

ECG Biometrics using RNN and CNN

David Belo (✉ dj.belo@fct.unl.pt)

Universidade Nova de Lisboa Faculdade de Ciencias e Tecnologia <https://orcid.org/0000-0002-5337-0430>

Nuno Bento

Universidade Nova de Lisboa Faculdade de Ciencias e Tecnologia

Hugo Silva

Instituto de Telecomunicacoes

Ana Fred

Instituto de Telecomunicacoes

Hugo Gamboa

Universidade Nova de Lisboa Faculdade de Ciencias e Tecnologia

Research

Keywords: Pattern Recognition, Deep Neural Networks, Biosignals, Electrocardiogram, Biometry, Identification, Authentication, Machine learning

Posted Date: January 30th, 2020

DOI: <https://doi.org/10.21203/rs.2.22270/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

RESEARCH

ECG Biometrics using RNN and CNN

David Belo^{1*}
, Nuno Bento¹
, Hugo Silva²
, Ana Fred²
and Hugo Gamboa¹

*Correspondence:

dj.belo@fct.unl.pt

¹LIBPhys - Nova University of
Lisbon, Departamento de Física -
Faculdade de Ciências e
Tecnologia - Universidade Nova de
Lisboa, Caparica, Portugal
Full list of author information is
available at the end of the article

Abstract

Background: Biometric Systems (BS) are based on a pattern recognition problem where the individual traits of a person are coded and compared. The Electrocardiogram (ECG) as a biometric emerged, as it fulfills the requirements of a BS.

Methods: Inspired by the high performance shown by Deep Neural Networks (DNN), this work proposes two architectures to improve current results in both identification and authentication: Temporal Convolutional Neural Network (TCNN) and Recurrent Neural Network (RNN). The last two results were submitted to a simple classifier, which exploits the error of prediction of the former and the scores given by the last.

Results: The robustness and applicability of these architectures were tested on Fantasia, MIT-BIH and CYBHi databases. The TCNN outperforms the RNN achieving 100%, 96% and 90% of accuracy, respectively, for identification and 0.0%, 0.1% and 2.2% equal error rate for authentication. When comparing to previous work, both architectures reached results beyond the state-of-the-art.

Conclusions:

Keywords: Pattern Recognition; Deep Neural Networks; Biosignals; Electrocardiogram; Biometry; Identification; Authentication; Machine learning

1 Introduction

Some traditional methods of identification and authentication, such as key-possession or username-password, may prove ineffective against terrorist and criminal acts [1]. Key areas of society demand reliable solutions for border control, criminal identification, and electronic transactions. Consequently, the use of individual traits, i.e. biometrics, is trending in both private and public sectors. A report from Acuity Market Intelligence in 2015 expected the biometrics market to grow not only in the USA and Europe but also in Asia and Africa. These numbers are prospected to increase from USD 10.74 Billion in 2015 to USD 32.73 Billion by 2022, reflecting the need for further developing these systems [2, 3].

1.1 Biometrics

Biometrical authentication (prove that the person is whom it claims) and identification (finding the registered person from sample) is a pattern recognition problem where physical or behavioral traits are submitted to a feature extraction module,

during the enrollment stage, and compared to the biometrical database of a population, during the verification phase [1, 4]. Also, the evaluation of a Biometric Systems (BS) is made in the dimensions of: universality; uniqueness; permanence; measurability; performance; acceptability, and; circumvention [5]. Given the success in fulfilling all these premises, the Electrocardiogram (ECG) morphological signature has emerged in the last decade, with the added effect in the difficulty of reproducing this signal.

1.2 ECG Biometrics

The ECG, representing the electrical activity of the heart, has a cyclic structure. Each cycle, corresponding to a heartbeat, presents a characteristic morphology composed by: (1) P wave - Contraction of the atria; (2) PQ interval - Time between the contraction of the atria and activation of the ventricles; (3) QRS complex - The combination of the Q, R and S waves that is associated to the contraction of the ventricles; (4) ST interval - Relaxation after the ventricles contraction; (5) T wave - repolarization of the ventricles in preparation for the next cycle of contractions [6]. The main challenge of non-invasive ECG measurements is the intra-subject variability due to the corruption of the signal (artifacts and noise) [6, 7], alterations to the physical state, affective status and drug effects[6–8].

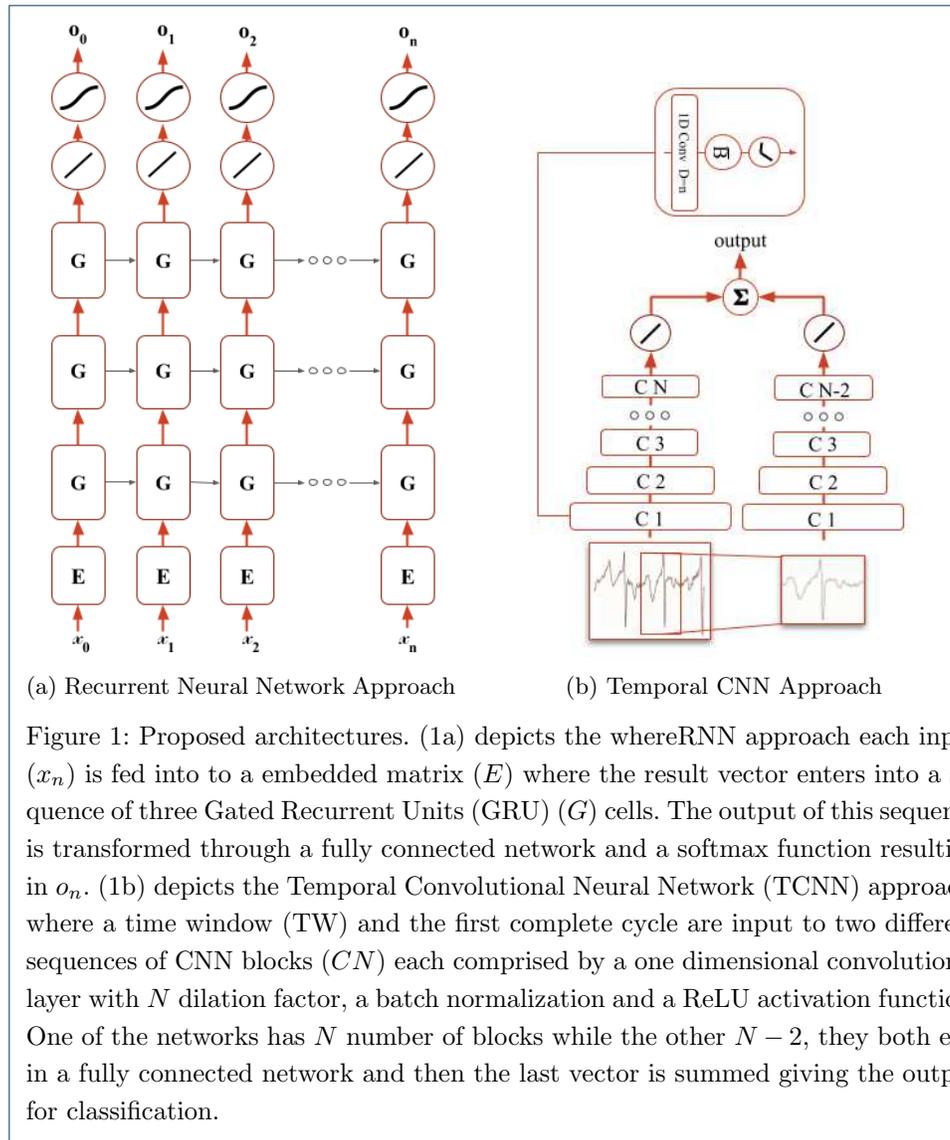
There are two main types of ECG biometric systems: fiducial-based, in which heartbeat templates are sampled from the ECG signals and have their relevant points and segments computed, and; non-fiducial-based, which considers a ECG signal as a whole and computes the features on this premise. Some studies use the combination of both methods, called partially-fiducial, to achieve better results [6, 9, 10].

1.3 Objective

In light of the current trend in Deep Neural Networks (DNN) and as a way to improve results in both identification and authentication without extracting human-crafted features, two Neural Networks (NN) architectures are proposed (Fig. ??). The first one is a direct application of a previous work [11], where we propose a non-fiducial system that uses a Recurrent Neural Networks (RNN) capable synthesizing ECG signals. The biometric solution is related to the hypothesis that if a model that is trained with the signal from person A, the prediction error can be used as a score, which will be higher when fed with signals for person B and C.

The second system is partial-fiducial and exploits the automatic feature learning ability of a Convolutional Neural Networks (CNN). This model outputs a score based on the probability function given by the network. Both scores are submitted to the same simple classifier, the Relative Score Threshold Classifier (RSTC) which compares the scores and finds the most probable source for the given sample or batch of samples.

Both methods will be tested in three well-known public databases: Fantasia, MIT-BIH and CYBHi [12–15]. This document is organized as follows: Section 2 addresses the state-of-the-art in ECG biometrics for each dataset; Section 3 describes the proposed techniques and system architecture; results and conclusions will be discussed in Sections 4 and 5, respectively.



2 Related Work

Recently, research in ECG biometry has been using DNN due to its popularity. Most architectures comprise CNN modules that both process, by automatically learning features, and classify. The inputs to these networks may vary, some examples are of *Labati et al. (2018)* [16], which uses the raw signal in a fiducial system, while *Zhang et al. (2017)* [17] describes a non-fiducial system that feeds the network with temporal frequency spectrograms obtained by discrete wavelet transform.

For benchmarking purposes, this section will address related work with respect to reference databases: the Fantasia and MIT-BIH datasets from Physionet, which are extensively used for biometrics research, and the CYBHi database, an off-person ECG database. This last database has been considered by *Merone et al. (2017)*[12] as one of the two best databases in terms of acquisition protocol and hardware for biometric studies, in comparison with fifteen databases.

