



Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Deep-ECG: Convolutional Neural Networks for ECG biometric recognition

Ruggero Donida Labati, Enrique Muñoz*, Vincenzo Piuri, Roberto Sassi, Fabio Scotti

Department of Computer Science, Università degli Studi di Milano, via Bramante, 65, Crema, CR I-26013, Italy

ARTICLE INFO

Article history:
Available online xxx

MSC:
41A05
41A10
65D05
65D17

Keywords:
Biometrics
ECG
Deep learning

ABSTRACT

Electrocardiographic (ECG) signals have been successfully used to perform biometric recognition in a wide range of applications. However, ECG-based biometric systems are usually less accurate than technologies based on other physiological traits. To increase their performance, it is necessary to study novel approaches. Deep learning methods, like Convolutional Neural Networks (CNNs), can automatically extract distinctive features, and have demonstrated their effectiveness for other biometric systems. In this paper, we present Deep-ECG, a CNN-based biometric approach for ECG signals. To the best of our knowledge, this is the first study in the literature that uses a CNN for ECG biometrics. Deep-ECG extracts significant features from one or more leads using a deep CNN and compares biometric templates by computing simple and fast distance functions, obtaining remarkable accuracy for identification, verification and periodic re-authentication. Furthermore, using a simple quantization procedure, Deep-ECG can obtain binary templates that can facilitate the use of ECG-based biometric systems in conjunction with cryptographic applications. We also propose a simple method to enlarge the training dataset of ECG samples, which can increase the performance of deep neural networks. We performed experiments on large sets of samples acquired in uncontrolled conditions, proving the accuracy and robustness of Deep-ECG in non-ideal scenarios. Furthermore, we evaluated the performance of Deep-ECG for the PTB Diagnostic ECG Database, obtaining identification accuracy better or comparable to the best performing methods in the literature, also for signals with different characteristics with respect to the ones used to train the CNN.

© 2018 Published by Elsevier B.V.

1. Introduction

The use of one-dimensional physiological signals, such as the electrocardiogram (ECG) or electroencephalogram [22], as biometric traits is attracting much attention from the scientific community. Physiological signals have important advantages [30]: (i) it is difficult to counterfeit them; (ii) all living individuals present them; (iii) they contain information about clinical status or psychological states, which can be useful for other applications; (iv) their acquisition for long periods of time does not require explicit actions from the user, which facilitates their application for periodic re-authentication.

Among physiological signal analysis techniques, ECG stands out on account of its maturity and diffusion. Furthermore, recent studies proved that ECG biometrics can obtain satisfactory accuracy for different kinds of applications [18]. However, ECG-based biometric systems have to cope with a range of issues [31], including: (i) high intraclass variability caused by the heart rate and performed activities;

(ii) lack of studies that demonstrate the discriminability in large scale datasets; (iii) long-term changes in the characteristics of an individual's ECG. In particular, ECG biometric systems can be affected by a significant performance decrease in uncontrolled acquisition scenarios [30].

In this paper, we present Deep-ECG, a novel ECG-based biometric recognition approach based on deep learning. We propose using a deep Convolutional Neural Networks (CNN) to extract features that permit to perform closed-set identification, identity verification and periodic re-authentication. To the best of our knowledge, this paper is the first study in the literature that uses a deep CNN in ECG-based biometric systems. Furthermore, Deep-ECG can create both real and binary templates. Binary templates permit to perform a fast matching based on the Hamming distance and enable the use of a wide set of template protection methods designed for binary templates, such as Iriscode [6]. Deep-ECG can use a single lead or fuse multiple leads to achieve higher accuracy.

The paper presents four main contributions. First, we present the first study on ECG biometrics based on deep CNNs. Second, the proposed approach can achieve a remarkable accuracy. Third, we introduce a method to binarize ECG templates that permits to re-

* Corresponding author.

E-mail address: enrique.munoz@unimi.it (E. Muñoz).

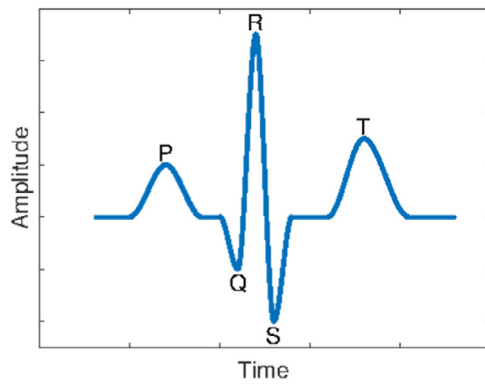


Fig. 1. Main components of an ideal heartbeat.

duce the matching time and apply template protection techniques. Fourth, we describe a method to extend ECG biometric datasets that permits to increase the recognition performance and robustness of deep CNNs.

We evaluated the performance of the proposed approach on datasets acquired in uncontrolled conditions and simulated different application scenarios, including closed-set identification, identity verification and periodic re-authentication. The obtained results proved the accuracy and robustness of Deep-ECG in non-ideal conditions. Furthermore, we proved that Deep-ECG can be used for ECG signals acquired in different conditions, achieving better or comparable accuracy with respect to the best performing state-of-the-art methods.

The paper is structured as follows. Section 2 analyzes the state of the art. Section 3 presents the proposed approach. Section 4 presents the performed experiments. Finally, Section 5 concludes the work.

2. Related works

The design of ECG-based biometric systems has to consider several aspects that can have a great influence in the final performance, such as the number of leads analyzed, the type of features exploited, and the type of classifier used [30].

One-lead systems are more widespread and can increase the user acceptance [28]. The use of multiple leads permits to increase the accuracy of the system [1].

Depending on the features that are used for recognition, ECG biometric systems can be classified into fiducial, non-fiducial and partially fiducial [30]. Fiducial points are six characteristic points (P, Q, R, S, T and U) present in all healthy ECG signals (see Section 3 and Fig. 1 for a detailed description). Partially fiducial methods usually locate the R point to segment the heartbeat waveform and analyze the full waveform or a part of it (the QRS complex) [2,7,39]. The search of R is generally easier than the search of other fiducial points, because it is the highest and sharpest peak in a heartbeat. In addition, the QRS complex is less sensible to physical and emotional variations with respect to the other portions of ECG signals [39].

