ECG-BASED BIOMETRICS USING RECURRENT NEURAL NETWORKS

Ronald Salloum and C.-C. Jay Kuo

Ming Hsieh Department of Electrical Engineering
University of Southern California, Los Angeles, CA
(rsalloum, jckuo)@usc.edu

ABSTRACT

In this paper, we propose the use of recurrent neural networks (RNNs) to develop an effective solution to two problems in electrocardiogram (ECG)-based biometrics: identification/classification and authentication. Different RNN architectures with various parameter settings were evaluated, including traditional, long short-term memory (LSTM), gated recurrent unit (GRU), unidirectional, and bidirectional networks. Unlike many existing methods, the RNN-based method does not require any feature extraction. The method was evaluated using two publicly available datasets: ECG-ID and MIT-BIH Arrhythmia (MITDB). For the identification problem, nearly 100% classification accuracy on the ECG-ID dataset was achieved, and similar results were observed for the MITDB dataset. For the authentication problem, an RNN was trained and the hidden state at the final time step was extracted to make a decision. We evaluated the effect of the training size on the equal error rate (EER), and showed that the EER drops from approximately 3.5% to nearly 0% as we increased the percentage of subjects used for training from approximately 15% to 80%.

Index Terms— RNN, ECG, Biometrics, Identification, Authentication

1. INTRODUCTION

Electrocardiogram (ECG)-based biometric systems have been used for two different applications: identification and authentication. For identification, the system is given input data and must output the identity of the unknown subject. On the other hand, in an authentication scenario, there is a claimed identity associated with the input data and the system must either accept or reject the claimed identity. There are numerous publications in the ECG identification and authentication literature, and a wide variety of machine learning and pattern recognition techniques have been utilized.

In this paper, we propose the use of recurrent neural networks (RNNs) to develop an effective solution to the two problems in ECG-based biometrics. Unlike many existing methods, the RNN-based method does not require any feature extraction. The ECG data is directly fed to the RNN. To the best of our knowledge, our study is the first one that applies RNNs to ECG-based biometrics. The performance of the proposed method is evaluated using two public datasets from the Physionet database, and it is shown that it outperforms existing methods on both datasets.

The rest of this paper is organized as follows. Related previous work is reviewed in Section 2. The RNN-based method is described in Section 3. Experimental results are presented in Section 4. Finally, concluding remarks are given in Section 5.

2. RELATED WORK

A review on existing methods for ECG-based biometrics can be found in [1], and more recently in [2]. They can be roughly grouped into two main categories based on the type of feature extraction performed – fiducial and non-fiducial methods. Fiducial methods require the identification of certain characteristic or anchor points on the ECG recordings while non-fiducial methods do not use characteristic points to generate the feature set. Some non-fiducial methods use autocorrelation, Fourier or wavelet coefficients as features while others use the raw heartbeat waveforms as feature vectors. There are also hybrid methods that combine fiducial and non-fiducial features. The existing methods can also be grouped based on their classification methods. Common classification methods include: k nearest neighbors, nearest center, linear discriminant analysis (LDA), feedforward neural networks, generative model classifiers (GMCs), support vector machines (SVMs), and match/similarity score classifiers.

Although a wide variety of techniques have been proposed, it is difficult to compare methods because their performance was reported on different datasets. Odinaka *et al.* [1] implemented and compared existing methods using a private dataset of recordings from 265 subjects. This comparative study indicated that most algorithms perform well when the training and testing data for a given subject come from the same session (referred to as within-session analysis). However, when training and testing data come from different sessions (referred to as across-session analysis), performance degradation occurs. For across-session analysis with training from multiple sessions, the top two methods had an equal error rate (EER) of 5.47% and 6.28%, while the average EER of the other methods was 20.75% (EER is the error rate at which the false acceptance and false rejection rates are equal). A more recent review paper [2] selected 18 journal publications, and computed a weighted average of classification accuracy and EER in identification and authentication scenarios, respectively. Each study's performance was weighted by the number of subjects used in the study against the total number of subjects of the selected journal publications. They found that the weighted average classification accuracy and EER is 94.95% and 0.92%, respectively.

