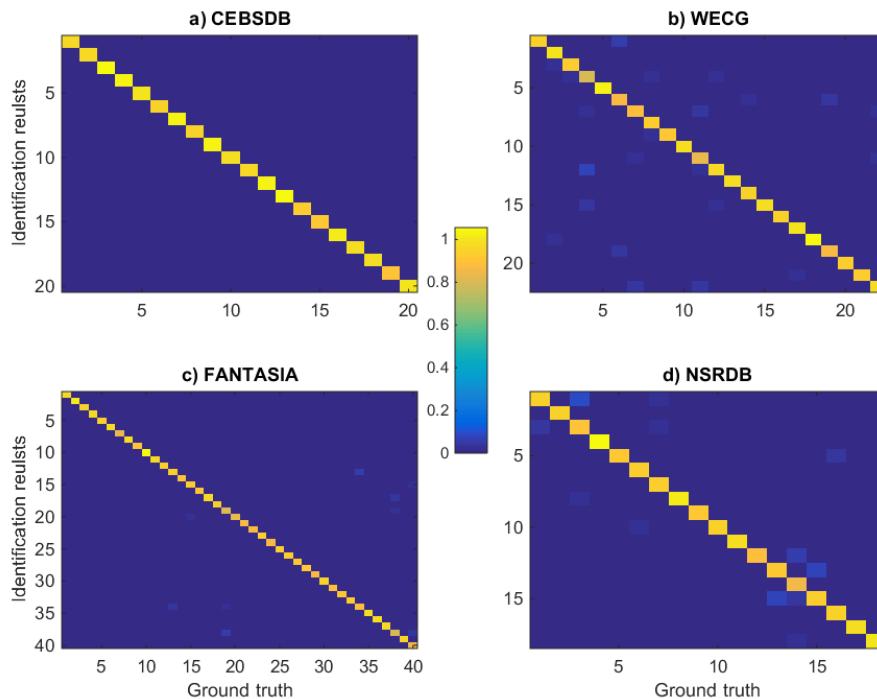


FIGURE 7. Confusion matrix for human identification based on testing data of four normal ECG datasets.

group 5-8, we can find dramatically improved identification rate resulting from the dropout regularization operation. As mentioned above, this is mainly benefitting from the fact that the specialization of neighboring neurons is suppressed to prevent a fragile model overfitting to the training data [31]. In such a way, other neurons are forced to learn to handle the responsibility of the discarded neurons, which enhances the generalization ability of the learned hierarchical features.

Therefore, the wavelet option 10 (Z_{3-9}) and CNN option 8 (deep CNN with 4 layers and the dropout operation) are selected as the final wavelet topology and parallel 1D-CNN topology, respectively. This topology selection process is based on the testing performance on the dataset CEBSDB, but is expected to be able to generalize to other seven normal and abnormal ECG datasets.

E. TOPOLOGY GENERALIZATION

The proposed MCNN algorithm is evaluated on eight ECG datasets. As mentioned above, the topologies of wavelet operation and neural network are determined based on the testing performance over the CEBSDB dataset, and are used on other datasets. The testing performance in terms of the confusion matrix of all eight datasets is given in Fig. 7 and Fig. 8, and corresponding identification rate is summarized in Table 2.

For four normal ECG datasets, the performance visualization in Fig. 7 clearly shows that the trained MCNN can effectively identify the human subjects (diagonal entries with a yellow color), with very little false positives or false negatives (non-zero off-diagonal entries). The top part of Table 2 also illustrates the high identification rate for these

TABLE 2. Identification rate of all datasets.

Abbreviations	# Subject	Identification rate (%)
CEBSDB	20	99.0
WECG	22	94.5
FANTASIA	40	97.2
NSRDB	18	95.1
Averaged @ normal datasets	25	96.5
STDB	28	90.3
MITDB	47	91.1
AFDB	23	93.9
VFDB	22	86.6
Averaged @ abnormal datasets	30	90.5
Averaged @ all datasets	28	93.5

four normal ECG datasets, from 94.5% to 99.0%. It is worth noting that there is still a high identification rate even for the FANTASIA dataset with a number of subjects as high as forty. Another thing worth noting is that a dataset with relatively less subjects does not necessarily correspond to a higher identification rate, such as the NSRDB dataset, due to the high variability of individual heart behaviors and blind segmentation operations. But these four datasets all correspond to an identification rate no less than 94.5%, and own an average identification rate of 90.5%, demonstrating the effectiveness of the proposed algorithm.