ECG-based biometric systems can also be classified taking into account the classifier used to perform recognition tasks [30]. All classifier-based systems share a common idea: they analyze the ECG signal to extract some relevant features to be used as input of the classifier. The features may be extracted using heterogeneous techniques, e.g. local binary patterns [27] or random projections [3]. These features can be elaborated using other methods, such as principal component analysis [16] or bag of words [3]. However, the transformation of a raw ECG signal into a suitable feature vector for classification has to be carefully engineered and requires

a considerable expert knowledge. Deep learning methods, such as CNNs, overcome this limitation by using multiple levels of representation that can exploit raw data to automatically discover the representations needed for recognition. The key aspect is that the design of the features does not require experts because it is possible to use data to automatically learn an optimized feature extraction function by using a general learning method [26].

In the literature, there is a relevant number of studies on methods based on deep learning techniques applied to biometrics, including face [31] and iris [12]. Regarding one-dimensional signals, deep learning techniques have obtained promising results for speaker recognition [13,34]. However, speech signals are very different from cardiac signals and these approaches are not suitable for ECG. Only a few studies have considered the application of deep learning strategies for ECG analysis, but they focus on the classification of heartbeats in healthy and non-healthy [29] using techniques such as CNNs [23,24,32], autoencoders [33], or deep belief networks [17]. To the best of our knowledge, only the work in [10] uses a deep learning technique for ECG biometrics. This work applies deep autoencoders and considers only the closed-set biometric identification task, thus requiring to re-train the system for enrolling new users. This study did not prove that deep-networks trained using samples acquired from a set of individuals can successfully be used for samples of unknown individuals. In contrast, Deep-ECG uses deep CNNs to extract features from samples of users not known during the training phase and can perform closed-set identification, identity verification and periodic re-authentication.

3. The proposed approach: Deep-ECG

A healthy ECG signal usually presents six fiducial points (Fig. 1). The QRS complex, which joins the Q, R and S points, usually appears as the central and most visually prominent part of the signal. It represents a single event, the depolarization of the right and left ventricles. The analysis of the QRS complex for biometric tasks can be advantageous since the QRS complex is the portion of the ECG that is less affected by variations due to physical activities or emotional states. Thus, Deep-ECG considers only the QRS complex.

Deep-ECG can work in three modalities: closed-set identification, identity verification and periodic re-authentication. In all cases, our approach extracts a set of m QRS complexes from ECG samples of short duration, and joins them in the signal V . In closed-set identification, a CNN processes V and indicates who is the closest registered user. In identity verification and periodic re-authentication, the CNN processes V to obtain a biometric template T . To compute the matching score, we use simple distance measures, such as, Euclidean or cosine distances. Furthermore, we propose a quantization technique to compute binary templates and a matching algorithm based on the Hamming distance.

Deep-ECG can be divided into the following steps:

1. signal preprocessing;
2. CNN feature extraction;
3. recognition:
 - Soft-max based identification;
 - Matching for identity verification and periodic re-authentication.

Fig. 2 shows the schema of the proposed approach in the different modalities.

3.1. Signal preprocessing

In this step, we enhance the ECG signals and extract the m most discriminative QRS complexes. We use the algorithms presented in [8,9].

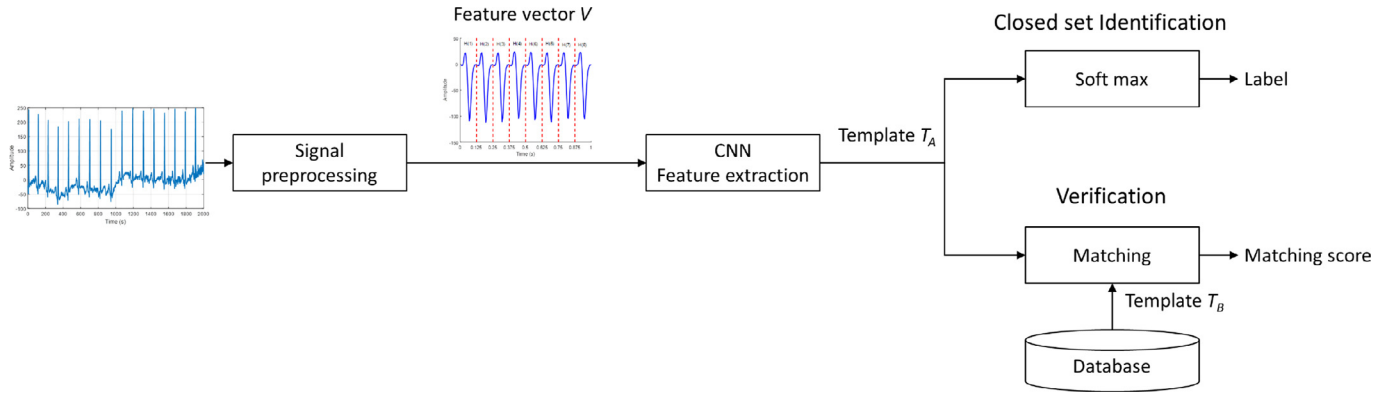


Fig. 2. Biometric verification process of the proposed approach. It can work in two modalities: (i) closed-set identification, and (ii) identity verification and periodical re-identification.

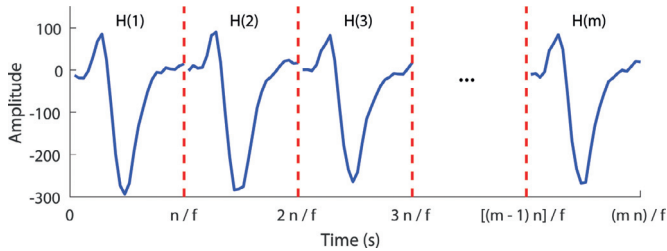


Fig. 3. Example of feature vector V , composed by m QRS complexes. The vector V is the input of the CNN.

First, we reduce the noise by applying a notch IIR filter and normalize the signal's baseline by using a third order high-pass Butterworth filter with cutoff frequency of 0.5 Hz.

