

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/343139887>

ECG Biometrics Using Deep Learning and Relative Score Threshold Classification

Article in Sensors · July 2020

DOI: 10.3390/s20154078

CITATIONS
0

READS
59

5 authors, including:



Some of the authors of this publication are also working on these related projects:

-  PersonAAL (Personalized web application to improve quality of life and remote care for older adults) [View project](#)
-  Arthrokinemat [View project](#)

Article

ECG Biometrics Using Deep Learning and Relative Score Threshold Classification

David Belo ^{1,*},[†] , Nuno Bento ^{1,†}, Hugo Silva ^{2,‡} , Ana Fred ^{2,‡}  and Hugo Gamboa ^{1,†} 

¹ LIBPhys, Physics Department, Faculty of Sciences and Technology, Nova University of Lisbon, 2825-149 Caparica, Portugal; n.bento@campus.fct.unl.pt (N.B.); hgamboa@fct.unl.pt (H.G.)

² Instituto de Telecomunicacões, Instituto Superior Técnico (IST), Technical University of Lisbon, 1049-001 Lisboa, Portugal; hsilva@lx.it.pt (H.S.); afred@lx.it.pt (A.F.)

* Correspondence: dj.belo@fct.unl.pt

† Current address: Faculdade de Ciências e Tecnologia - Nova University of Lisbon, 2829-516 Caparica, Portugal.

‡ Current address: Av. Rovisco Pais, 1049-001 Lisbon, Portugal.

Received: 25 June 2020; Accepted: 20 July 2020; Published: 22 July 2020



Abstract: The field of biometrics is a pattern recognition problem, where the individual traits are coded, registered, and compared with other database records. Due to the difficulties in reproducing Electrocardiograms (ECG), their usage has been emerging in the biometric field for more secure applications. Inspired by the high performance shown by Deep Neural Networks (DNN) and to mitigate the intra-variability challenges displayed by the ECG of each individual, this work proposes two architectures to improve current results in both identification (finding the registered person from a sample) and authentication (prove that the person is whom it claims) processes: Temporal Convolutional Neural Network (TCNN) and Recurrent Neural Network (RNN). Each architecture produces a similarity score, based on the prediction error of the former and the logits given by the last, and fed to the same classifier, the Relative Score Threshold Classifier (RSTC). The robustness and applicability of these architectures were trained and tested on public databases used by literature in this context: Fantasia, MIT-BIH, and CYBHI databases. Results show that overall the TCNN outperforms the RNN achieving almost 100%, 96%, and 90% accuracy, respectively, for identification and 0.0%, 0.1%, and 2.2% equal error rate (EER) for authentication processes. When comparing to previous work, both architectures reached results beyond the state-of-the-art. Nevertheless, the improvement of these techniques, such as enriching training with extra varied data and transfer learning, may provide more robust systems with a reduced time required for validation.

Keywords: deep learning; biometrics; electrocardiogram; convolutional neural network; recurrent neural network; authentication; identification; artificial neural networks; biosignal; RLTC

1. Introduction

Some traditional methods of identification and authentication, such as key-possession or username-password, may prove ineffective against terrorist and criminal acts. Key areas of society, such as border control, criminal identification, and electronic transactions, demand more reliable solutions provided by [Biometric Systems \(BS\)](#). Accordingly, both the public and private sectors have been increasing the investment in these solutions in the last decade [1,2].

The field of biometrics is a pattern recognition problem, where the individual traits of a person are coded, stored into a database, and compared with the recorded biometric data according to two approaches: identification (finding the registered person from a sample) and verification, also known as authentication (prove that the person is whom it claims) [1,3,4]. These systems [BS](#) may be evaluated

in several dimensions, such as universality, uniqueness, permanence, measurability, performance, acceptability, and circumvention [3,5].

Since the application of elementary fingerprint recognition in 1883, research has been searching for different and more reliable types of biometric data [6]. Nowadays, fingerprints, faces, hands, veins, voices, irises, and other biological and behavioral features are used to identify subjects [7,8]. An example of another potentially useful biometric signature is physiological signals. These are electric signals generated from the electrochemical changes in nerve cells, muscles, or gland cells, which may be acquired from surface electrodes in contact with the skin. Some examples of physiological signals used in biometrics are: the [Electrocardiogram \(ECG\)](#), extracted from the heart; the [Electromyogram \(EMG\)](#), from muscle activity; the [Electrodermal Activity \(EDA\)](#), which tracks changes in skin conductivity, expressing the activity of the sympathetic nervous system; and the [Electroencephalogram \(EEG\)](#) [9], extracted from brain activity [3,10].