Several authors have used feedforward neural networks for ECG-based biometrics [2]. Unlike feedforward networks, RNNs have feedback connections that make them suitable for processing sequential data such as an ECG signal. For applications in which there are long-term dependencies, traditional RNNs are difficult to train because they suffer from vanishing and exploding gradients [3]. The exploding gradient problem is typically addressed by clipping gradients whose norms exceed a threshold, which is known as gradient clipping [4]. Networks that use long short-term memory (LSTM) units [5] or gated recurrent units (GRUs) [6] have been proposed to address the vanishing gradient problem. LSTM-based networks have recently been used in a wide range of applications, including phoneme classification [7], handwriting recognition [8], speech recognition [9], and speaker verification [10]. Dropout, which is a regularization technique used to reduce overfitting, can also be utilized in RNNs. Zaremba *et al.* [11] proposed a method in which dropout is applied only to the non-recurrent connections. Bidirectional RNNs [7], which can be used to incorporate future context in addition to past context, have also been utilized in certain applications.

3. PROPOSED RNN METHOD

3.1. Preprocessing and Segmentation

The ECG recordings used in this work are from the publicly available ECG-ID [12] and MIT-BIH Arrhythmia (MITDB) [13] datasets, which are part of the Physionet database [14]. Analysis was performed separately for the two datasets. Given an ECG recording, the first step is to segment the recording into individual heartbeat waveforms. Since the R peak is the most prominent peak, it can be used as a marker of a given heartbeat waveform. The R peaks were detected using the Pan-Tompkins algorithm [15]. Once the peaks are detected, a certain number of samples before and after a given R peak are concatenated, forming a vector which represents the heartbeat waveform. For the ECG-ID dataset, 150 samples before/after the R peak were selected, while 125 samples were selected for the MITDB dataset (please note that the two datasets have different sampling rates, as discussed in Section 3.4). After the segmentation procedure is completed, each individual heartbeat waveform is z-score standardized. Finally, a certain number of consecutive heartbeat waveforms are grouped to form a given input sequence, where the number of heartbeats in the input sequence is a hyperparameter.

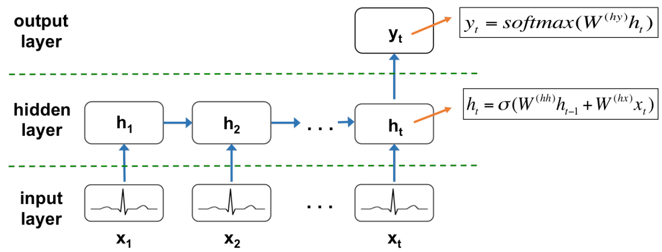

Fig. 1. Block diagram of a traditional RNN applied to an input sequence of t heartbeats $(x_1 x_2 ... x_t)$ from a given subject. The hidden state and output at the final time step are denoted as h_t and y_t, respectively. The weight matrices $W^{(hh)}$, $W^{(hx)}$, and $W^{(hy)}$ are parameters optimized during training, and σ represents an element-wise non-linearity such as $tanh$ or ReLU.

3.2. Identification Procedure

For the identification scenario, the ECG recordings are divided into training and testing sets. The division depends on the type of analysis performed. For within-session analysis, training and testing data for a given subject are obtained from the same recording or session. In contrast, for across-session analysis, training and testing data for a given subject are obtained from different recordings. Each training or testing sequence is of size $N \times D$, where N is the number of heartbeats in a given sequence and D is the dimension of each heartbeat waveform. One-hot encoding is used for the sequence labels. Specifically, if a given sequence is from the i^{th} subject, then the corresponding label is an M-length vector, where M is the total number of subjects, and the j^{th} element of this vector is given by $z_j = \begin{cases} 1 & j = i \\ 0 & j \neq i \end{cases}$. The parameters or weights of an RNN with a given architecture are optimized using the training set. The hidden state at the final time step, or summary vector, is multiplied by a weight matrix (optimized during the training process) and fed to the softmax function to yield a probability distribution over the set of subjects (referred to as class probabilities). Once the RNN has been successfully trained, the weights are fixed, and testing sequences are fed to the trained network. A classification decision for each testing sequence is made by selecting the class with the highest assigned probability. A block diagram illustrating how a traditional RNN is used to classify a sequence of heartbeats is shown in Figure 1. In a traditional RNN, the hidden state at a given time step is computed as a linear combination of the previous hidden state and the current input. GRU and LSTM networks have similar block diagrams, however the update of the hidden state is more complex.