2.0.1 Fantasia

Fantasia is a database from PhysioNet which was created under the auspices of the National Center for Research Resources of the National Institutes of Health [18]. This dataset contains 20 subjects aged between 21 and 34 years old and 20 aged between 68 and 81, during 120 min of continuous supine resting ECG recording while watching the Disney movie Fantasia [13].

This database was used by *Tantawi et al. (2013)*[19], which created a model that uses a reduced version of a set of fiducial features by applying Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), information-gain ratio and rough sets (RS) achieving an accuracy of $90 \pm 8\%$ in PCA, False Rejection Rate (FRR) 5% for LDA and False Negative Rate (FNR) of 4% for RS [19]. The same team published a method that extracts and decomposes R-R intervals using Discrete Wavelet Transform (DWT) and obtained an accuracy of 95.89%, FRR 0% and FNR 5% [20]. The work of *Gargulio et al. (2015)*[9] analyses the influence of the HRV in the QT intervals and suggests their correction to improve performance, leading to identification rates with MultiLayer Perceptron (MLP) and Support Vector Machine (SVM) between 97% and 99% [9]. The study described above by *Zhang et al. (2017)*[17] has archived an accuracy of 97.2% for the Fantasia dataset.

2.0.2 MIT-BIH

The MIT-BIH Database has been available since 1999 in PhysioNet. The basic subset of this database, MIT-BIH Arrhythmia, contains ECG records from 47 subjects with 360Hz of sample frequency and 11-bit resolution, from Boston's Beth Israel Hospital. More subjects, with no significant arrhythmic episodes, were added to this dataset, 18 subjects in MIT-BIH Normal Sinus and 7 individuals in MIT-BIH Long-Term [14].

This database has been used not only for validating arrhythmia detectors and other cardiac dynamics research, but also in the biometrics field. The MIT-BIH Normal Sinus Rhythm dataset was used by *Shen et al. (2011)* [21], *Rahbi (2013)* [22] and *Sidek et al. (2014)* [23] to validate their models. *Shen et al. (2011)* [21] proposes a Piecewise Linear Representation for feature extraction and classification using the similarity measures from the Dynamic Time Warping (DTW) method. The results for identification for was 100% and the minimum of the half total error (HTER) of 0.2% [24].

The accuracy in *Rahbi (2013)* [22] reached 99.07%, using a combination of ten features from heartbeat and sixty from coefficients of the Hermite Polynomials Expansion as input to a Hidden Markov Model (HMM) [22].

Several classification methods applied to characteristics of the QRS complex were compared by *Sidek et al. (2014)* [23] achieving the accuracy of 98.3% for the Bayes Network, 99.07% for Naive Bayes, 99.07% for Multilayer Perceptron and 99.07% for k-Nearest Neighbor from the same database [23].

A semi-fiducial approach was suggested by *Ye & Coimbra (2012)* [24] in which a Wavelet Transform was applied together with a Independent Component Analysis (ICA) analysis to detect each heartbeat and insert this information together with RR intervals and fed to a SVM classifier. For the selected 23 records with normal sinus rhythm from the Arrhythmia dataset, they reached 86.4% accuracy for subject

evaluation, even though the focus was to classify heartbeat classes. *Zhang (2017)* got accuracy of 90.3% for the 47 individuals of MIT-BIH [17].

2.0.3 CYBHi

The *Check Your Biosignals Here initiative* (CYBHi) dataset is a system that acquired ECG without the use of skin electrodes. In the work of *Silva et al. (2013)* [15] states that it was "devised a data acquisition framework and experimental setup, for large scale data collection from a large group of subjects through an easily repeatable and efficient procedure". Both sensors had hand-shaped support synchronized via syncPLUX synchronization kit, in which the electrodes were placed on the fingers while the subjects were seated for 2 min in a resting position. The used data consists of long-term sessions, where two acquisitions were made with 3 months difference, including 63 participants, 14 males, and 49 females (18–24 years old) [15]. For the purpose of this paper, the long-term sessions from the first acquisition will be named M1 while the second acquisition will be named M2.

The outcome for authentication of this study was an Equal Error rate (EER) of 9.1% using a SVM classifier against the correlation between ECG templates [15]. In another study that used this database, *Lourenço et al. (2015)* [5], extracted the features from the ECG morphology and used the segmentation of the RR signals using a Nearest Neighbor Algorithm (NNA) and obtained with the Euclidean distance classifier an identification rate of 95.2% for the first test and 90.2% for the second test. As for *Luz et al. (2018)* [25], described before, archived EER values of: M1 vs M1 of 1.33%; M1 vs M2 12.78%; M2 vs M1 13.93%.

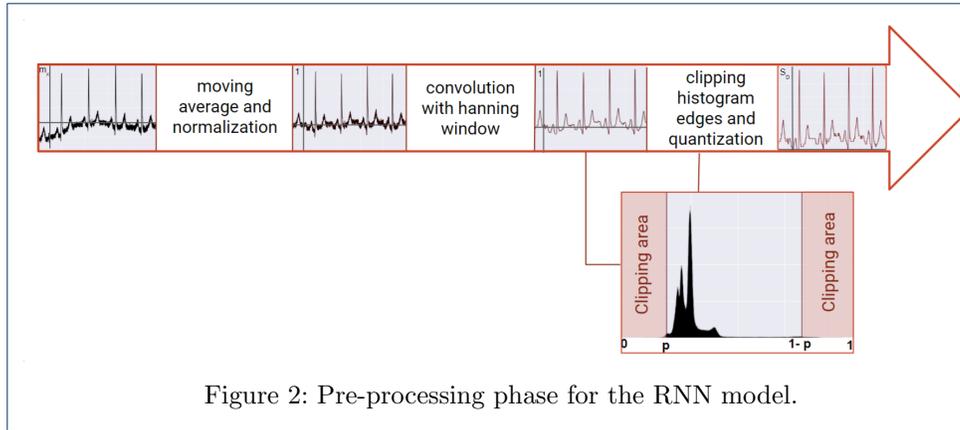
3 Methods

As stated before two architectures were compared and explored using the same classification method, the . The setup comprises the TCNN network and a RNN network, both are trained to create an internal representation of the waves. The difference between both architectures is while the TCNN scores are given by the output given by the last layer, the RNN approach the score is given by the prediction error.

In order to train and test models, heavily corrupted windows are removed using two different methods that assume that most of the signal is clean: (1) The first extracts the windows that are between intervals around a percentage of the median of the standard deviation value and the median of the mean value of each window, depending on the database noise factor; (2) The second uses the algorithm developed by *Rodrigues et al. (2017)* [7] relying on the standard deviation feature. This method was used for the CYBHi database where the first method was inefficient due to the low signal-to-noise ratio.

3.1 Recurrent Neural Network Approach

In the RNN approach, data is firstly pre-processed (Fig. 2), and later transformed by an embedded matrix - E , three sequential GRUs - G , a dense layer with linear activation and a softmax node [11].



3.2 Pre-processing

The pre-processing step starts with removing the moving average and normalized by a moving absolute maximum window and subtracted by a moving average window, followed by a convolution with a *Hanning* window. The benefit of this approach is to mitigate the unwanted frequencies from the signal acquisition. Due to the nature of the GRU networks, a quantization of the signal must be performed, which consists of reducing the continuous set of the amplitude values into a limited set of integer values with the dimension of S_D . But in order to reduce the influence of artifacts, and extract only the significant information, the signal amplitude was clipped, according to a value of confidence. To achieve this, the edges of the amplitude histogram outside a parameterized value of confidence were removed. The value of this parameter was typically 0.5% but could change depending on the corruption of the signal. Finally, the resulting signal (x) is transformed using the following equation:

$$x_n = \text{round} \left(\left[\frac{s_n - \min(s)}{\max(s - \min(s))} \right] \cdot [S_D - 1] \right) \quad (1)$$

where x_n is the n -th sample of the input vector. The variable $k \in \{0, 1, \dots, S_D - 1\}$ represents the number of possible integer values x_n may take. The higher the value of the signal dimension (S_D), the higher the detail is retained from the signal. The last step is the segmentation of the signal into sliding windows of size W with overlap, depending on the signal length and TWs.

3.3 Architecture

The architecture depicted in Fig. 1a, comprises one embedding matrix, commonly used for text processing. It functions as a translation mechanism between the input and the three GRU layers (G). The input is an integer value corresponding to the x_n -th column of the matrix E . It starts as a matrix with random values but optimizes as training progresses, adjusting itself to the input along with the other parameters.

The GRU is a type of RNN, characterized by learning the sequential patterns where the current state depends on the previous ones while learning short-term and long-term dependencies by memory management [26, 27].

In our previous work [11], this architecture could synthesize biosignals and prove by the fact that the prediction error of a model is lower when the input signal is the one that trained the model in relation to the signals with other origins. In the case of biometry, this notion is expanded to the individual source of the signal, where this network may be able to learn the individual intricacies of each ECG. Following this reasoning, the produced dissimilarity score ($S(p, i, w)$) is given by the following equation:

$$S(p, i, w) = \frac{e_p}{\max(e(:, i, w))} \quad (2)$$

, where e_p is the prediction error made to the predicted individual, i.e. predictor - p - and w is the time-window of the signal that belongs to the individual i . The expression $\max(e(:, i, w))$ represents the maximum for the same window w for all predictors.

3.4 Temporal Convolutional Neural Network

The second proposed method is a two-stream TCNN [28], which uses 1-dimensional convolutional layers with dilated convolutions to learn temporal patterns. It combines predictions from two different inputs: ECG non-fiducial window segments and a sub-segment of the same window containing only one full cycle centered in R peak of the QRS complex.