1.1. [ECG Biometrics](#)

Given its morphological signature and advances in sensing devices, the [ECG](#) has emerged as a biometric modality with the promise of robustness against circumvention attacks and the ability for continuous pervasive acquisition scenarios. Notwithstanding this, using non-invasive [ECG](#) presents its challenges, such as noise, artifacts, and intra-subject variability [10,11]. The computation of features may prove even more arduous when the contamination is present in characteristic points of the waveform [12].

The sources of noise and more abrupt changes of the signal (artifacts) include electrode material, sensor location, power-line interference, movement artifacts, and instrumentation of the devices [13]. The intra-subject variability may be due to changes in the health status of the patient, heart-rate variability, physical exercise, affective and emotional states, and drug consumption [10,11]. Nevertheless, the [ECG](#) remains stable over the years, sufficiently to allow recognition with an error rate of 3% [14].

The morphology of the [ECG](#) cycle is related to its electrophysiological activity and comprises the P, PQ interval, QRS complex, ST interval, and T waves [11,14]. According to the used features, [ECG](#) biometric systems can be categorized as fiducial-based, in which heartbeat templates are composed of features related to the relevant points of each cycle, and non-fiducial-based, which consider the [ECG](#) signal as a whole, computing its features accordingly [1]. The inclusion of both types of elements, to achieve better results, defines these systems as partially-fiducial [10,15,16].

1.2. [Objective](#)

[ECG](#) biometric systems struggle to find their way outside controlled settings due to the intra-variability existent in real-life scenarios. Therefore, this work aims to provide one step further in the abstraction of the individualized [ECG](#) signal to be used in real-life scenarios using state-of-the-art technologies such as [Deep Neural Networks \(DNN\)](#).

In order to achieve this goal, two systems with distinct architectures are proposed (Figure 1) as an effort to improve performance in both identification and authentication processes without extracting human-crafted features: (1) a non-fiducial system which uses a [Recurrent Neural Networks \(RNN\)](#) capable synthesizing [ECG](#) signals, outputting a score based on the error of prediction; (2) a partial-fiducial system which uses a [Temporal Convolutional Neural Network \(TCNN\)](#), giving a score based on the output of the last layer. Both scores computed and fed to the [Relative Score Threshold Classifier \(RSTC\)](#), which classifies a single sample window or a chunk of windows. Both systems will be benchmarked in three public databases used in [BS](#) literature: Fantasia ([17]), MIT-BIH ([18]), and CYBHi ([13,19]).