As second step, we estimate the position of the n R fiducial points by using an automatic labeling tool (Vision Premier, SpaceLab-Burdick Inc.). After that, we select a time window of 0.125 s for each R point, obtaining a vector H of n QRS complexes. We extract the m most discriminative complexes, obtaining feature vector V . To estimate the quality of a QRS complex, we calculate its correlation with the average QRS pattern of H , \overline{QRS} . Subsequently, we compute the cross-correlation between \overline{QRS} and every QRS complex of H , obtaining the vector C . We concatenate the m QRS complexes with the maximum values of C , obtaining the vector V . If the number i of QRS signals belonging to H is less than m , we complete V by replicating the QRS complex with the maximum value of C . Fig. 3 shows an example of feature vector V .

Before processing V using the CNN, we subtract from it \bar{V} , the mean value of the signals used for training, and multiply the result by 256.

3.2. Deep-ECG CNN feature extraction

Deep-ECG can work for closed-set identification, identity verification and periodic re-authentication. These tasks share the feature extraction module based on a deep CNN. In closed-set identification and during training, Deep-ECG processes the features computed by the CNN using a Soft-max layer that returns the user identity. In verification and periodic re-authentication, Deep-ECG uses a biometric matcher to compare templates obtained from the data computed by the CNN.

3.2.1. CNN architecture

Fig. 4 shows the architecture of the CNN used by Deep-ECG. It is composed of six convolution layers that use ReLU (Rectified Linear Units) neurons, three max-pooling layers, three LRN (Local Re-

sponse Normalization) layers, one dropout layer, a fully connected layer, and a Soft-max layer (for training and closed-set identification). ReLU neurons permit to add non-linearity to the network, favoring a deeper representation. LRN layers are useful when combined with ReLU neurons, since they have unbounded outputs. LRN layers permit to detect high frequency features with large activations with regard to their neighborhood. At the same time, they reduce the impact of large responses in a local neighborhood. In this way, it is possible to emulate the useful neurobiological concept of lateral inhibition. In addition, we use the dropout regularization, which permits to improve generalization and avoid co-adaptation by randomly setting a fraction of the activations to zero. The last layer of the network is a soft-max layer used to train the CNN and to perform closed-set identifications. The output of this layer is an integer label corresponding to the identifier of the user. Considering a training dataset composed of samples pertaining to n_u users, the output of the CNN is an integer value $\in [1, 2, \dots, n_u]$.

Convolutional layers process the input signal x by convolving it with a bank of K filters f , using biases b . As a result, they obtain an output signal y . Here

$$x \in \mathbb{R}^{H \times W \times D}, f \in \mathbb{R}^{H' \times W' \times D \times D''}, y \in \mathbb{R}^{H'' \times W'' \times D''}, \quad (1)$$

where H , W and D are the height, width and depth dimensions, respectively. In the basic configuration of the convolutional layer, for each coordinate (i, j, d) , the output is computed as follows:

$$y_{i''j''d''} = b_{d''} + \sum_{i'=1}^{H'} \sum_{j'=1}^{W'} \sum_{d'=1}^D f_{i'j'd'} \times x_{i''+i'-1, j''+j'-1, d'+d''}. \quad (2)$$

In some layers it is necessary to perform a padding of the input signal x or a subsampling stride of the output. In particular, we consider top-bottom-left-right paddings ($P_h^-, P_h^+, P_w^-, P_w^+$) and strides (S_h, S_w). In these cases, the output is calculated as follows (where x is implicitly expanded using zeros as needed):

$$y_{i''j''d''} = b_{d''} + \sum_{i'=1}^{H'} \sum_{j'=1}^{W'} \sum_{d'=1}^D f_{i'j'd'} \times x_{S_h(i''-1)+i'-P_h^-, S_w(j''-1)+j'-P_w^+, d'+d''}. \quad (3)$$

The pooling layers use a max-pooling operator to reduce the feature space. This operator calculates, for each feature channel, the maximum response in a $H' \times W'$ patch, using the following approach:

$$y_{i''j''d} = \max_{1 \leq i' \leq H', 1 \leq j' \leq W'} x_{i''+i'-1, j''+j'-1, d}. \quad (4)$$

The LRN layers apply the following operator:

$$y_{ijk} = x_{ijk} \left(\kappa + \alpha \sum_{t \in G(k)} x_{ijt}^2 \right)^{-\beta}, \quad (5)$$

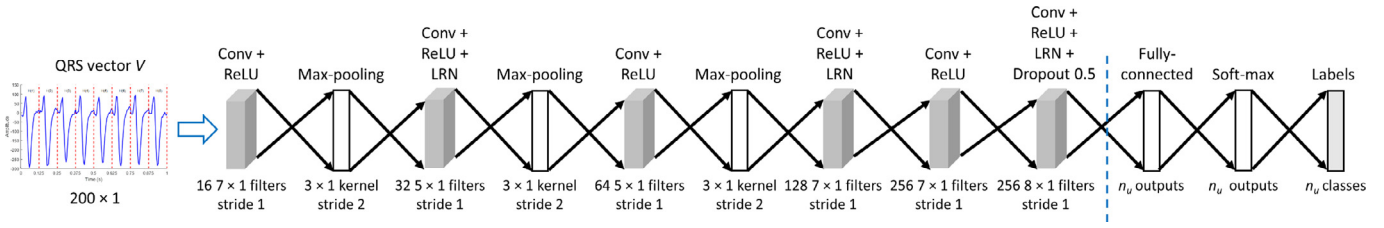


Fig. 4. Schema of the proposed deep learning approach based on CNN. In our approach, we use the CNN as a feature extractor for computing biometric templates to be used for identity verification and periodic re-authentication. In these cases, we remove the fully connected layer of the trained CNN. The fully connected layer is used only to train the network and for closed-set identification.

where, for each output channel k , $G(k) \subset 1, \dots, D$ is the corresponding set of input channels. All LRN layers in Deep-ECG use the parameters $D = 5$, $\kappa = 1$, $\alpha = 2 \times 10^{-4}$ and $\beta = 0.75$.