This architecture is depicted in Figure 1.3 (b), where each individual network is composed of several convolutional layers with 24 kernels of size 4, followed by batch normalization and a ReLU activation function. At the end of each network, two fully connected layers, the first with 256 units while the second with the number of individuals are applied to the end of each network producing a logit vector for each network. The scale differences between the input used by each network and the progressive dimensionality reduction of the TCNN generate variability in the number of layers. For the Fantasia and MIT-BIH databases, the number of convolutional layers is 6 non-fiducial network and 4 for the individual cycle network. Due to the increase of sampling frequency of the CYBHi database, the number of layers is adjusted to 8 and 6, respectively.

These networks are trained independently using cross-entropy loss function and Adam optimizer. After trained, the logit vectors from both networks are fused with the sum rule. The score, in this case, is given by subtracting one to the normalized output vector (o_n), given by:

$$S(p, i, w) = 1 - \frac{o_p}{\max(o(:, i, w))} \quad (3)$$

3.5 Relative Score Threshold Classification

The RSTC is a simple method that classifies choosing the lowest normalized similarity score ($\bar{S}(p, i, b)$) for each batch of windows (b), consequently, S is a 3 dimension tensor with dimensions ($M \times I \times B$), where B is the size of the batch. This measurement is obtained by normalizing for each predictor the minimum of a set of windows. Therefore the minimum is the result of the following function:

$$S_{p,i,b} = \min \left[\left\{ S_{p,i,B \cdot b}, S_{p,i,B \cdot b + 1}, \dots, S_{p,i,B \cdot b + B} \right\} \right] \quad (4)$$

and the normalization for each b in relation to the p is made with the following rule:

$$\bar{S}_{p,i,b} = \frac{S_{p,i,b} - \min(S_{:,i,b})}{\max[S_{:,i,b} - \min(S_{:,i,b})]} \quad (5)$$

, where minimum and maximum values are calculated in respect to all predictors for each batch of windows. Considering that $\bar{S}(p, i, b) \in [0, 1]$ the value 0 encodes to the most probable class and when closer to 1, the lower probability. The class of the individual is given by $C_{i,b}^I$:

$$C_{i,b}^I = \arg \min_p \left[\bar{S}_{:,i,b} \right] \quad (6)$$

3.6 Evaluation

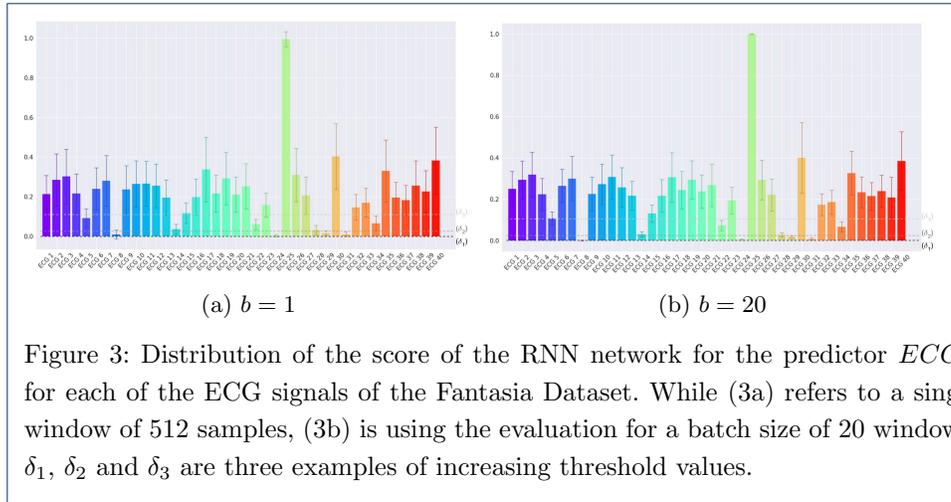
The identification evaluation made by the multimodal classifier (C^I) is given by the accuracy, specificity, and sensibility. The authentication evaluation is made by the EER that is obtained when FNR = False Positive Rate (FPR) for each binary classifier (C^A). The authentication mode is evaluated with this measurement, as these systems prioritize the minimization of the possibility of imposters that can access the system and the number of the rightful individuals that cannot log in. Each binary classifier is made by comparing the score for all predictors p for each individual i for all windows w with a changing threshold (φ). This system is fine-tuned for each individual and the output gives if that specific b of windows belongs to claimed individual:

$$C_{p,i,b}^A = \bar{S}(p, i, b) > \delta, \quad C^A \in 0, 1 \text{ and } \bar{S} \in [0, 1] \quad (7)$$

The Receiver Operating Characteristic (ROC) can be calculated with an array of those thresholds and the EER. Since each individual has its own fine-tuned ROC, the mean and standard deviation must be calculated and used as validation. An example of how these thresholds are obtained is depicted in Fig. 3

4 Results

This section will present the results for each database, ordered in increasing average intra-variability across subjects. Apart from the validation for identification and authentication modalities, the Fantasia dataset will be the benchmark to test the used methods and robustness of the proposed algorithms.



4.0.1 Fantasia

The first hour of the Fantasia dataset was separated after 20 minutes for cross-validation. The first part was used for testing and all data was segmented in windows of 512 samples (approximately 2s) with an overlap of 67%. The selected windows were on average 1464 windows of 3787 per subject. The rejection algorithm had a mean and standard deviation tolerance of 0.2 and 0.05, respectively.

The score distribution per subject when the test set is fed into the RNN model trained with ECG_8 is depicted in Fig. 3. The score in the right is given for all batches of windows for $B = 1$ and the left for $B = 20$.

When analyzing this example, it can be observed that with the increase in the number of windows included in each batch, both mean and variability are reduced. The three thresholds presented in Fig. 3 will have a different binary classification for the predicted ECG_8 when submitted to different values of B . In the case of (a), δ_1 threshold will give a positive value for most of the ECG_{23} , which is false, but both δ_2 and δ_3 will include the ECG_8 batches as well, but also the number of imposters increase with the threshold value.

When the B increases to 20, the variance and mean value of ECG_8 will decrease significantly and δ_1 will now classify correctly most, or even all, the ECG_8 batches, while reducing the number of ECG_{23} classified wrongly, ensuring lower values for the FNR and FPR. Both δ_2 and δ_3 produce more imposters, while not increasing the correct classification. The ROC curve is generated for each predictor and for each batch size, producing different values for EER that can be interpreted for authentication systems.

The identification rate, i.e. accuracy for the identification modality, for both algorithms are depicted in Figs. 4a. Close inspection reveals that both algorithms increase accuracy with time per batch, but the TCNN algorithm starts with higher accuracy, being outperformed by RNN after approximately 1 minute of signal per batch. While the RNN algorithm increases until reaching 100 % at approximately 112 s, the TCNN algorithm reaches 99.1% after approximately 90 s, but its curve displays a stabler behavior. The best results are presented in the confusion matrices of Fig. ?? for RNN and Fig. ?? TCNN approaches.

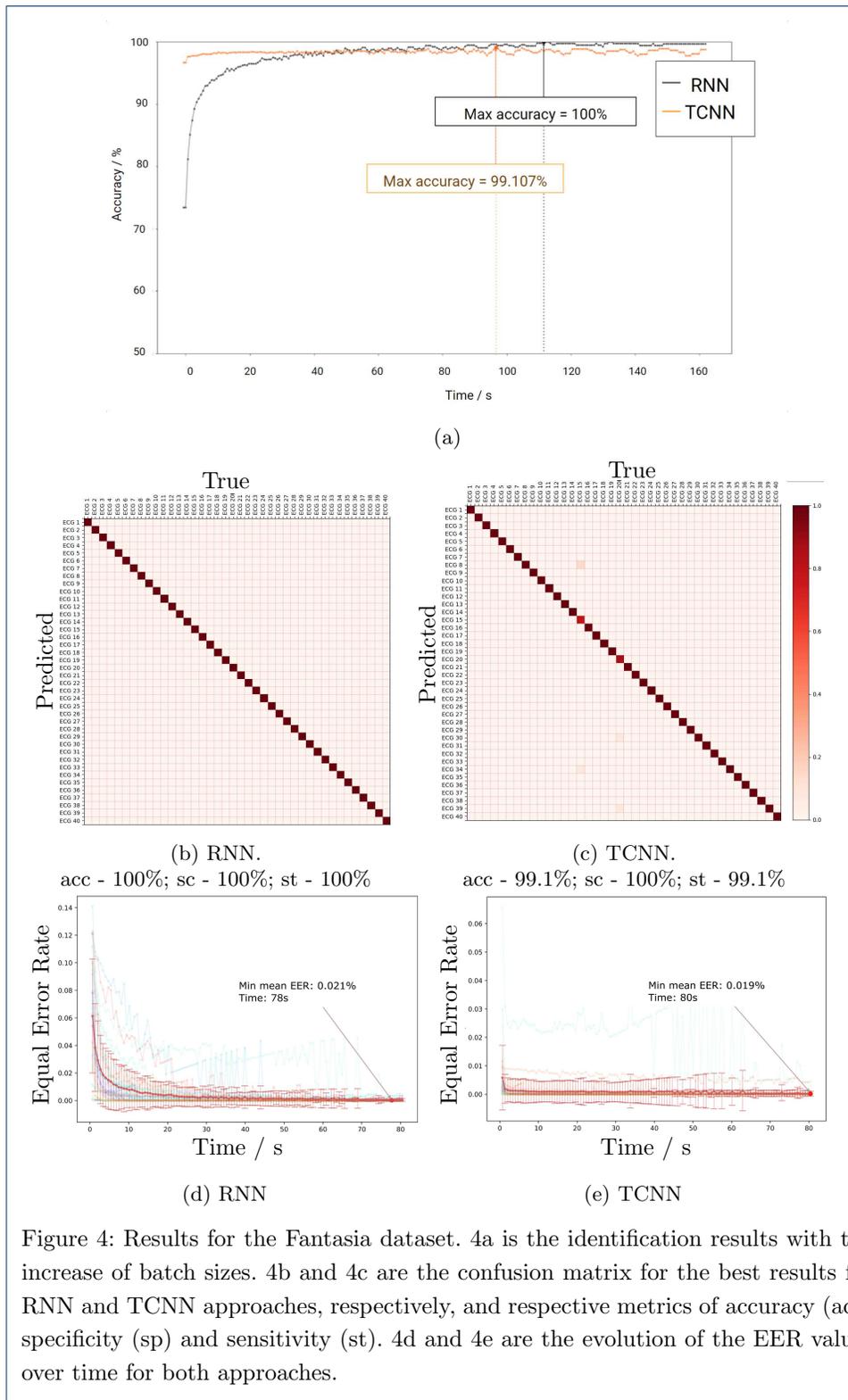


Figure 4: Results for the Fantasia dataset. 4a is the identification results with the increase of batch sizes. 4b and 4c are the confusion matrix for the best results for RNN and TCNN approaches, respectively, and respective metrics of accuracy (acc) specificity (sp) and sensitivity (st). 4d and 4e are the evolution of the EER values over time for both approaches.