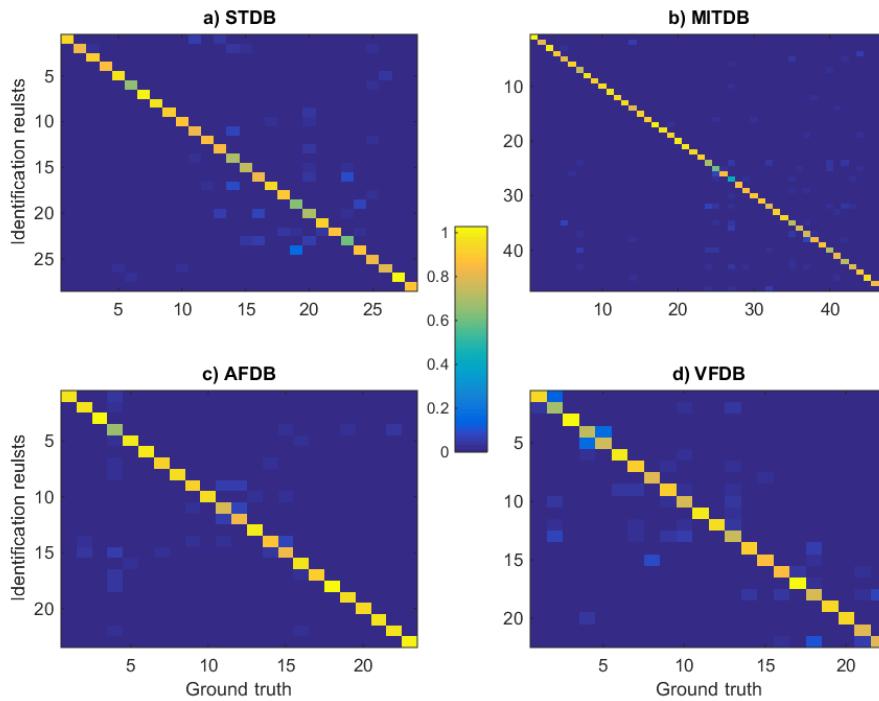


FIGURE 8. Confusion matrix for human identification based on testing data of four abnormal ECG datasets with severe heart diseases.

For four abnormal ECG datasets with severe heart diseases, the confusion matrices in Fig. 8 show that there are slightly increased false positives and false negatives, due to a much higher variability of the heartbeat morphologies. As shown in Table 1, there are difference kinds of heart diseases in these four datasets, corresponding to ST depression/elevation, arrhythmia with other 18 kinds of diseases, atrial fibrillation, malignant ventricular arrhythmia, respectively. Therefore, it is much more challenging to learn the underlying patterns of these time-varying abnormal heartbeat behaviors. The bottom part of Table 2 gives the identification rate for these four abnormal datasets, which is from 86.6% to 93.9%. The average identification rate is 93.5%, which is still an attractive result, considering that our user identification approach is directly performed on randomly chosen ECG segments using automatically learned features from raw data without feature engineering.

F. COMPARISON WITH STATE-OF-THE-ART WORKS

The proposed approach is further compared with other state-of-the-art works, as shown in Table 3, in terms of if they support blind signal processing (BSP), what kind of datasets are chosen for evaluation and what the average identification rate is. From the table, we can see the proposed approach is completely based on blind signal processing techniques, meaning that it is unnecessary to consider complicated and data-dependent heartbeat identification and segmentation steps. The blind signal processing not only avoids heavy algorithm engineering effort, but also guarantees a good generalization ability when we apply the algorithm to different datasets with diverse signal behaviors.

TABLE 3. Performance comparison with state-of-the-art works.