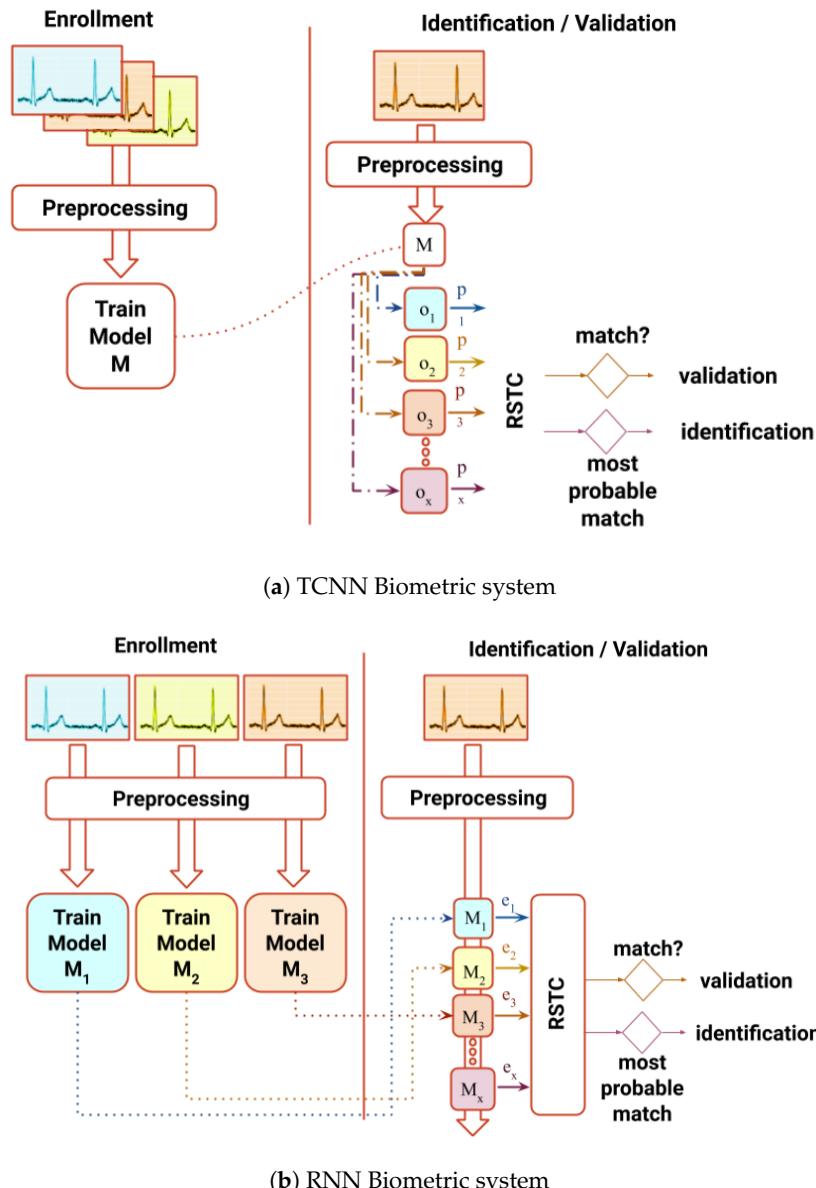


Figure 1. Proposed biometric systems. In (a) the trained model extracts features and has an output which is used to calculate the probabilities— p —for each class, as for (b) the error of prediction— e —of each model is used. **RSTC** is the classifier that validates or identifies which person does the input sample belong.

2. Related Work

The literature that refers to the application of **DNN** architectures in **ECG** biometry is mainly comprised of feed-forward or **Convolutional Neural Networks (CNN)** architectures. The practice of using human-extracted features may be seen in [20] where a temporal frequency spectrograms are produced with discrete wavelet transform. The extracted features from these time-frequency matrices are fed into a feed-forward network for classification. As for the authors of [21], the feature extraction module is bypassed, using the template of the **ECG** cycle as input to a feed-forward network, for both feature extraction and classification. **CNN** architecture is also used in this context, in the works of [22,23], this architecture is explored in both authentication and authentication biometric processes. These fiducial systems feed their networks with the average cycle of the signal, based on the neighborhood of the R peak of the **ECG** signal. The systems developed by [24–26] also use

DNN technologies, but since they are trained and tested with the benchmarked datasets, they will be discussed later in the "Fantasia", "MIT-BIH", and "CYBHi" subsections.

The improvements of these methods cover the generalization of the abstract notion of an individualized ECG by employing one-dimensional time-series, as opposed to images or human-extracted features and limiting the need to extract the QRS complexes, by developing non-fiducial and semi-fiducial systems. This claim is empowered by including an off-person database which has high intra-variability in most of the subjects. For benchmarking purposes, the rest of this section will address related work using the referenced databases: Fantasia, MIT-BIH, and CYBHi.

2.1. Fantasia

PhysioNet was created under the auspices of the National Center for Research Resources of the National Institutes of Health [27] and comprises several public physiological databases, such as MIT-BIH and Fantasia. The Fantasia 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 its homonym Disney movie [17].

This database was adopted by [28], 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 $90 \pm 8\%$ for PCA, False Rejection Rate (FRR) 5% for LDA, and False Negative Rate (FNR) of 4% for RS [28]. The same team published a method that extracts and decomposes R-R intervals using Discrete Wavelet Transform (DWT) and achieved 95.89% accuracy, FRR 0%, and FNR 5% [29]. The work of [15] includes the analyses of the influence of the heart rate variability in the QT intervals and suggests their correction to improve performance, leading to the identification rates with MultiLayer Perceptron (MLP) and Support Vector Machine (SVM) of 97% and 99%, respectively. The study made by [24] feeds spectrograms' extracted features into a feed-forward network, such as [20], achieving an accuracy of 97.2% for the Fantasia dataset; this work also has a reference to the MIT-BIH database.