The ReLU layers utilize a non-saturating activation function $y_i = \max(0, x_i)$. The Softmax classifier calculates its output as $y_j = \frac{e^{x_j}}{\sum_{i=1}^n e^{x_i}}$, where n is the number of inputs to the neuron.

The number of layers, size of the convolution kernels, max-pooling kernels and stride have been tuned empirically.

3.2.2. Training of the CNN

To train the CNN, we used Stochastic gradient descent with momentum [25]. We initialized the weights of the CNN filters using random sampling from a Gaussian distribution with zero mean and 0.01 standard deviation. We trained the CNN for 500 epochs. We used a learning rate of 0.001, weight decay of 0.0005 and batch size of 500 samples.

A CNN requires a great number of samples to produce a general model that can be used for biometric recognition. To increase the number of samples used for training, we propose to consider signals from different leads as distinct traits. This is a procedure commonly used for other biometric modalities, such as fingerprint and iris. Nevertheless, to the best of our knowledge, this is the first work that attempts this technique for ECG biometrics.

3.2.3. Deep-ECG templates

Deep-ECG can compute two types of templates, real or binary features. Binary templates, such as Iriscode, have very interesting properties. First, using them we can perform a fast matching based on the Hamming distance. Furthermore, they enable the use of a wide set of template protection methods. Hence, a binary template for ECG biometrics is a desirable result that can have a deep impact in its development [7,15].

Real features take values in the range $[-\infty, +\infty]$ and correspond to the activation of the last LRN layer.

Binary features take values in the range $[0,1]$ and are obtained by quantizing the real features. In particular, we intend to create a feature set in which the probability that a given feature takes a value of 1 is approximately 0.5. To do so, we compute a binary template T in the following way:

$$T(i) = \begin{cases} 1, & \text{if } r_i \geq t_i \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

where, r_i is real feature i and t_i is the threshold used for feature i , calculated as $t_i = \text{median}(R_i)$, being R_i the set of all values of i th real feature, estimated using the training set.

3.3. Deep-ECG matching for identity verification and periodic re-authentication

Deep-ECG computes the matching scores between the templates T_A and T_B using simple distance functions for templates

composed of real numbers. We have tested different distance functions, such as: Euclidean distance, cosine correlation, and Spearman distance. The first distance function is the one that obtained the best performance. In the following, we will refer to the configuration that uses real features and Euclidean distance matching as Deep-ECG-R. For binary templates, Deep-ECG computes the Hamming distance between T_A and T_B . In the following, we will refer to this configuration as Deep-ECG-B.

3.4. Deep-ECG fusion of multiple leads

In the literature, there are different strategies for biometric fusion [35]. Since the study of the best fusion method is out of the scope of this paper, we have considered simple matching-score level methods that permit to increase the recognition accuracy, but do not require a training process. In particular, we evaluated the minimum, maximum and mean methods. Considering the results of previous studies [8,9] and those obtained by Deep-ECG, we selected $S_{Fused} = \text{mean}(S_{lead_X}, S_{lead_Y}, S_{lead_Z})$.

4. Experimental results

This section presents the datasets, testing procedures, and the performance achieved by Deep-ECG in terms of closed-set identification, identity verification, periodic re-authentication, and computational time. This section also compares the performance of Deep-ECG with a wide set of state-of-the-art methods, showing that our approach achieved the best identification accuracy.

4.1. Datasets

To evaluate the capability of Deep-ECG to work in realistic conditions, with a relevant number of users, samples acquired during long periods of time and with continuity, we chose the database E-HOL-03-0202-003 (Intercity Digital Electrocardiogram Alliance—IDEAL) [37]. In contrast, most datasets in the literature include few users or short sessions [28], and cannot be used to train and test properly our approach.

We considered two public datasets to extract training and test sets, namely, E-HOL-03-0202-003 (Intercity Digital Electrocardiogram Alliance—IDEAL) database [37] and PTB Diagnostic ECG Database (Physionet) [14].

E-HOL-03-0202-003 database contains digital signals from 202 healthy individuals captured from Holter recordings (SpaceLab-Burdick Inc.) of around 24 h. We discarded the recordings of 17 users that were corrupted by noise and artifacts. During the acquisitions, the users could act freely, without restrictions or control on the performed activities. The database contains the same proportion of males and females. The ECG signals have a sampling frequency of 200 Hz, an amplitude resolution of $10 \mu V$, and have been captured using a three pseudo-orthogonal leads configuration (X, Y and Z).

PTB Diagnostic ECG Database contains 549 acquisitions from 290 subjects, with different time durations of around 2 min. This

database has been collected in a medical ambulatory. Only 52 volunteers are healthy and the rest of the individuals suffer of heart diseases. For each subject, there are from one to five acquisitions. Each acquisition includes 15 simultaneously measured ECG signals: the conventional 12 leads together with the 3 Frank orthogonal leads X, Y and Z (which we used in our analysis). The signals have a sampling frequency of 1000 Hz and an amplitude resolution of 0.5 μ V. To perform tests simulating realistic application scenarios and use data similar to that adopted by other studies in the literature, we considered only signals acquired from the 52 healthy volunteers. To use the signals of this database with CNNs trained for the ECG E-HOL-03-0202-003 database, we preprocessed the ECGs by re-sampling their frequency to 200 Hz.

To create biometric samples for training and test, we divided the ECG signals into slots of 10 s. The samples have fixed length of 10 s and are not temporally aligned in terms of fiducial points. The samples for which the segmentation algorithm was not able to find at least one R fiducial point were discarded. A period of 10 s is typically used for medical analysis and represents a good trade-off for accuracy [28].

For each sample, we extracted a feature vector V of $m = 8$ QRS samples. Previous studies proved that 8 heartbeats are a balanced trade-off between accuracy and usability [30].

4.1.1. Training dataset

To obtain a robust CNN, it is necessary to use a training set composed of a wide number of heterogeneous samples.

We divided ECG E-HOL-03-0202-003 database in two disjoint groups of users, one used for training and the other for testing. The training dataset, DB_H_T includes samples collected from 93 users, extracted from all the available leads (X, Y and Z). DB_H_T is composed of 1000 acquisitions per user. To increase the number of data usable to train the CNN, we considered the 3 leads as separated traits, thus simulating 279 users. The total number of training samples is 277,563.