3.3. Authentication Procedure

For the authentication scenario, the ECG recordings are divided into three datasets: training, enrollment, and evaluation. The subjects used during training are different than those used during enrollment/evaluation. Furthermore, the subjects in the enrollment and evaluation sets are identical. Training of the RNN is performed in the way described in the previous subsection, with the softmax layer present. After the RNN has been successfully trained, the softmax layer is removed. The testing phase consists of two stages: enrollment and evaluation. In the enrollment phase, the trained network is used to extract the summary vector (i.e. the hidden state at the final time step) from each enrollment sequence. For each subject, there are multiple enrollment sequences, each producing a different summary vector, and the enrollment model of a given subject is taken to be the average of these vectors. The subjects in the evaluation set are identical to those in the enrollment set, except new sequences are fed to the RNN. An evaluation sequence from a given subject is fed to the network, yielding a summary vector. This summary vector is compared to each enrollment model (one for each subject in the enrollment/evaluation set) using a similarity metric (e.g. cosine distance), and an authentication decision for each comparison is made based on a given threshold. The claimed identity is accepted or rejected depending on whether the distance is less than or greater than the given threshold, respectively. The advantage of the summary vector approach is that the network can be trained using a given number of subjects, and testing (enrollment/evaluation) can be performed on any other subjects that were not necessarily part of the training set. In this method, the trained RNN can be viewed as a feature extractor, in which the summary vector represents the extracted feature vector; similar techniques have been used for speaker verification [10].

3.4. Dataset and Implementation

The ECG-ID dataset contains 310 recordings, obtained from 90 subjects (44 male and 46 female). Each recording is ECG lead I, recorded over a duration of 20 seconds, and digitized at 500 Hz with 12-bit resolution over a nominal $\pm 10\,mV$ range. Two recordings per subject were used for the analysis, because only a small subset of the subjects have more than 2 recordings. The MITDB dataset contains 48 two-channel recordings, obtained from 47 subjects (25 male and 22 female). Each subject has only one recording available, except for one subject that has two recordings (records 201 and 202); only record 201 was used for this subject. The recordings were digitized at 360 samples per second per channel with 11-bit resolution over a $10mV$ range. Only the upper signal was used because QRS complexes are usually prominent in the upper signal; a notable exception is record 114, for which the signals are reversed. Both the ECG-ID and MITDB datasets contain pre-filtered data, which was used in this study.

R peak detection and generation of the training and testing data were performed in Matlab, while implementation and training of RNNs were performed using TensorFlow [16]. The cost function used during training is the cross-entropy error, and optimization was performed using the Adam algorithm [17] with a learning rate of 0.0001. We evaluated different RNN architectures with different parameter settings.

Table 1. ECG-ID Within-Session Analysis (Identification): Classification accuracy for selected parameter settings.

Type of cell/unit	Input sequence length (in number of beats)	Number of hidden layers	Classification Accuracy
Traditional	3	1	92.8%
Traditional	9	1	93.3%
GRU	3	1	95.2%
GRU	9	1	96.7%
LSTM	3	1	98.2%
LSTM	3	2	97.8%
LSTM	9	1	100%

The architectures we tested are LSTM, GRU, and traditional RNNs, and we tested both unidirectional and bidirectional architectures. The RNN parameters are the number of layers, number of cells or units per layer, and dropout rate. Hyperparameter optimization was performed using random search.

For the identification problem, the classification accuracy was computed using the testing set. A classification decision is made for each testing sequence, and the classification accuracy is defined to be the percentage of correctly classified testing sequences. For the authentication scenario, the equal error rate (EER) was computed in the following manner. Given a summary vector obtained from the evaluation data of a particular subject, this vector is compared to each enrollment model (one model for each subject in the enrollment/evaluation set) using cosine distance as a similarity metric. An authentication decision for each comparison is then made based on a given threshold. Thus, for each summary vector obtained from the evaluation dataset, there is a certain number of each of the following quantities: true acceptance, true rejection, false acceptance, false rejection. Each of these 4 quantities is summed across the set of summary vectors obtained from the evaluation dataset, and the false acceptance rate (FAR) and false rejection rate (FRR) are computed. The threshold is then varied to yield a plot of FAR and a plot of FRR, and the EER is taken to be the intersection of these two curves.