The results for the authentication mode is depicted in Figs. 4d and ??, which shows the evolution of the EER per time in each batch. Both RNN and TCNN achieve values very close to 0% at 80 s, but the TCNN reaches those values with a lower standard deviation faster than the RNN counterpart.

4.0.2 MIT-BIH

All the MIT-BIH signals were resampled to a frequency of 250 Hz to ensure that the same sampling frequency is maintained in all datasets for a rigorous comparison. The rejection rate was 42% with the parameters for tolerance for average and standard deviation of 0.9 and 0.5, respectively.

The identification test performed in the MIT-BIH database, observed in Fig. 5a, shows that the accuracy curves have similar behavior as the ones in the Fantasia dataset. For the sake of comparison with the bibliography, an analysis of the Normal Sinus Rhythm dataset was also made in detail (Fig. 5a). Even though the RNN approach does not go above 80% for all the MIT-BIH, the confusion matrix (Fig. 5b) shows the expected diagonal. Unfortunately, some models generalize the ECG signals confusing the right class, probably due to the loss of information during resampling, due to the sensibility to noise or to due to the symptomatic behavior of the arrhythmia events. The last option is supported by the fact that the Normal Sinus Rhythm subset reached 100% accuracy.

As for the TCNN even though the 96.4% is reached after more than 2 min for the full MIT-BIH database, one can observe that values close to 96% are reached after 10 s, reaffirming its robustness to noise and stability. The confusion matrix in Fig. ?? shows that only two predictors display a faulty behavior, probably due to the presence of arrhythmia events. When comparing the individuals from the Normal Sinus Rhythm the accuracy reached 100% after approximately 10 s (Fig. 5a).

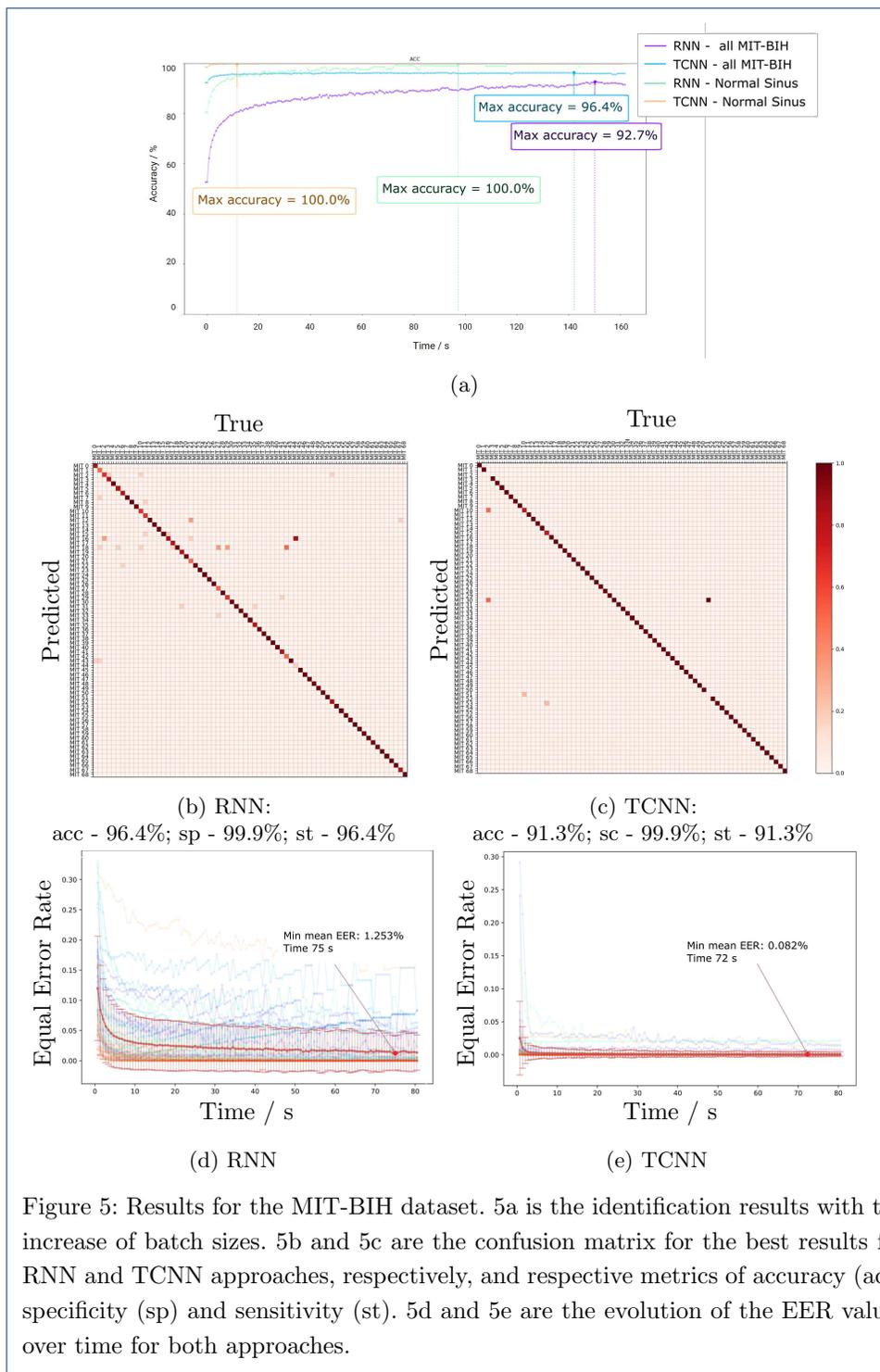
The authentication evaluation of all the MIT-BIH individuals is displayed in Figs. 5d and ?. These figures show that the TCNN approach reaches EER values close to 0% with 72 s in each batch and 1.7% for RNN. Even though the best result for TCNN is after 1 min, the values are quite low from the start. For the Normal Sinus Rhythm database (Fig. 5a) the TCNN reached 0% of EER, while the RNN reached the value of 0.6%.

4.0.3 CYBHi

As mentioned before the CYBHi database comprises two moments with a time distance of two months. Therefore, both algorithms are evaluated with all the combinations of both moments (M1 and M2) in terms of cross-validation, simulating the enrolment and verification environments of a biometric system. This means that each model was trained using the first element and tested with the second of the following sets: "M1 vs M1", "M2 vs M2", "M1 vs M2" and "M2 vs M1". When the training and testing moment was the same, the data was separated by 50% and the segmentation was made with an 89% of overlap to increase the number of available time-windows.

The clustering method rejected $8.93 \pm 22.2\%$ of the total signal length on the first session and $9.97 \pm 24.2\%$ on the second.

In Fig. 6a shows that the decrease of accuracy reflects the signal corruption that this dataset faces. Since this dataset displays a high variability within subjects due



to the low Signal-to-Noise Ratio (SNR), the robustness of the method is imperative for good results. The TCNN algorithm archives results above 90% for "M1 vs M1" and 100 % for "M2 vs M2", even if the amount of information was significantly lower than the previous datasets. The decrease of crossing both moments as significant as expected, revealing accuracies above 75 % for both "M1 vs M2" and "M2 vs M1". The low accuracy presented by the RNN approach reflects the sensitivity to noise.

Fig. 6b shows that the EER archive 4.3% and 6.1% for "M1 vs M1" and "M2 vs M2", respectively, using the RNN models, while TCNN displays results close to 0%, probability due the temporal proximity of the training and testing sets. As for the evaluation across different moments, the EER for the RNN approach is 18.3% and 18.8% for "M1 vs M2" and "M2 vs M1", respectively and the 2.2% and 4.1% for the TCNN in the same order. These results, even though they are based in a higher quantity of information, they outperform current state-of-the-art for off-person measurements.

4.1 Comparative Study

Table 1 summarizes the results per database for the consulted bibliography. This table shows that the TCNN approach outperforms most of the other studies in both identification and authentication paradigms. The robustness of this architecture is mainly due to the combination of bringing together both fiducial and non-fiducial machine-learned characteristics of the signal through the fusion layer.