To validate the CNN and evaluate its accuracy for closed-set identification on Holter signals, we used 70% of the samples of each user of DB_H_T for training and the remaining 30% for validation.

4.1.2. Test datasets

To test the proposed approach, we created three datasets from the samples related to the 92 users not included in DB_H_T . Thus, the testing datasets are composed of samples pertaining to different users with respect to DB_H_T . The datasets are described in the following:

- DB_H_S : we created this dataset using samples from the E-HOL-03-0202-003 database to check the identity verification accuracy of Deep-ECG in short term verifications. It contains a total of 8280 samples from 92 users and 3 leads. For each user, we sampled the leads X, Y and Z at 30 instants of time. We obtained a total of 90 samples per user (30 samples per lead), acquired consecutively during a time span of 300 s.
- DB_H_L : we created this dataset using samples from the E-HOL-03-0202-003 database to check the identity verification accuracy of Deep-ECG in long term verifications. It contains a total of 8280 samples from 92 users and 3 leads. For each user, we sampled the leads X, Y and Z at 30 instants of time. We obtained a total of 90 samples per user (30 samples per lead), acquired with a distance of 300 s between each other during a time span of 150 min.
- DB_H_C : we created this dataset using samples from the E-HOL-03-0202-003 database to check the periodic re-authentication accuracy of Deep-ECG. It contains a total of 138,000 samples from 92 users and 3 leads. For each user, we sampled the leads

X, Y and Z at 500 instants of time. We obtained a total of 1500 samples per user (500 samples per lead), acquired with a distance of 60 s between each other during a time span of 500 min.

- DB_PTB : we created this dataset using samples from the PTB Diagnostic ECG Database to evaluate the applicability of Deep-ECG on signals obtained in different conditions (ambulatory vs. Holter) and with different equipments from the ones used to train the CNN. Since the PTB Diagnostic ECG Database has been used by many studies on biometric recognition methods based on ECG signals, we created this dataset also to compare the performance of Deep-ECG with a wide set of state-of-the-art methods. It contains a total of 912 samples from 52 healthy users and 3 leads (X, Y and Z).

4.2. Evaluation procedure

We used different protocols to evaluate the performance of the proposed approach. To test the closed-set identification accuracy, we used a subset of DB_H_T . In particular, the validation set was split in two sets, one used for validation and the other for testing. The used figures of merit are the Cumulative Match Characteristic (CMC) curve and the Rank(N) [21].

To test the identity verification accuracy, we used the datasets DB_H_S and DB_H_L . For each dataset and lead, we had 40,020 genuine comparisons and 3,767,400 impostors. We evaluated the achieved performance using the Receiver Operating Characteristic (ROC) curve and the Equal Error Rate (EER) [21].

To test the periodic re-authentication, we used the dataset DB_H_C . For each user, we included the first template in the gallery, and we considered the rest of the templates as probes. We compared all enrolled templates with all test templates, for a total of 45,803 genuine comparisons and 4,176,455 impostor comparisons. We evaluated the achieved performance in term of EER.

4.3. Identification accuracy and training strategies

In this section we evaluate the performance of Deep-ECG for closed-set identification, and check the benefits of using an augmented dataset that considers signals from different leads as distinct traits. In particular, we considered three strategies: (i) a different CNN for each lead, trained with data obtained only from that lead; (ii) a single CNN trained using information obtained from a single lead; (iii) a single CNN trained using data obtained from all leads, considering signals from different leads as distinct traits.

The obtained results illustrate that the most robust option is option (iii). Regarding closed-set identification, following option (i) or (ii), using a dataset that contains only the samples coming from lead X, Deep-ECG managed to obtain a Rank(1) error of 1.96%, with 93 different classes. While using the augmented training dataset introduced in option (iii), Deep-ECG managed to obtain a Rank(1) error of 2.07%, with 279 different classes, which is a marginal error increase compared to the increase in the number of classes. Fig. 5 shows the CMC obtained using option (iii).

In addition, we evaluated the three strategies for identity verification by using the lead X of the samples of DB_H_S . The first and second strategies obtained an EER of 3.81%, while the augmented dataset (third strategy) achieved EER of 3.37% (see Table 1). Furthermore, the features obtained using the augmented dataset can be used also for the different leads. For instance, for the lead Z, the third strategy achieved EER of 4.18% (Table 1), while the CNN trained with the lead X and tested with the lead Z (second strategy) achieved EER of 11.79%.

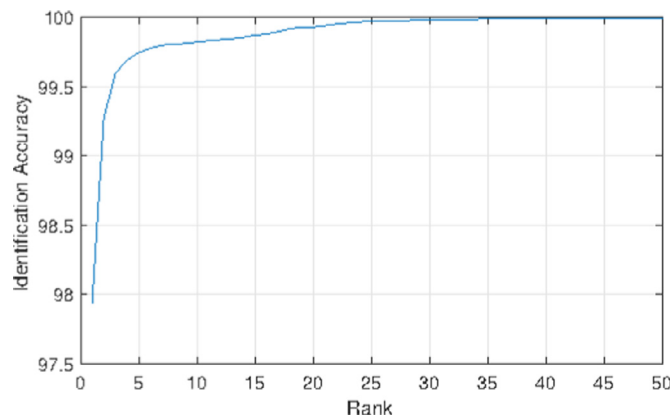


Fig. 5. CMC achieved using a training set including samples from three leads. Each lead is considered as a distinct trait.

Table 1
Verification accuracy of Deep-ECG for *DB_H_S*.