Methods	BSP	Datasets	ID Rate
Proposed	Yes	18 to 47 subjects @ 8 normal/abnormal DS	93.5%
Yao et al. [13]	No	20 subjects @ 1 normal DS	91.5%
Tan et al. [14]	No	10 subjects @ 1 abnormal DS	91.7%
Lourenco et al. [15]	No	16 subjects @ 1 normal DS	94.3%
Ye et al. [16]	No	18 to 65 subjects @ 3 normal/abnormal DS	85.1%
Ting et al. [17]	No	13 subjects @ 1 abnormal DS	87.5%
David et al. [39]	No	26 to 51 subjects @ 2 normal DS	99.7%
Adrian et al. [40]	No	50 subjects @ 1 normal DS	89.0%

Notes. BSP: blind signal processing; DS: dataset; ID: identification.

Moreover, we have evaluated our approach on as many as eight datasets, which is much more thorough than other works. One thing worth noting is that although Ye et al. evaluated their algorithm on three datasets, the identification rate is much lower than our approach. It could be because of the deep learning algorithm used in our study which can more effectively learn the underlying feature patterns from the data. Another thing worth noting is that the work of David et al. shows a high accuracy, however, they still need to firstly identify each heartbeat that needs lot of algorithm engineering compared with our blind signal processing techniques, and they only evaluated the method on two healthy datasets which are of much lower diversity compared with our evaluation method. Overall, the proposed algorithm owns

more advantages and high enough performance compared with these previously reported studies.

G. FURTHER DISCUSSION AND FUTURE WORK

ECG-based biometric identification is an emerging application and attracting more and more attentions. But we also notice that compared with currently mature techniques such as fingerprint-based identification, more efforts still need to be put in several aspects. For example, although more and more ECG datasets are available now benefitting from great advancement of data acquisition systems, the amount of data is still not comparable to the fingerprint data. Usually, hundreds or thousands of records of fingerprint can be obtained [41], which is much higher than the dataset sizes as shown in Table 3. Moreover, there are also more diversity induced by different heart health conditions as shown in Fig. 2, which brings more challenges to the generalization ability of the identification algorithms (one of the major goals of this study). In future, we will consider more datasets corresponding to diverse health conditions and sensor placement methods [42]. Another interesting work is to lower the power consumption of the identification algorithm for long-term wearable application scenarios [10], [43], [44].

Besides, we will further consider more data representation methods and neural network topologies, to explore the interesting connection between the data representation /network topology and the feature learning ability.

Moreover, investigations [45] have also found that the cardiac organisms react physiologically to stimulations such as danger and threat, and corresponding ECG signal also changes due to the cardiac defense mechanism (CDR). Therefore, identifying these external stimulation-related ECG patterns in future work is also expected to further improve the user identification performance.

IV. CONCLUSION

In this paper, we have proposed a novel multiresolution convolutional neural network for biometric human identification applications. Focusing on existing challenges, we have introduced blind signal processing and automatic feature learning techniques to effectively lower the algorithm engineering effort and also highly enhance the generalization ability of the algorithm. The proposed methodology can also be generalized to other quasi-periodical biometric signal-based user identification applications, such as photoplethysmogram (PPG), ballistocardiograph (BCG) and body movements (walking, running, etc.).

Our contributions include: 1) blindly select the signal segment for user identification purpose, which effectively avoids complicated and data-dependent signal event identification (e.g., ECG R peaks) and segmentation effort; 2) enrich the time-frequency representation by transforming data from the time domain to the wavelet domain, and remove phase difference among random-chosen signal segments by the auto-correlation approach; 3) introduce a parallel 1D-CNN to automatically learn multi-scale feature hierarchies from the

wavelet domain raw data, which can greatly lower the feature engineering effort and also capture intrinsic feature patterns more effectively; 4) explore the effect of many wavelet and CNN topologies to the ECG-based user identification task and determine a good combination; 5) evaluate the proposed algorithm extensively on eight ECG datasets and illustrate the advantages over state-of-the-art works. This study is expected to demonstrate that the proposed blind signal processing and deep learning techniques can effectively lower the algorithm engineering effort and provide a good generalization ability, for the biometric human identification applications.