4. EXPERIMENTAL RESULTS

4.1. Identification Analysis

Several existing methods have used the MITDB and ECG-ID datasets. Of those that used the MITDB dataset, the reported classification accuracy ranged from 93.1% to 99.57% (fusion of two leads) [18, 19, 20], and of those that used the ECG-ID dataset, the reported accuracy ranged from 82.3% to 96% [12, 19]. The RNN-based method outperforms these existing methods on both datasets. Both within-session and across-session analysis were performed; for the MITDB dataset, only within-session analysis was performed (only one recording per subject was available). In MITDB within-session analysis, 18 training and 18 testing beats per subject were used. In the case of ECG-ID within-session analysis, for a given subject, 9 training and 9 testing beats from each session were

Table 2. ECG-ID Across-Session Analysis (Identification): Classification accuracy for selected parameter settings.

Type of cell/unit	Input sequence length (in number of beats)	Number of hidden layers	Classification Accuracy
Traditional	3	1	91.1%
Traditional	9	1	91.7%
GRU	3	1	93.5%
GRU	9	1	94.4%
LSTM	3	1	97.0%
LSTM	3	2	96.7%
LSTM	9	1	100%

Table 3. MITDB Within-Session Analysis (Identification): Classification accuracy for selected parameter settings.

Type of cell/unit	Input sequence length (in number of beats)	Number of hidden layers	Classification Accuracy
Traditional	3	1	93.3%
Traditional	9	1	93.6%
GRU	3	1	95.7%
GRU	9	1	96.8%
LSTM	3	1	98.6%
LSTM	3	2	97.9%
LSTM	9	1	100%

used, for a total of 18 training and 18 testing beats per subject. In ECG-ID across-session analysis, for a given subject, 18 training beats were taken from one session and 18 testing beats were taken from the other session.

Tables 1-3 show the classification accuracy for selected RNN architectures and parameter settings. Please note that the results shown are for unidirectional networks with zero dropout. The results show that 100% classification accuracy is achieved on both datasets, using a unidirectional LSTM network with 1 hidden layer (250 hidden units) and zero dropout. As the results indicate, LSTM networks performed better than GRU and traditional RNNs. Furthermore, it can be seen that as the length of the input sequence increases, the improvement in performance is more significant for LSTM networks, as compared to GRU and traditional RNNs. In addition, we found that bidirectional architectures, dropout, and multiple layers did not significantly improve performance. Also, feedforward neural network architectures with different parameter settings were evaluated, and the classification accuracy was slightly lower than that of traditional RNNs.

4.2. Authentication Analysis

For the ECG-ID dataset, 9 beats were taken from each session for a given subject in the training set, for a total of 18 training beats per subject, while for each subject in the enrollment/evaluation set, 18 enrollment beats were taken from one session and 18 evaluation beats were taken from the other session. For the MITDB dataset, 18 beats were taken for a given subject in the training set, and for each subject in the enrollment/evaluation set, 18 beats were taken for enrollment and 18 beats were taken for evaluation. We observed patterns

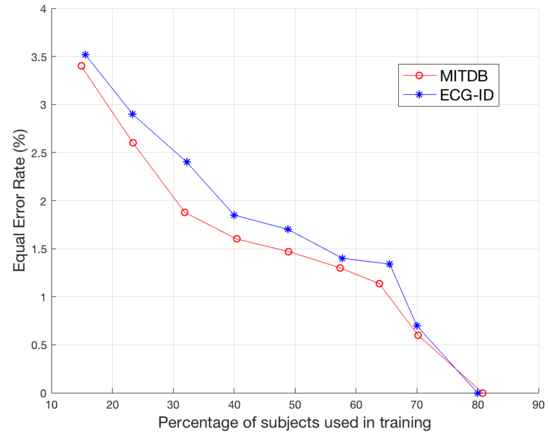

Fig. 2. Authentication analysis: Equal error rate (%) as a function of the percentage of subjects used in training, for the ECG-ID and MITDB datasets.

similar to those discussed in Section 4.1. In addition, we evaluated the effect of varying the percentage of subjects used for training. Figure 2 shows the EER as a function of the percentage of subjects used for training, for both the ECG-ID and MITDB datasets. The results shown are for input sequences consisting of 9 heartbeats. The figure shows that the EER decreases as the training size increases, and furthermore, that the proposed method is able to achieve 0% EER when the percentage of subjects used for training is approximately 80%.