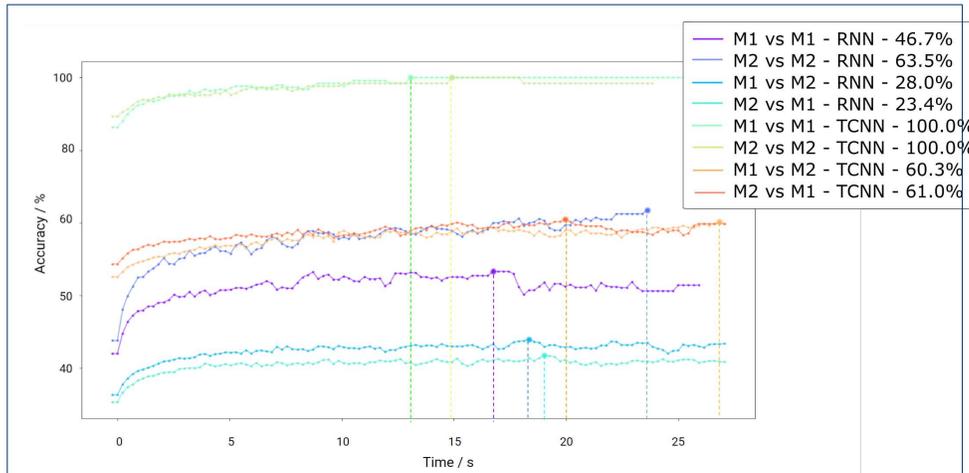
Unfortunately, the computational time increases significantly with the number of samples per window and the number of individuals and with it some limitations on providing a biometric system in a real scenario:

- when a new person is added to the model, all the training must be repeated;
- this approach is viable for a suitable number of individuals, but to use in a scale of millions, the training and testing process will be costly or even impossible with state of the art computational power and memory
- Another disadvantage is the need to extract the peak information, which sometimes is not feasible, especially when the slow waves (such as T or U) reach higher values than the R wave.

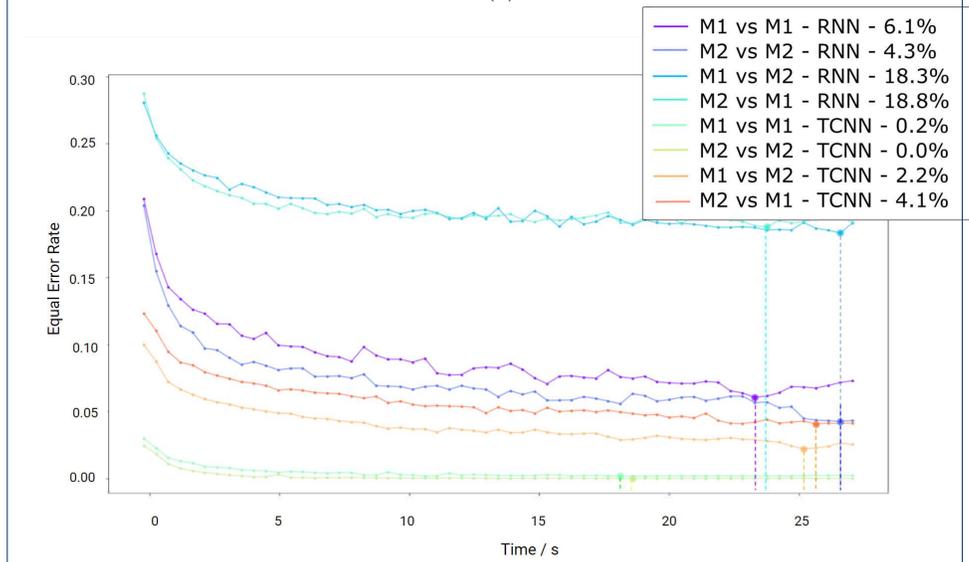
Even though the RNN achieves state of the art results for the Fantasia and Rhythm of the MIT-BIH, the sensibility to changes in the signal is higher than the aforementioned architecture. The most prominent possibilities for this sensibility is:

- The noise may have a high impact in the quantization process, especially due to random artifacts that shift the maximum and minimum values of the signal;
- The loss of information due to the preprocessing stage caused by the decimate method, which could be important in establishing the difference between individuals.
- The need for higher quantities of data for training, which could contain more angles of the ECG mechanism.

Another limitation for this system is the required computational power, due to its recursive nature, it is impossible to parallelize between sequential samples, leading to the training of 512 windows with 512 samples an average of 35.5 hours with an Nvidia GTX 1080 Ti. In terms of testing, the score calculation for each window is just on the scale of milliseconds.



(a)



(b) EER values through time.

Figure 6: Results for the CYBHi dataset for each approach. M1 represents the first moment and M2 represents the second moment.

Table 1: Fantasia Dataset

| Study | Database | Acc | EER |
|-------|-------------------|--------------|--------------|
| [19] | Fantasia | 90.0% | — |
| [20] | | 95.9% | — |
| [9] | | 99.0% | — |
| RNN | | 100% | 0.02% |
| TCNN | | 99.1% | 0.02% |
| [29] | MIT-BIH Sinus | 98.1% | — |
| [30] | | 99.6% | — |
| [29] | | 94.5% | — |
| [22] | | 99.0% | — |
| [23] | | 99.1% | — |
| RNN | | 100% | 0.6% |
| TCNN | 100% | 0.0% | |
| RNN | MIT-BIH | 79.8% | 2.4% |
| TCNN | Arr, Sin, Long | 96.3% | 0.1% |
| [15] | CYBHi M1 vs M1 | 94.4% | — |
| [5] | | 95.2% | — |
| [25] | | — | 1.3% |
| RNN | | 46.7% | 6.1% |
| TCNN | | 100% | 0.2% |
| [5] | CYBHi M1 vs M2 | 90.2% | — |
| [25] | | — | 12.8% |
| RNN | | 28.0% | 18.3% |
| TCNN | | 60.3% | 2.2% |
| [25] | CYBHi M2 vs M1 | — | 14.0% |
| RNN | | 23.5% | 18.8% |
| TCNN | | 61.0% | 4.1% |
| RNN | CYBHi | 63.5% | 4.3% |
| TCNN | M2 vs M2 | 100% | 0.0% |

The main advantages of the RNN architecture are:

- It is unnecessary to train all the models again when adding a new person to the biometric system as the individual traits of one individual would be preserved in a single model that is independent of the others;
- The non-fiducial approach removes the need to extract the R peak from the signal;
- This approach achieves excellent results on signals with good SNR, therefore this system is viable for good quality acquisitions.
- The possibility of generating the signal using the same method as in *Belo et al. (2017)* [11] to observe the differences with the individual, in order to track models that did not learn correctly.

5 Conclusions

This study has also shown the usefulness of DNN in enhancing current biometric systems. These technologies allied with the massive collection and storage of biological data will provide powerful and robust systems, increasing the level of security, reducing the concerns of counterfeiting actions.

Even though the RNN and TCNN systems provided good results, their limitations could be surpassed if both architectures were combined. Other state-of-the-art downsampling methods could also be applied in order to mitigate the loss of information while increasing the training speed. The noise contamination is also a major concern, by employing other filtering methods the system could have a better grasp of the intricate characteristics, only present in a clean signal.

The next steps to implement in a real-life scenario would be the introduction of transfer learning in the training process, the acquisition of more time windows per

person, the introduction of more individuals, and the increase of the number of moments through time. The study of the impact of the noise and the scalability of these systems would give an added value.

One factor that could give a higher generalization and detection capabilities would be the inclusion of different electrode configurations for the clinical-grade ECG so that the DNN model learns the different angles of the signal gaining robustness over the different placements. The analysis of different emotional and mental states could also be very important, as the ECG signal can change under different conditions.

Competing interests

The authors declare that they have no competing interests.

Author's contributions

David Belo and Hugo Gamboa conceived the presented idea. David Belo and Nuno Bento developed the theory and performed the computations. Hugo Silva and Ana Fred verified the analytical methods. Hugo Gamboa supervised the findings of this work. All authors discussed the results and contributed to the final manuscript.

Author details

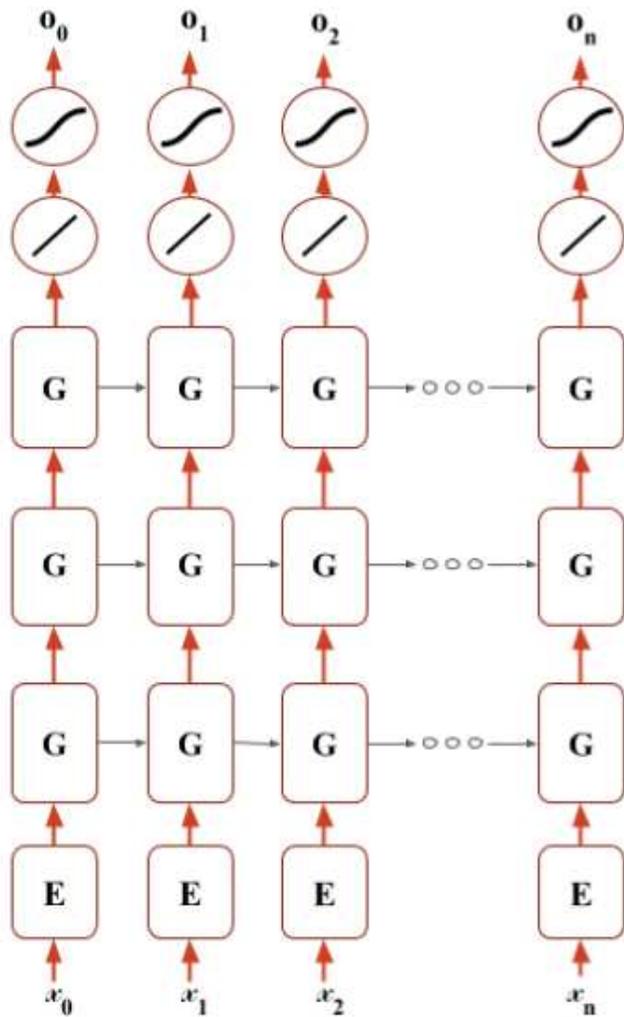
¹LIBPhys - Nova University of Lisbon, Departamento de Física - Faculdade de Ciências e Tecnologia - Universidade Nova de Lisboa, Caparica, Portugal. ²Instituto de Telecomunicações - Instituto Superior Técnico (IST) - Technical University of Lisbon, Av. Rovisco Pais, 1, Lisbon, Portugal. ³Instituto de Telecomunicações (IT) - Instituto Superior Técnico (IST) - Technical University of Lisbon, Av. Rovisco Pais, 1, Lisbon, Portugal.