	Identity verification performance (EER)			
	Lead X	Lead Y	Lead Z	Fusion (X, Y, Z)
Deep-ECG-R	3.37%	4.31%	4.18%	1.36%
Deep-ECG-B	3.51%	4.86%	4.15%	1.05%
Correlation	8.16%	11.40%	9.01%	5.92%

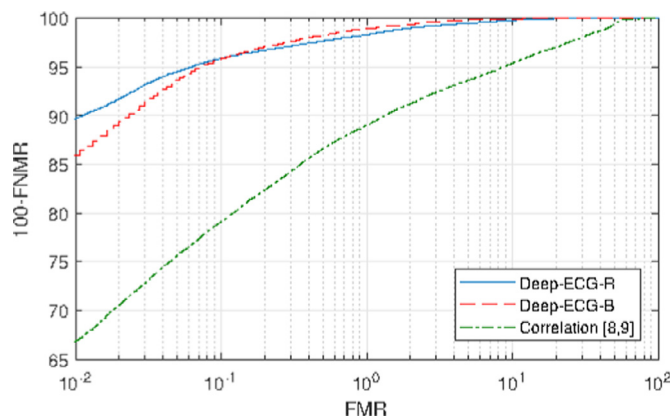


Fig. 6. ROC curves of Deep-ECG-R Fusion, Deep-ECG-B Fusion and a correlation-based method for *DB_H_S*. The configurations Deep-ECG-R and Deep-ECG-B achieved the best performance for all the points of the curve.

4.4. Identity verification accuracy

In this section, we evaluate the performance of Deep-ECG for identity verification. We evaluate the two kinds of templates proposed for Deep-ECG: real (Deep-ECG-R) and binary (Deep-ECG-B). We compared the obtained results with a recent correlation-based method [8,9].

The first experiment analyzes the performance of Deep-ECG with samples acquired within a short period of time, using *DB_H_S*. Table 1 and Fig. 6 present the results obtained by Deep-ECG in its two variants and the correlation-based method. Deep-ECG clearly outperformed the correlation-based method with both feature configurations and for every lead. Deep-ECG-R obtained slightly better performance for single leads. However, the binarization of the features still produced robust templates, which, in addition, achieved better performance using the fusion of three leads.

Table 2 and Fig. 7 present the results of Deep-ECG for identity verification with samples acquired after a longer period of time, from *DB_H_L*. We can observe a rise in EER, which is due to the increased difficulty of matching ECG signals acquired after a long

Table 2
Verification accuracy of Deep-ECG for *DB_H_L*.

	Identity verification performance (EER)			
	Lead X	Lead Y	Lead Z	Fusion (X, Y, Z)
Deep-ECG-R	5.81%	5.83%	5.68%	2.75%
Deep-ECG-B	5.55%	6.42%	5.76%	2.26%
Correlation	10.66%	13.15%	11.65%	5.95%

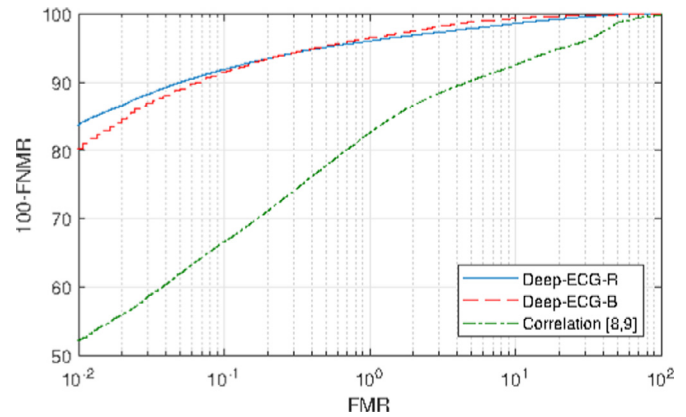


Fig. 7. ROC curves of Deep-ECG-R Fusion, Deep-ECG-B Fusion and a correlation-based method for *DB_H_L*. ROC curves of Deep-ECG-R Fusion, Deep-ECG-B Fusion and a correlation-based method for *DB_H_S*. The configurations Deep-ECG-R and Deep-ECG-B achieved the best performance for all the points of the curve.

Table 3
Re-authentication accuracy of Deep-ECG for *DB_H_C*.

	Periodic re-authentication performance (EER)			
	Lead X	Lead Y	Lead Z	Fusion (X, Y, Z)
Deep-ECG-R	5.79%	7.33%	6.25%	2.90%
Deep-ECG-B	5.41%	7.69%	6.15%	2.15%
Correlation	11.81%	16.57%	12.75%	8.31%

period of time. However, Deep-ECG is still capable of obtaining high performance, and its results do not worsen as much as the correlation-based method.

4.5. Periodic re-authentication accuracy

In this section, we analyze the performance of Deep-ECG in periodic re-authentication tasks using *DB_H_C*. In this test, we evaluated the capability of Deep-ECG of properly matching a wide set of genuine samples for each user (around 500 samples per users) acquired in uncontrolled conditions and performing heterogeneous activities in a time span of 8 h.

Table 3 presents the obtained results. As in previous studies [8,9], the results show that the periodic re-authentication accuracy is less than the one of identity verification performed in a short period of time because of the great variability of genuine samples. Nonetheless, Deep-ECG achieved EER equal to 2.90%, which is a satisfactory performance for many application scenarios. This result proves the robustness of the method to the various activities performed by people during the day. Moreover, the achieved accuracy is always better than that of the baseline method. In this context, periodic re-enrollment techniques [8] could further increase the accuracy.

Table 4

Identification accuracy of different methods for the PTB Diagnostic ECG Database.

Method	Subjects	Accuracy (%)
DWT [36]	90	97.7
Morphology + PCA [20]	20	95
Morphology [20]	29	100
Heartbeat selection [40]	74	98.8
Phase space [11]	100	99
Eigenpulse [19]	43	100
Matching pursuit [41]	20	95.3
Bag of Words [38]	100	99.48
Bag of Words [4]	100	99.72
Fiducial [5]	51	98.85
Non-fiducial [5]	51	99.39
Deep-ECG	52	100

4.6. Comparison with methods in the literature and applicability to different kinds of ECG signals

We compared the performance of Deep-ECG with other state-of-the-art methods for DB_PTB, which is composed of samples with different characteristics with respect to the Holter signals used to train the CNN. Similarly to most of the papers in the literature, we evaluated the accuracy of Deep-ECG for identification scenarios.

We adopted a fine-tuning strategy [26], by using the CNN trained for DB_H_T as a starting point. We replaced the last fully connected layer and the Soft-max layer, trained to classify samples from 276 individuals, with a new fully connected layer and a Soft-max layer performing a soft-max classification of 52 classes. Finally, we performed a fine tuning with a maximum of 300 training epochs.