2.2. 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 360 Hz 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 seven individuals in MIT-BIH Long-Term [18].

This database has been used not only for validating arrhythmia detectors and other cardiac dynamics analysis but also in the biometrics field. The MIT-BIH Normal Sinus Rhythm subset was adopted by [30–32] to validate their models. The feature extraction method proposed by [30] is comprised of Piecewise Linear Representation and Dynamic Time Warping (DTW) methods, and the computation of similarity measurements is used for classification. Overall, the identification results were close to 100% with the minimum of the half total error (HTER) of 0.2%. As for [31], the accuracy has reached 99.07% by using a combination of ten features from heartbeat and sixty from coefficients of the Hermite Polynomials Expansion fed into a Hidden Markov Model (HMM). In another case, several classification methods were applied with characteristics of the QRS complex by [32] 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 [32].

Combined subsets of the MIT-BIH were used by [24,33]. The first uses a semi-fiducial approach in which a Wavelet Transform was applied together with an Independent Component Analysis (ICA) analysis to detect each heartbeat. This information concatenated with the R-R intervals information and fed into a SVM classifier. For the selected 23 records, which include the Normal Sinus Rhythm and the Arrhythmia subsets, this work reached 86.4% accuracy for subject evaluation, even though the focus was to classify heartbeat classes. The mentioned work from [24] achieved

90.3% of accuracy for 47 individuals of MIT-BIH. The work developed by [25] claims to achieve a 98.55 % of accuracy in identification paradigm using a bi-directional **Gated Recurrent Units (GRU)** network model incorporating for 47 records of the MIT-BIH Arrhythmia Database. In this article, they compare several architectures, including **CNN** and **Long Short-Term Memory (LSTM)** models with a fiducial approach.

2.3. CYBHi

The *Check Your Biosignals Here initiative* (CYBHi) dataset was acquired with an off-person procedure without the use of skin electrodes. This database has been considered by [13] as one of the two best public databases in terms of acquisition protocol and hardware for biometric studies, in comparison with fifteen databases. The setup of the acquisition system consists of a pair of synchronized sensors that were hand-shaped, in which the electrodes were placed on the fingers. The data consists of two 2-min acquisitions three months apart of 63 participants, 14 males, and 49 females (18–24 years old) [19]. For this paper, the sessions will be named M1 and M2.

In literature, the outcome for authentication of [19] was an **Equal Error rate (EER)** of 9.1% using an **SVM** classifier against the correlation between features extracted from ECG cycle templates. Concerning [5], the extracted features from the **ECG** morphology are joined with the features extracted from the R-R intervals with **Nearest Neighbor Algorithm (NNA)** and used for classification using a Euclidean distance classifier. The identification rate of this study is 95.2% for the first test and 90.2% for the second test. The work of [26] uses two **CNN** networks are fused together, one fed with the raw **ECG** cycle, while the other uses the cycle's spectrogram, described before, achieved **EER** values of M1 vs. M1 of 1.33%; M1 vs. M2 12.78%; M2 vs. M1 13.93%.

3. Materials and Methods

As stated before, two architectures were explored and compared using the same classification method, the **RSTC**. The setup comprises **RNN** and **TCNN** approaches without the use of human-extracted features. Even though both procedures are trained to create an internal representation of the **ECG** waves, they have different outputs for calculating the same similarity score. While, for the **RNN** approach, the score is based on the prediction error, the **TCNN** score is given by the output given by the last layer, the logits layer. Before training and testing models, heavily corrupted windows were removed by using the algorithm developed by [11] using the standard deviation feature.

Cross-validation was used for training and testing. For Fantasia and MIT-BIH, the training was made within the first 33% of the **ECG** records, while the rest was reserved for the testing set. As for the CYBHi database, the training and testing sets were divided for 50% of the dataset when each session was tested, while, when crossing both sessions, 100% of the respective session was used. More details will be given in the Results section. All of the inter-sectioned windows were removed to ensure that both datasets did not share information between them.