References

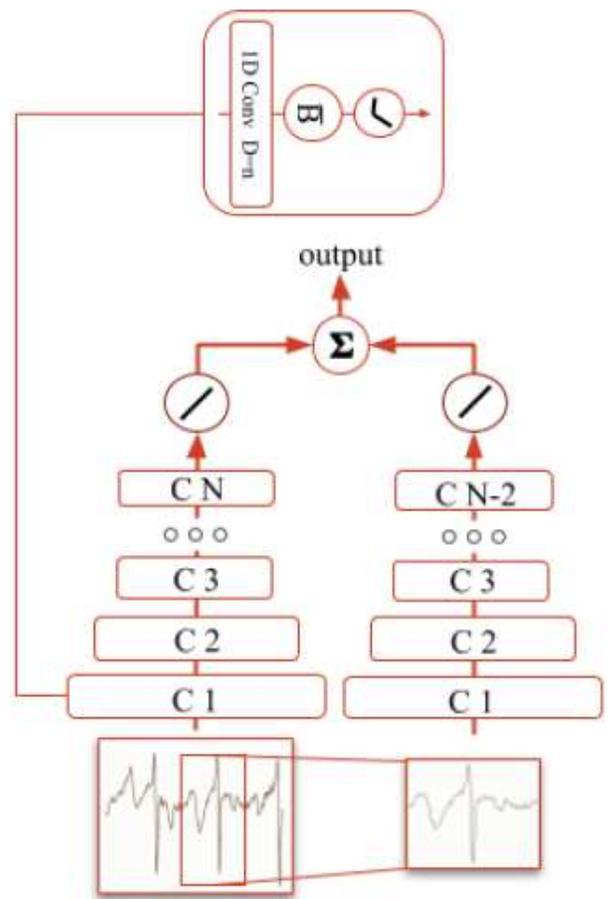
- Unar, J.A., Seng, W.C., Abbasi, A.: A review of biometric technology along with trends and prospects. *Pattern Recognition* **47**(8), 2673–2688 (2014). doi:10.1016/j.patcog.2014.01.016
- MarketsandMarkets: Report: Biometric system market by authentication type, component, function, application, and region - global forecast to 2022. Technical Report 4397168, MarketsandMarkets institution, Hadapsar, Pune-411013, India (November 2016). <http://www.marketsandmarkets.com/Market-Reports/next-generation-biometric-technologies-market-697.html>
- Caldwell, T.: Market report: border biometrics. *Biometric Technology Today* **2015**(5), 5–11 (2015)
- Nait-Ali, A.: Hidden biometrics: Towards using biosignals and biomedical images for security applications. 7th International Workshop on Systems, Signal Processing and their Applications, WoSSPA 2011, 352–356 (2011). doi:10.1109/WOSSPA.2011.5931509
- Lourenço, A., Silva, H., Fred, A.: ECG-based biometrics: A real time classification approach. *IEEE International Workshop on Machine Learning for Signal Processing, MLSP*, 1–6 (2012). doi:10.1109/MLSP.2012.6349735
- Nait-Ali, A.: Hidden biometrics: Towards using biosignals and biomedical images for security applications. In: *Proc. IEEE In'l Workshop on Systems, Signal Processing and Their Applications (WOSSPA)*, pp. 352–356 (2011). IEEE
- Rodrigues, J., Belo, D., Gamboa, H.: Noise detection on ECG based on agglomerative clustering of morphological features. *Computers in biology and medicine* **87**, 322–334 (2017)
- Wübbeler, G., Stavridis, M., Kreisler, D., Boussejot, R.D., Elster, C.: Verification of humans using the electrocardiogram. *Pattern Recognition Letters* **28**(10), 1172–1175 (2007). doi:10.1016/j.patrec.2007.01.014
- Gargiulo, F., Fratini, A., Sansone, M., Sansone, C.: Subject identification via ECG fiducial-based systems: Influence of the type of QT interval correction. *Computer Methods and Programs in Biomedicine* **121**(3), 127–136 (2015). doi:10.1016/j.cmpb.2015.05.012
- Da Silva, H.P., Fred, A., Lourenço, A., Jain, A.K.: Finger eeg signal for user authentication: Usability and performance. In: *2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pp. 1–8 (2013). IEEE
- Belo, D., Rodrigues, J., Vaz, J.R., Pezarat-Correia, P., Gamboa, H.: Biosignals learning and synthesis using deep neural networks. *BioMedical Engineering OnLine* **16**(1), 115 (2017). doi:10.1186/s12938-017-0405-0
- Merone, M., Soda, P., Sansone, M., Sansone, C.: ECG databases for biometric systems: A systematic review. *Expert Systems with Applications* **67**, 189–202 (2017). doi:10.1016/j.eswa.2016.09.030
- Iyengar, N., Peng, C., Morin, R., Goldberger, A.L., Lipsitz, L.A.: Age-related alterations in the fractal scaling of cardiac interbeat interval dynamics. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* **271**(4), 1078–1084 (1996)
- Moody, G.B., Mark, R.G.: The impact of the MIT-BIH arrhythmia database. *IEEE Engineering in Medicine and Biology Magazine* **20**(3), 45–50 (2001)
- da Silva, H.P., Lourenço, A., Fred, A., Raposo, N., Aires-de-Sousa, M.: Check Your Biosignals Here: A new dataset for off-the-person ECG biometrics. *Computer Methods and Programs in Biomedicine* **113**(2), 503–514 (2014). doi:10.1016/j.cmpb.2013.11.017
- Labati, R.D., Muñoz, E., Piuri, V., Sassi, R., Scotti, F.: Deep-ecg: Convolutional neural networks for eeg biometric recognition. *Pattern Recognition Letters* (2018). doi:10.1016/j.patrec.2018.03.028
- Zhang, Q., Zhou, D., Zeng, X.: Heartid: A multiresolution convolutional neural network for eeg-based biometric human identification in smart health applications. *IEEE Access* **5**, 11805–11816 (2017). doi:10.1109/ACCESS.2017.2707460. cited By 0

18. Goldberger, A.L., Amaral, L.A., Glass, L., Hausdorff, J.M., Ivanov, P.C., Mark, R.G., Mietus, J.E., Moody, G.B., Peng, C.-K., Stanley, H.E.: Physiobank, physiotoolkit, and physionet components of a new research resource for complex physiologic signals. *Circulation* **101**(23), 215–220 (2000)
19. Tantawi, M., Revett, K., Salem, A.-b.: ECG based Biometric Recognition using Wavelets and RBF Neural Network 1 Introduction 2 Related Work. 7th European Computing Conference, 100–104 (2013)
20. Tantawi, M.M., Revett, K., Salem, A., Tolba, M.F.: Fiducial feature reduction analysis for electrocardiogram (ECG) based biometric recognition. *Journal of Intelligent Information Systems* **40**(1), 17–39 (2013). doi:10.1007/s10844-012-0214-7
21. Shen, J., Bao, S.D., Yang, L.C., Li, Y.: The PLR-DTW method for ECG based biometric identification. Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, 5248–5251 (2011). doi:10.1109/IEMBS.2011.6091298
22. Rabhi, E., Lachiri, Z.: Biometric personal identification system using the ECG signal. In: Proceedings in Computing in Cardiology Conference (CinC), 2013, pp. 507–510 (2013). IEEE
23. Sidek, K.A., Mai, V., Khalil, I.: Data mining in mobile ECG based biometric identification. *Journal of Network and Computer Applications* **44**, 83–91 (2014). doi:10.1016/j.jnca.2014.04.008
24. Ye, C., Kumar, B.V., Coimbra, M.T.: Heartbeat classification using morphological and dynamic features of ECG signals. *IEEE Transactions on Biomedical Engineering* **59**(10), 2930–2941 (2012)
25. da Silva Luz, E.J., Moreira, G.J.P., Oliveira, L.S., Schwartz, W.R., Menotti, D.: Learning deep off-the-person heart biometrics representations. *IEEE Transactions on Information Forensics and Security* **13**(5), 1258–1270 (2018). doi:10.1109/TIFS.2017.2784362
26. Greff, K., Srivastava, R.K., Koutník, J., Steunebrink, B.R., Schmidhuber, J.: LSTM: A search space odyssey. *IEEE transactions on neural networks and learning systems* **28**(10), 2222–2232 (2017)
27. Wu, Z., King, S.: Investigating gated recurrent neural networks for speech synthesis. *CoRR* **abs/1601.02539** (2016)
28. Bai, S., Kolter, J.Z., Koltun, V.: An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271* (2018)
29. Wang, Y., Agrafioti, F., Hatzinakos, D., Plataniotis, K.N.: Analysis of Human Electrocardiogram for Biometric Recognition. *EURASIP Journal on Advances in Signal Processing* **2008**(1), 148658 (2008). doi:10.1155/2008/148658
30. Fatemian, S.Z., Hatzinakos, D.: A new ECG feature extractor for biometric recognition. DSP 2009: 16th International Conference on Digital Signal Processing, Proceedings, 1–6 (2009). doi:10.1109/ICDSP.2009.5201143. IEEE

Figures



(a) Recurrent Neural Network Approach



(b) Temporal CNN Approach

Figure 2

Proposed architectures. (1a) depicts the whereRNN approach each input (x_n) is fed into to a embedded matrix (E) where the result vector enters into a se- quence of three Gated Recurrent Units (GRU) (G) cells. The output of this sequence is transformed through a fully connected network and a softmax function resulting in o_n . (1b) depicts the Temporal Convolutional Neural Network (TCNN) approach, where a time window (TW) and the first complete cycle are input to two different sequences of CNN blocks (CN) each comprised by a one dimensional convolutional layer with N dilation factor, a batch normalization and a ReLU activation function. One of the networks has N number of blocks while the other N - 2, they both end in a fully connected network and then the last vector is summed giving the output for classification.