5. CONCLUSION

This paper has demonstrated that an LSTM-based RNN is a more effective tool for ECG-based biometric identification and authentication, as compared to existing methods that used the ECG-ID or MITDB datasets. For both applications, the proposed method does not require extraction of fiducial features nor features such as autocorrelation and wavelet coefficients; the ECG data is directly fed to the RNN. For the identification problem, an RNN is trained as a classifier, and 100% classification accuracy was achieved for both datasets. For the authentication scenario, a summary vector approach was utilized. The primary advantage of this technique is that authentication can be performed on any number of subjects not seen during training. We evaluated the effect of the training size on the EER, and found that the EER drops to 0% when the percentage of subjects used for training is increased to approximately 80%. The results on the MITDB dataset, which contains recordings of subjects with arrhythmia, suggest that the proposed method is robust to some abnormal cardiac conditions. A more extensive analysis is needed to evaluate the robustness of the proposed method to other abnormal cardiac conditions, as well as different physiological conditions. This work indicates that LSTM-based RNNs are a promising direction for both ECG-based biometric identification and authentication.

6. REFERENCES

[1] Ikenna Odinaka, Po-Hsiang Lai, Alan D Kaplan, Joseph A O'Sullivan, Erik J Sirevaag, and John W Rohrbaugh, "Ecg biometric recognition: A comparative analysis," *Information Forensics and Security, IEEE Transactions on*, vol. 7, no. 6, pp. 1812–1824, 2012.

[2] Antonio Fratini, Mario Sansone, Paolo Bifulco, and Mario Cesarelli, "Individual identification via electrocardiogram analysis," *Biomedical engineering online*, vol. 14, no. 1, pp. 78, 2015.

[3] Yoshua Bengio, Patrice Simard, and Paolo Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE transactions on neural networks*, vol. 5, no. 2, pp. 157–166, 1994.

[4] Razvan Pascanu, Tomas Mikolov, and Yoshua Bengio, "On the difficulty of training recurrent neural networks.," *ICML (3)*, vol. 28, pp. 1310–1318, 2013.

[5] Sepp Hochreiter and Jürgen Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[6] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio, "Learning phrase representations using rnn encoder-decoder for statistical machine translation," *arXiv preprint arXiv:1406.1078*, 2014.

[7] Alex Graves and Jürgen Schmidhuber, "Framewise phoneme classification with bidirectional lstm and other neural network architectures," *Neural Networks*, vol. 18, no. 5, pp. 602–610, 2005.

[8] Alex Graves and Jürgen Schmidhuber, "Offline handwriting recognition with multidimensional recurrent neural networks," in *Advances in neural information processing systems*, 2009, pp. 545–552.

[9] Alan Graves, Abdel-rahman Mohamed, and Geoffrey Hinton, "Speech recognition with deep recurrent neural networks," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*. IEEE, 2013, pp. 6645–6649.

[10] Georg Heigold, Ignacio Moreno, Samy Bengio, and Noam Shazeer, "End-to-end text-dependent speaker verification," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2016, pp. 5115–5119.

[11] Wojciech Zaremba, Ilya Sutskever, and Oriol Vinyals, "Recurrent neural network regularization," *arXiv preprint arXiv:1409.2329*, 2014.

[12] Tatiana S Lugovaya, "Biometric human identification based on ecg," 2005.

[13] George B Moody and Roger G Mark, "The impact of the mit-bih arrhythmia database," *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, no. 3, pp. 45–50, 2001.

[14] Ary L Goldberger, Luis AN Amaral, Leon Glass, Jeffrey M Hausdorff, Plamen Ch Ivanov, Roger G Mark, Joseph E Mietus, George B Moody, Chung-Kang Peng, and H Eugene Stanley, "Physiobank, physiotoolkit, and physionet components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[15] Jiapu Pan and Willis J Tompkins, "A real-time qrs detection algorithm," *IEEE transactions on biomedical engineering*, , no. 3, pp. 230–236, 1985.

[16] Martın Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, et al., "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," *arXiv preprint arXiv:1603.04467*, 2016.

[17] Diederik Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[18] Can Ye, Miguel Tavares Coimbra, and BVK Vijaya Kumar, "Investigation of human identification using two-lead electrocardiogram (ecg) signals," in *Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on*. IEEE, 2010, pp. 1–8.

[19] Muhammad Najam Dar, M Usman Akram, Anam Usman, and Shoab A Khan, "Ecg biometric identification for general population using multiresolution analysis of dwt based features," in *2015 Second International Conference on Information Security and Cyber Forensics (InfoSec)*. IEEE, 2015, pp. 5–10.

[20] Khairul A Sidek, Ibrahim Khalil, and Herbert F Jelinek, "Ecg biometric with abnormal cardiac conditions in remote monitoring system," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 44, no. 11, pp. 1498–1509, 2014.