Experiments have been performed using the same validation strategy adopted in [5]. Specifically, we divided the set of ECGs into a training set composed of 30% of the samples of each individual and a test set composed of 70% of the samples of each individual. We repeated the process 50 times by randomly selecting the samples and computed the mean identification error. Only lead X was considered. Deep-ECG achieved a mean Rank(1) error of 0.00%.

Table 4 compares the achieved results with that of Deep-ECG. All the compared methods have been evaluated using different numbers of leads, samples, users, and temporal durations. Furthermore, they have been validated using heterogeneous strategies. Nevertheless, we choose a widely used validation strategy, achieving accuracy in line with the best performing methods in the literature. In our opinion, this result is particularly relevant and proves the applicability of Deep-ECG for ECG signals collected in heterogeneous application scenarios.

4.7. Computational time

We executed the tests using a PC with 3.5GHz Intel (R) Core (TM) i7-7800X CPU, RAM 32 GB, GPU NVIDIA TITAN X (Pascal) with 12 GB of memory. The operating system was Windows 10 professional 64 bit. We implemented all the methods using Matlab 2017b and MatConvNet toolbox. The computational time needed to train the CNN was about 9 h and 22 min. The mean time to compute the vector V of QRS complexes was 0.65 s. The mean time to extract Deep-ECG-R features was 0.31 ms, and Deep-ECG-B features was 0.32 ms. The mean matching time for Deep-ECG-R features was 0.09 μ s, and for Deep-ECG-B features was 0.13 μ s. We believe that the computational cost permits to use the proposed approach in live biometric recognition systems deployed in real scenarios. Nevertheless, the computational time could be reduced using compiled languages, such as C or C++, because Matlab is a prototype-oriented and non-optimized environment.

5. Conclusion

This work introduced an ECG-based biometric recognition approach, called Deep-ECG. This approach is based on deep Convolutional Neural Networks and can be used in three common biometric tasks: closed-set identification, identity verification and periodic re-authentication. Deep-ECG analyzes sets of QRS complexes extracted from ECG signals, and produces a set of features extracted using a deep CNN. Deep-ECG can extract two kinds of templates: real and binary, which are matched using the Euclidean and Hamming distances, respectively.

To evaluate the performance of Deep-ECG, we tested it using wide sets of samples acquired in uncontrolled conditions for closed-set identification, identity verification and periodic re-authentication. In all cases, Deep-ECG obtained remarkable accuracy, using real as well binary templates. Results proved the feasibility of the proposed approach for samples acquired in non-ideal conditions.

Furthermore, we evaluated the performance of Deep-ECG with samples acquired in scenarios different from the one used to train the CNN (ambulatory at rest vs. Holter during real life activity), obtaining identification accuracy better or comparable to the best performing methods in the literature.

Future work should regard the design of more complex matchers that permit to increase the recognition accuracy. In addition, the use of Deep-ECG binary features to implement template protection methods should be analyzed in more detail.

Acknowledgments

This work was funded in part by European Union within Integrated Project “ABC gates for Europe” (FP7-SEC-2012-312797), Italian Ministry of Research within PRIN project “COSMOS” (201548C5NY).

We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan X Pascal GPU used for this research, within the project “Deep Learning and CUDA for advanced and less-constrained biometric systems”.

Data used for this research were provided by the Telemetric and Holter ECG Warehouse of the University of Rochester (THEW), NY.

References

- [1] F. Agraftioti, D. Hatzinakos, Fusion of ECG sources for human identification, in: 3rd International Symposium on Communications, Control and Signal Processing, 2008 (ISCCSP 2008), 2008, pp. 1542–1547.
- [2] A. Bonissi, R. Donida Labati, L. Perico, R. Sassi, F. Scotti, L. Sparagino, A preliminary study on continuous authentication methods for photoplethysmographic biometrics, in: Proceedings of the 2013 IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications, 2013, pp. 28–33.
- [3] I.B. Ciocoiu, ECG biometrics using bag-of-words models, in: 2015 International Symposium on Signals, Circuits and Systems (ISSCS), 2015, pp. 1–4.
- [4] I.B. Ciocoiu, Comparative analysis of bag-of-words models for ECG-based biometrics, IET Biom. 6 (6) (2017) 495–502.
- [5] D.P. Coutinho, H. Silva, H. Gamboa, A. Fred, M. Figueiredo, Novel fiducial and non-fiducial approaches to electrocardiogram-based biometric systems, IET Biom. 2 (2) (2013) 64–75.
- [6] J. Daugman, Information theory and the iriscodes, IEEE Trans. Inf. Forensics Secur. 11 (2) (2016) 400–409.
- [7] R. Donida Labati, V. Piuri, R. Sassi, F. Scotti, Heartcode: a novel binary ECG-based template, in: Proceedings of the IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications, 2014, pp. 86–91.
- [8] R. Donida Labati, V. Piuri, R. Sassi, G. Sforza, F. Scotti, Adaptive ECG biometric recognition: a study on re-enrollment methods for QRS signals, in: Proceedings of the IEEE Workshop on Computational Intelligence in Biometrics and Identity Management, 2014, pp. 30–37.
- [9] R. Donida Labati, R. Sassi, F. Scotti, ECG biometric recognition: permanence analysis of QRS signals for 24 hours continuous authentication, in: Proceedings of the IEEE International Workshop on Information Forensics and Security, 2013, pp. 31–36.
- [10] A. Eduardo, H. Aidos, A.L.N. Fred, ECG-based biometrics using a deep autoencoder for feature learning: an empirical study on transferability, in: Proceedings of the INSTICC International Conference on Pattern Recognition Applications and Methods (ICPRAM), 2017, pp. 463–470.