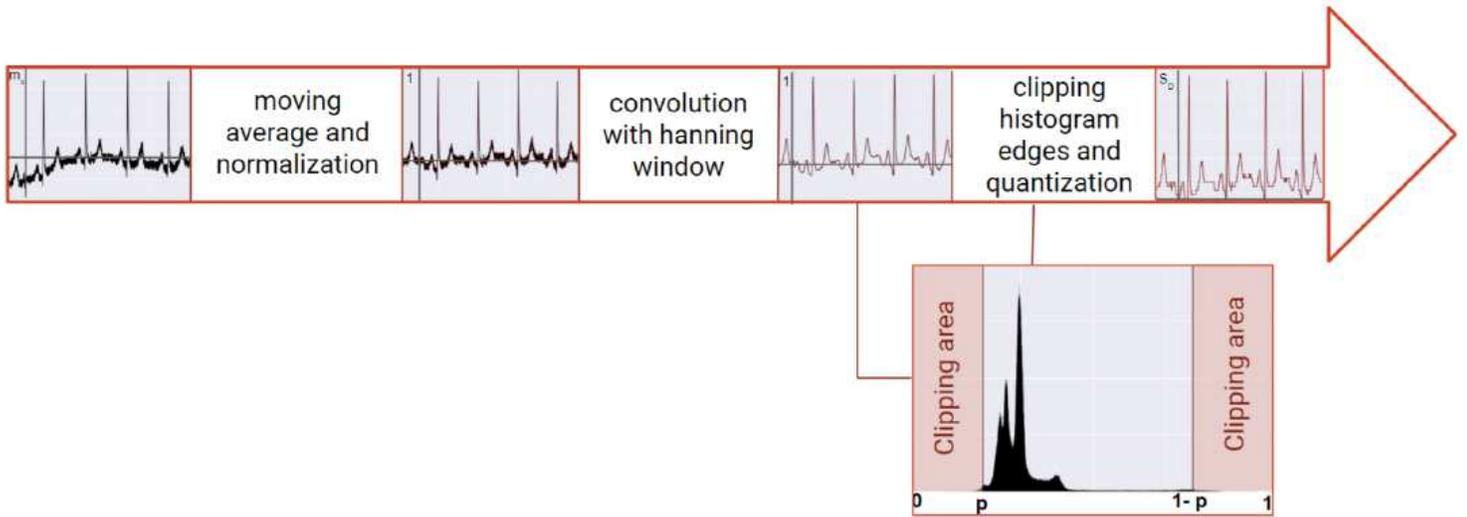


Figure 4

Pre-processing phase for the RNN model.

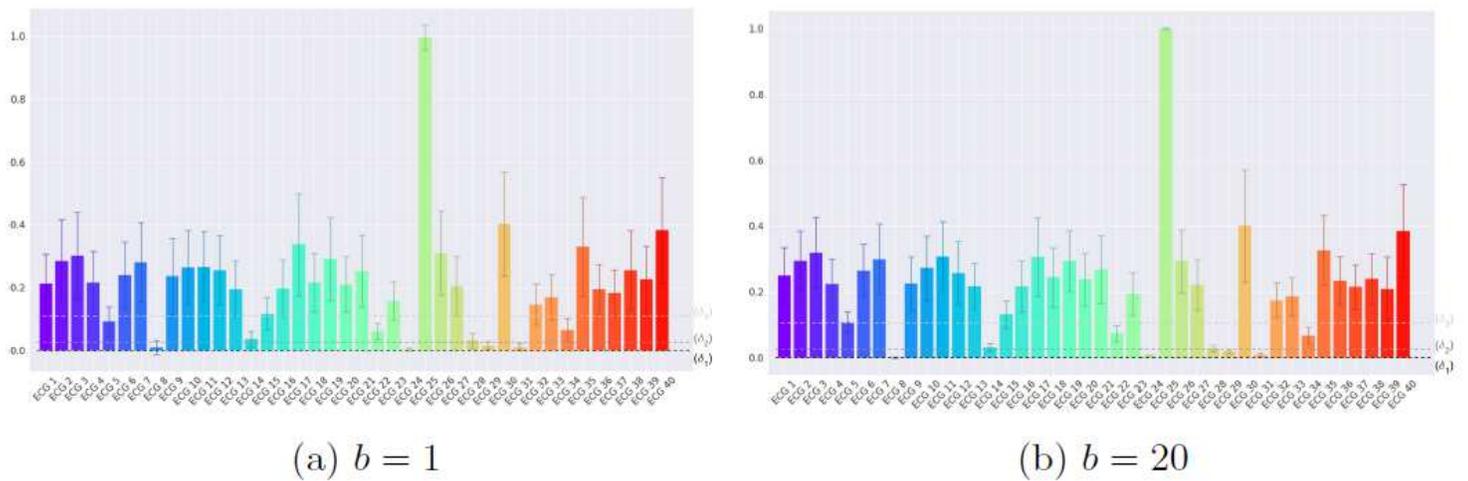
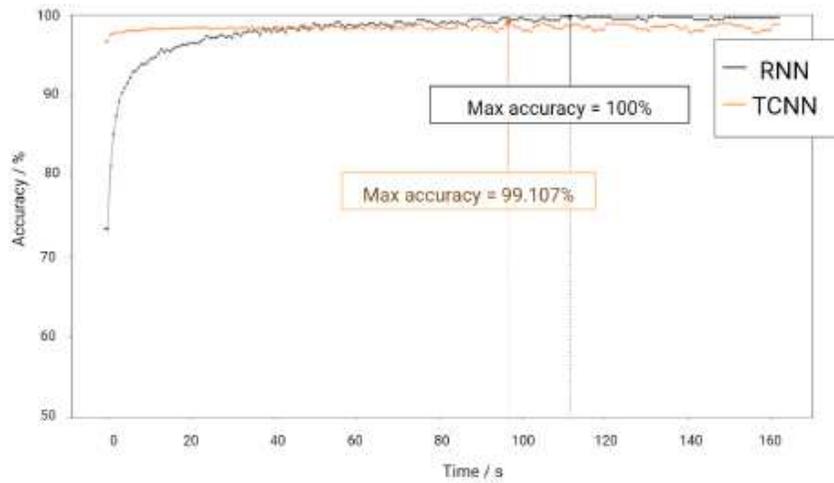
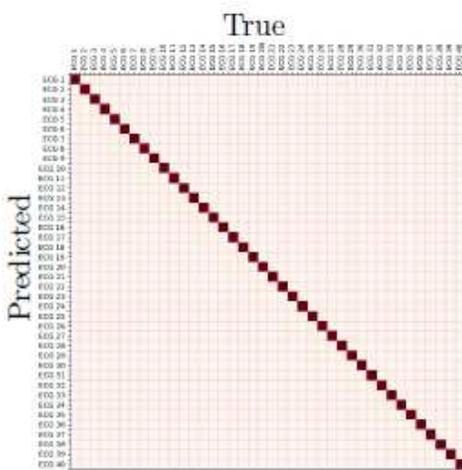


Figure 6

Distribution of the score of the RNN network for the predictor ECG 8 for each of the ECG signals of the Fantasia Dataset. While (3a) refers to a single window of 512 samples, (3b) is using the evaluation for a batch size of 20 windows. δ_1 , δ_2 and δ_3 are three examples of increasing threshold values.

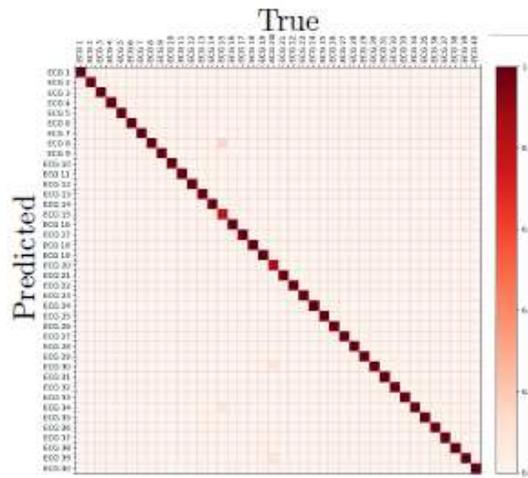


(a)



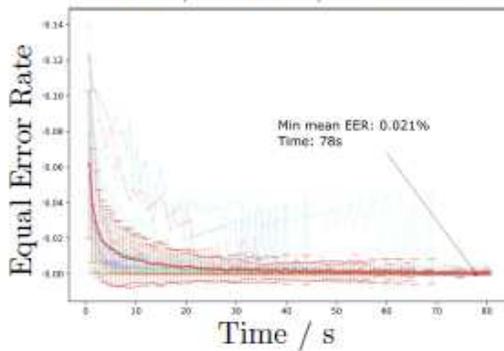
(b) RNN.

acc - 100%; sc - 100%; st - 100%

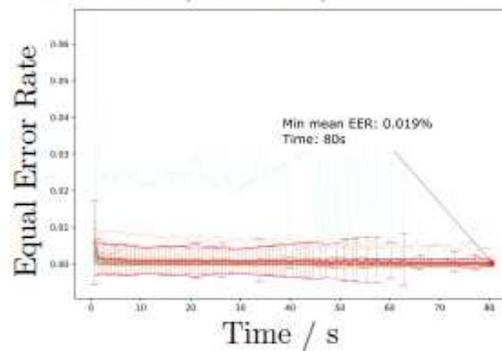


(c) TCNN.