REFERENCES

- [1] G. Fortino, R. Giannantonio, R. Gravina, P. Kuryloski, and R. Jafari, "Enabling effective programming and flexible management of efficient body sensor network applications," *IEEE Trans. Human-Mach. Syst.*, vol. 43, no. 1, pp. 115–133, Jan. 2013.
- [2] R. Gravina, P. Alinia, H. Ghasemzadeh, and G. Fortino, "Multi-sensor fusion in body sensor networks: State-of-the-art and research challenges," *Inf. Fusion*, vol. 35, pp. 68–80, May 2017.
- [3] A. Lounis, A. Hadidj, A. Bouabdallah, and Y. Challal, "Secure and scalable cloud-based architecture for e-health wireless sensor networks," in *Proc. 21st Int. Conf. Comput. Commun. Netw. (ICCCN)*, 2012, pp. 1–7.
- [4] J. Unar, W. C. Seng, and A. Abbasi, "A review of biometric technology along with trends and prospects," *Pattern Recognit.*, vol. 47, no. 8, pp. 2673–2688, 2014.
- [5] I. Odinaka, P.-H. Lai, A. D. Kaplan, J. A. O'Sullivan, E. J. Sirevaag, and J. W. Rohrbaugh, "ECG biometric recognition: A comparative analysis," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 6, pp. 1812–1824, Dec. 2012.
- [6] Y. M. Chi, T.-P. Jung, and G. Cauwenberghs, "Dry-contact and noncontact biopotential electrodes: Methodological review," *IEEE Rev. Biomed. Eng.*, vol. 3, pp. 106–119, Oct. 2010.
- [7] S. I. Safie, J. J. Soraghan, and L. Petropoulakis, "ECG biometric authentication using pulse active width (PAW)," in *Proc. IEEE Workshop Biometr. Meas. Syst. Secur. Med. Appl. (BIOMS)*, Sep. 2011, pp. 1–6.
- [8] F. Camento, A. Lourenço, H. Silva, and A. Fred, "Review and comparison of real time electrocardiogram segmentation algorithms for biometric applications," in *Proc. 6th Int. Conf. Health Inform. (HEALTHINF)*, 2013. [Online]. Available: <https://pdfs.semanticscholar.org/788a/62ac3567e7793cfdf14aae65296101c43042.pdf>
- [9] Q. Zhang, D. Zhou, and X. Zeng, "A novel framework for motion-tolerant instantaneous heart rate estimation by phase-domain multi-view dynamic time warping," *IEEE Trans. Biomed. Eng.*, to be published.
- [10] Q. Zhang, C. Zahed, V. Nathan, D. A. Hall, and R. Jafari, "An ECG dataset representing real-world signal characteristics for wearable computers," in *Proc. IEEE Biomed. Circuits Syst. Conf. (BioCAS)*, Oct. 2015, pp. 1–4.
- [11] M. Tantawi, K. Revett, A.-B. Salem, and M. F. Tolba, "ECG based biometric recognition using wavelets and RBF neural network," in *Proc. 7th Eur. Comput. Conf. (ECC)*, 2013, pp. 100–105.
- [12] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, "Learning and transferring mid-level image representations using convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1717–1724.
- [13] J. Yao and Y. Wan, "A wavelet method for biometric identification using wearable ECG sensors," in *Proc. 5th Int. Summer School Symp. Med. Devices Biosensors (ISSS-MDBS)*, 2008, pp. 297–300.
- [14] X. Tang and L. Shu, "Classification of electrocardiogram signals with RS and quantum networks neural," *Int. J. Multimedia Ubiquitous Eng.*, vol. 9, no. 2, pp. 363–372, 2014.
- [15] A. Lourenço, H. Silva, and A. Fred, "Unveiling the biometric potential of finger-based ECG signals," *Comput. Intell. Neurosci.*, vol. 2011, Jun. 2011, Art. no. 720971.
- [16] C. Ye, M. T. Coimbra, and B. V. K. Vijaya Kumar, "Investigation of human identification using two-lead electrocardiogram (ECG) signals," in *Proc. 4th IEEE Int. Conf. Biometrics, Theory Appl. Syst. (BTAS)*, Sep. 2010, pp. 1–8.
- [17] C.-M. Ting and S.-H. Salleh, "ECG based personal identification using extended Kalman filter," presented at the IEEE 10th Int. Conf. Inf. Sci., Signal Process. Appl., Kuala Lumpur, Malaysia, 2010, pp. 774–777.