- [11] S.-C. Fang, H.-L. Chan, Human identification by quantifying similarity and dissimilarity in electrocardiogram phase space, *Pattern Recognit.* 42 (9) (2009) 1824–1831.
- [12] A. Gangwar, A. Joshi, Deepirisnet: deep iris representation with applications in iris recognition and cross-sensor iris recognition, in: *Proceedings of the IEEE International Conference on Image Processing*, 2016, pp. 2301–2305.
- [13] O. Ghahabi, J. Hernando, Deep belief networks for i-vector based speaker recognition, in: *Proceedings of the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 1700–1704.
- [14] A.L. Goldberger, L.A.N. Amaral, L. Glass, J.M. Hausdorff, P.C. Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, C.-K. Peng, H.E. Stanley, Physiobank, physiotoolkit, and physionet, *Circulation* 101 (23) (2000) e215–e220.
- [15] S. Hari, F. Agraftoti, D. Hatzinakos, Design of a Hamming-distance classifier for ECG biometrics, in: *Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013, pp. 3009–3012.
- [16] L.S. Hou, K. Subari, S. Syahril, QRS-complex of ECG-based biometrics in a two-level classifier, in: *Proceedings of the 2011 IEEE Region 10 Conf-(TENCON)*, 2011, pp. 1159–1163.
- [17] M. Huanhuan, Z. Yue, Classification of electrocardiogram signals with deep belief networks, in: *Proceedings of the IEEE International Conference on Computational Science and Engineering*, 2014, pp. 7–12.
- [18] J. Irvine, S. Israel, A sequential procedure for individual identity verification using ECG, *EURASIP J. Adv. Signal Process.* 2009 (1) (2009) 243215.
- [19] J.M. Irvine, S.A. Israel, W.T. Scruggs, W.J. Worek, Eigenpulse: robust human identification from cardiovascular function, *Pattern Recognit.* 41 (11) (2008) 3427–3435.
- [20] S.A. Israel, J.M. Irvine, A. Cheng, M.D. Wiederhold, B.K. Wiederhold, ECG To identify individuals, *Pattern Recognit.* 38 (1) (2005) 133–142.
- [21] A. Jain, P. Flynn, A. Ross, *Handbook of Biometrics*, Springer, 2007.
- [22] W. Khalifa, A. Salem, M. Roushdy, K. Revett, A survey of EEG based user authentication schemes, in: *Proceedings of the 8th International Conference on Informatics and Systems (INFOS)*, 2012, pp. 55–60.
- [23] S. Kiranyaz, T. Ince, M. Gabbouj, Real-time patient-specific ECG classification by 1-d convolutional neural networks, *IEEE Trans. Biomed. Eng.* 63 (3) (2016) 664–675.
- [24] S. Kiranyaz, T. Ince, R. Hamila, M. Gabbouj, Convolutional neural networks for patient-specific ECG classification, in: *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2015, pp. 2608–2611.
- [25] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, in: F. Pereira, C.J.C. Burges, L. Bottou, K.Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems*, vol. 25, Curran Associates, Inc., 2012, pp. 1097–1105.
- [26] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, *Nature* 521 (7553) (2015) 436–444.
- [27] W. Louis, M. Komeili, D. Hatzinakos, Continuous authentication using one-dimensional multi-resolution local binary patterns (1dmrlbp) in ECG biometrics, *IEEE Trans. Inf. Forensics Secur.* 11 (12) (2016) 2818–2832.
- [28] M. Merone, P. Soda, M. Sansone, C. Sansone, ECG Databases for biometric systems: a systematic review, *Expert Syst. Appl.* 67 (2017) 189–202.
- [29] S. Min, B. Lee, S. Yoon, Deep learning in bioinformatics, *Brief Bioinf.* 2009 (1) (2016) 243215.
- [30] I. Odinaka, P.-H. Lai, A. Kaplan, J. O'Sullivan, E. Sirevaag, J. Rohrbaugh, ECG Biometric recognition: a comparative analysis, *IEEE Trans. Inf. Forensics Secur.* 7 (6) (2012) 1812–1824.
- [31] O.M. Parkhi, A. Vedaldi, A. Zisserman, Deep face recognition, in: *British Machine Vision Conference*, 2015.
- [32] B. Pourbabaee, M.J. Roshkhari, K. Khorasani, Feature learning with deep convolutional neural networks for screening patients with paroxysmal atrial fibrillation, in: *Proceedings of the International Joint Conference on Neural Networks*, 2016, pp. 5057–5064.
- [33] M.A. Rahhal, Y. Bazi, H. AlHichri, N. Alajlan, F. Melgani, R. Yager, Deep learning approach for active classification of electrocardiogram signals, *Inf. Sci.* 345 (C) (2016) 340–354.
- [34] F. Richardson, D. Reynolds, N. Dehak, Deep neural network approaches to speaker and language recognition, *IEEE Signal Process. Lett.* 22 (10) (2015) 1671–1675.
- [35] A.A. Ross, K. Nandakumar, A.K. Jain, *Handbook of Multibiometrics*, vol. 6, Springer Science & Business Media, 2006.
- [36] M.M. Tantawi, K. Revett, A.-B. Salem, M.F. Tolba, A wavelet feature extraction method for electrocardiogram (ECG)-based biometric recognition, *Signal Image Video Process.* 9 (6) (2015) 1271–1280.
- [37] University of Rocher Medical Center, Telemetric and Holter ECG Warehouse, E-hol-03-0202-003, <http://thew-project.org/Database/E-HOL-03-0202-003.html>.
- [38] J. Wang, M. She, S. Nahavandi, A. Kouzani, Human identification from ECG signals via sparse representation of local segments, *IEEE Signal Process. Lett.* 20 (10) (2013) 937–940.
- [39] Y. Wang, F. Agraftoti, D. Hatzinakos, K.N. Plataniotis, Analysis of human electrocardiogram for biometric recognition, *EURASIP J. Adv. Signal Process.* 2008 (1) (2008) 148658.
- [40] G. Wbbeler, M. Stavridis, D. Kreiseler, R.-D. Boussetlot, C. Elster, Verification of humans using the electrocardiogram, *Pattern Recognit. Lett.* 28 (10) (2007) 1172–1175.
- [41] Z. Zhao, L. Yang, ECG identification based on matching pursuit, in: *4th International Conference on Biomedical Engineering and Informatics*, vol. 2, 2011, pp. 721–724.