acc - 99.1%; sc - 100%; st - 99.1%



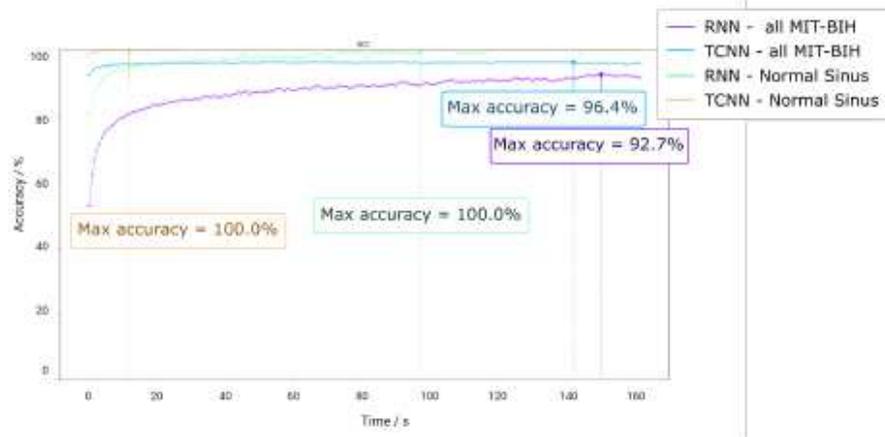
(d) RNN



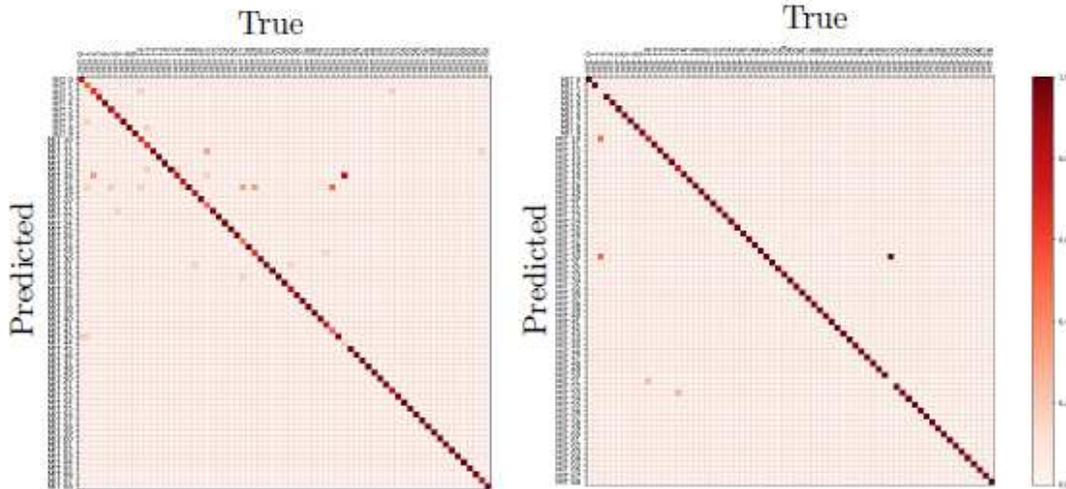
(e) TCNN

Figure 8

Results for the Fantasia dataset. 4a is the identification results with the increase of batch sizes. 4b and 4c are the confusion matrix for the best results for RNN and TCNN approaches, respectively, and respective metrics of accuracy (acc) specificity (sp) and sensitivity (st). 4d and 4e are the evolution of the EER values over time for both approaches.



(a)

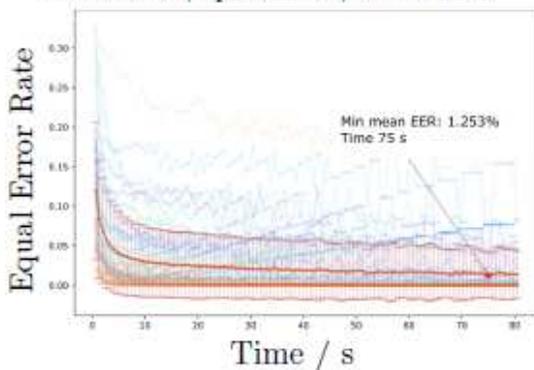


(b) RNN:

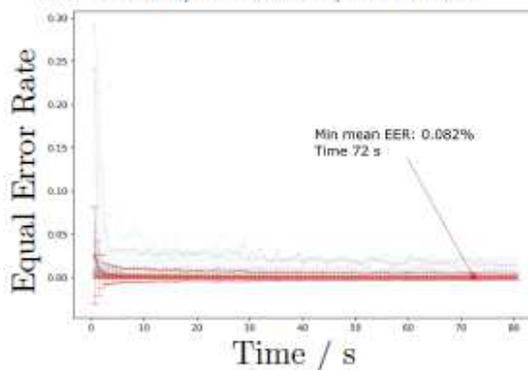
acc - 96.4%; sp - 99.9%; st - 96.4%

(c) TCNN:

acc - 91.3%; sc - 99.9%; st - 91.3%



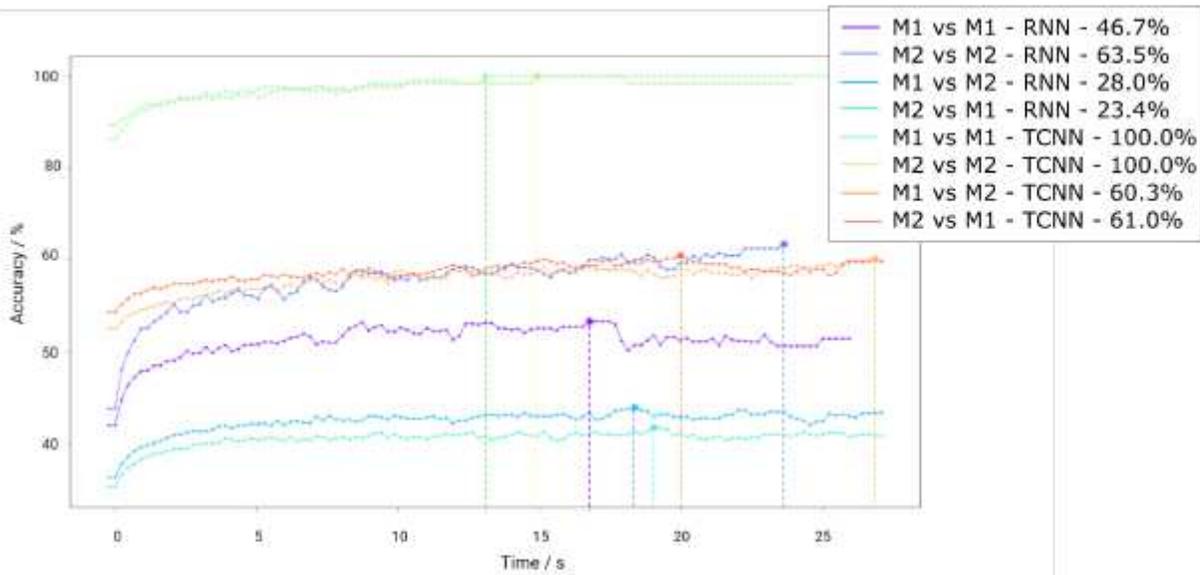
(d) RNN



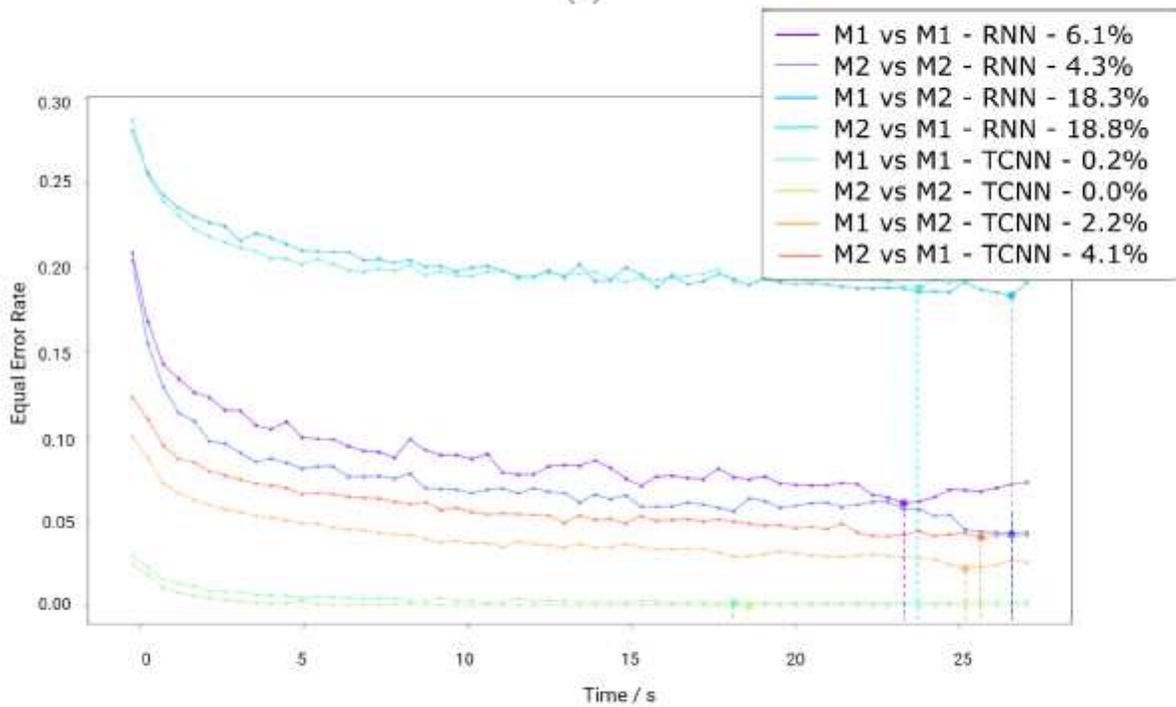
(e) TCNN

Figure 10

Results for the MIT-BIH dataset. 5a is the identification results with the increase of batch sizes. 5b and 5c are the confusion matrix for the best results for RNN and TCNN approaches, respectively, and respective metrics of accuracy (acc) specificity (sp) and sensitivity (st). 5d and 5e are the evolution of the EER values over time for both approaches.



(a)



(b) EER values through time.

Figure 12

Results for the CYBHi dataset for each approach. M1 represents the first moment and M2 represents the second moment.