- [18] Q. Zhang, D. Zhou, and X. Zeng, "A novel machine learning-enabled framework for instantaneous heart rate monitoring from motion-artifact-corrupted electrocardiogram signals," *Physiol. Meas.*, vol. 37, no. 11, pp. 1945–1967, 2016.
- [19] S. Mukhopadhyay, S. Biswas, A. B. Roy, and N. Dey, "Wavelet based QRS complex detection of ECG signal," *Int. J. Eng. Res. Appl.*, vol. 2, no. 3, pp. 2361–2365, 2012.
- [20] M. A. García-González, A. Argelagós, M. Fernández-Chimeno, and J. Ramos-Castro, "Differences in QRS locations due to ECG lead: Relationship with breathing," in *Proc. 13th Medit. Conf. Med. Biol. Eng. Comput.*, 2014, pp. 962–964.
- [21] M. A. García-González, A. Argelagós-Palau, M. Fernandez-Chimeno, and J. Ramos-Castro, "A comparison of heartbeat detectors for the seismocardiogram," in *Proc. Comput. Cardiol. Conf. (CinC)*, 2013, pp. 461–464.
- [22] N. Iyengar, C. K. Peng, R. Morin, A. L. Goldberger, and L. A. Lipsitz, "Age-related alterations in the fractal scaling of cardiac interbeat interval dynamics," *Amer. J. Physiol.-Regulatory, Integr. Comparative Physiol.*, vol. 271, pp. R1078–R1084, Oct. 1996.
- [23] A. L. Goldberger *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.
- [24] P. Albrecht, *ST Segment Characterization for Long Term Automated ECG Analysis*. Cambridge, MA, USA: Massachusetts Institute of Technology, Dept. Elect. Eng. Comput. Sci., 1983.
- [25] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH arrhythmia database," *IEEE Eng. Med. Biol. Mag.*, vol. 20, no. 3, pp. 45–50, May/Jun. 2001.
- [26] G. B. Moody and R. G. Mark, "A new method for detecting atrial fibrillation using RR intervals," *Comput. Cardiol.*, vol. 10, no. 1, pp. 227–230, 1983.
- [27] S. D. Greenwald, "The development and analysis of a ventricular fibrillation detector," M.S. thesis, Dept. Elect. Eng. Comput. Sci., Massachusetts Inst. Technol., Cambridge, MA, USA, 1986.
- [28] H. Tanaka, K. D. Monahan, and D. R. Seals, "Age-predicted maximal heart rate revisited," *J. Amer. College Cardiol.*, vol. 37, no. 1, pp. 153–156, 2001.
- [29] J. B. Allen, "Short term spectral analysis, synthesis, and modification by discrete Fourier transform," *IEEE Trans. Acoust., Speech Signal Process.*, vol. 25, no. 3, pp. 235–238, Jun. 1977.
- [30] M. K. Kiymik, I. Guler, A. Dizibuyuk, and M. Akin, "Comparison of STFT and wavelet transform methods in determining epileptic seizure activity in EEG signals for real-time application," *Comput. Biol. Med.*, vol. 35, no. 7, pp. 603–616, 2005.
- [31] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.
- [32] L. Deng *et al.*, "A deep convolutional neural network using heterogeneous pooling for trading acoustic invariance with phonetic confusion," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2013, pp. 6669–6673.
- [33] M. Bojarski *et al.* (2016). "End to end learning for self-driving cars." [Online]. Available: <https://arxiv.org/abs/1604.07316>
- [34] J. D. WICHARD *et al.*, "CNN in drug design—Recent developments," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2015, pp. 405–408.
- [35] Keras: Deep Learning Library for Theano and TensorFlow, accessed on Dec. 2016. [Online]. Available: <https://keras.io/>
- [36] Nvidia. GeForce GTX 960M GPU, accessed on Dec. 2016. [Online]. Available: <http://www.geforce.com/hardware/notebook-gpus/geforce-gtx-960m/specifications>
- [37] Nvidia. CUDA Parallel Computing Platform. [Online]. Available: http://www.nvidia.com/object/cuda_home_new.html
- [38] S. Sra, *Optimization for Machine Learning*. Cambridge, MA, USA: MIT Press, 2012.
- [39] D. Pereira Coutinho, H. Silva, H. Gamboa, A. Fred, and M. Figueiredo, "Novel fiducial and non-fiducial approaches to electrocardiogram-based biometric systems," *IET Biometrics*, vol. 2, no. 2, pp. 64–75, Jun. 2013.
- [40] A. D. C. Chan, M. M. Hamdy, A. Badre, and V. Badee, "Wavelet distance measure for person identification using electrocardiograms," *IEEE Trans. Instrum. Meas.*, vol. 57, no. 2, pp. 248–253, Feb. 2008.
- [41] Biometric Special Databases and Software, accessed on Dec. 2016. [Online]. Available: <https://www.nist.gov/itl/iad/image-group/resources/biometric-special-databases-and-software>
- [42] Q. Zhang, D. Zhou, and X. Zeng, "Highly wearable cuff-less blood pressure and heart rate monitoring with single-arm electrocardiogram and photoplethysmogram signals," *Biomed. Eng. Online*, vol. 16, no. 1, p. 23, 2017.
- [43] J. Birjandtalab, Q. Zhang, and R. Jafari, "A case study on minimum energy operation for dynamic time warping signal processing in wearable computers," in *Proc. IEEE Int. Conf. Pervas. Comput. Commun. Workshops (PerCom Workshops)*, Mar. 2015, pp. 415–420.
- [44] J. Eriksson *et al.*, "International workshop on the impact of human mobility in pervasive systems and applications (PerMoby 2016)—Welcome and committees: Welcome to PerMoby 2016," in *Proc. IEEE Int. Conf. Perv. Comput. Commun. Workshops (PerCom Workshops)*, Mar. 2016, pp. 1–2.
- [45] R. Gravina *et al.*, "Automatic methods for the detection of accelerative cardiac defense response," *IEEE Trans. Affect. Comput.*, vol. 7, no. 3, pp. 286–298, Jul./Sep. 2016.



QINGXUE ZHANG (M'14) received the B.S. and M.S. degrees in science and technologies of electronics from Xi'an Jiaotong University, Xi'an, China, in 2004 and 2007, respectively, and the Ph.D. degree in electrical engineering from The University of Texas at Dallas, Richardson, TX, USA, in 2017. From 2007 to 2013, he was with Hisilicon Co., Ltd., as a System Architect and a Team Lead. His research interests include smart health and wearable computers, with emphasis on wearable prototyping, signal processing/machine learning algorithms, and emerging applications.



of Things, VLSI design, circuit and systems, and CAD tools and algorithms.

Dr. Zhou received the IEEE Circuits and Systems Society Darlington Award in 1993, the National Science Foundation (NSF) Young Investigator Award in 1994, the Outstanding Overseas Young Investigator Award from NSF of China in 2000, the Changjiang Honor Professor from the Ministry of Education Department of China in 2002, and the Grand Award HDTV Terrestrial Broadcasting ASIC Chip: Chinese Television Number One from the Electronics Society of China in 2005, and was selected as the Thousand People Plan Professor from China in 2010.



XUAN ZENG (M'97) received the B.S. and Ph.D. degrees in electrical engineering from Fudan University, Shanghai, China, in 1991 and 1997, respectively.

She was a Visiting Professor with the Department of Electrical Engineering, Texas A&M University, College Station, TX, USA, in 2002, and with the Department of Microelectronics, Technische Universiteit Delft, Delft, The Netherlands, in 2003. She was the Director of the State Key Laboratory of ASIC and System, Fudan University, from 2008 to 2012. She is currently a Full Professor with the Department of Microelectronics, Fudan University. Her research interests include smart city, smart health, machine learning, big data, design for manufacturability, high-speed interconnect analysis and optimization, analog behavioral modeling, circuit simulation, and ASIC design.

Dr. Zeng received the Chinese National Science Funds for Distinguished Young Scientists in 2011 and the First-Class of Natural Science Prize of Shanghai in 2012. She was the Changjiang Distinguished Professor with the Ministry of Education Department of China in 2014. She is currently an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS PART II: EXPRESS LETTERS.