3.1. Recurrent Neural Network Approach

The synthesizing of **ECG** signals in the previous work [34] provided a hypothesis that, if the model is trained with the signal from subject A and fed with the signal from subject B, the prediction error will be higher than if fed with the signal from the original subject A. Subsequent to the observation of the results while testing this premiss, it was confirmed that it could be further explored. The hyper-parameter tuning was made through careful observation and continuous try-and-error procedures in search of better results using a subset of the Fantasia dataset.

In the **RNN** approach, data are firstly pre-processed (Figure 2), and later transformed by an embedded matrix— E , three sequential **GRUs**— G , a dense layer with linear activation and a softmax node [34].

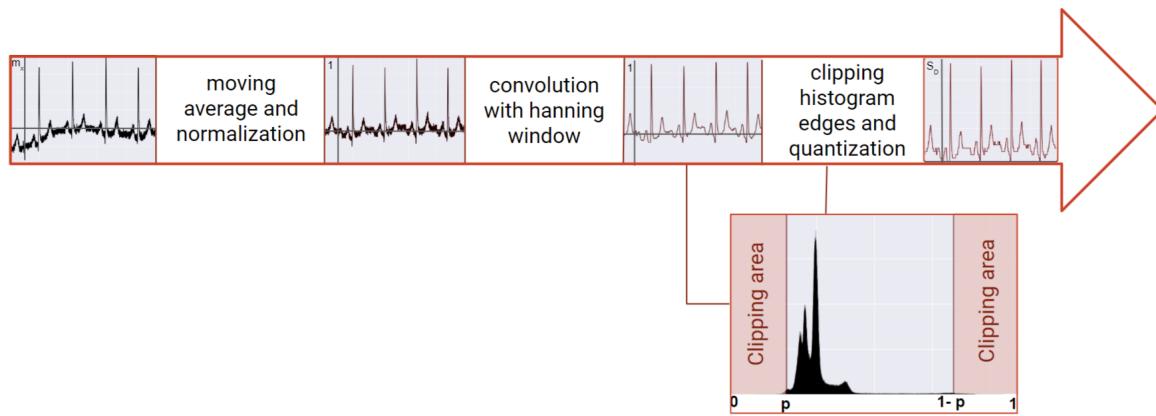


Figure 2. Pre-processing phase for the [RNN](#) procedure.

3.2. Pre-Processing

The pre-processing step starts with removing the moving average, followed by a convolution with a *Hanning* window and normalized by a moving absolute maximum window. The benefit of this approach is to mitigate the unwanted low and high frequencies resultant from the signal acquisition. Following this procedure, the edges of the signal's amplitude histogram were clipped, according to a value of confidence, to reduce the influence of artifacts. This value was typically 0.5% but could increase depending on the level of corruption of the signal. Due to the nature of the [GRU](#) networks, the quantization of the signal (s) must be performed. This step consists of reducing the dimension of the possible amplitude values. After the transformation, the signal will be comprised of S_D (signal dimension) number of possible values, which was set to 512. The higher the value of the signal dimension (S_D), the higher the detail is retained from the signal. In summary, the resultant signal (x) is obtained with 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 and s_n is the raw n -th sample of the raw signal (s). The variable $k \in \{0, 1, \dots, S_D - 1\}$ represents the number of possible integer values x_n may take.

Finally, the signal was segmented into windows of size W with overlap between them. The overlap percentage was 67% for the "Fantasia" and "MIT-BIH" datasets and 89.1% for the "CYBHi" datasets, to increase the amount of data.

3.3. Architecture

The architecture depicted in Figure 3a comprises one embedding matrix (E), 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 that maps to the corresponding x_n -th column of the matrix E . This matrix initialized randomly, but it is optimized during the training process, adjusting itself to the possible input values along with the other parameters.

As stated before, in [34], this architecture is shown to synthesize biosignals, validated by the prediction error. As the signal which trained the model is fed into itself, the prediction error is lower in contrast to the signals with other origins. In the case of biometry, this notion is expanded to the specific source of the signal, i.e., this network should be able to learn the [ECG](#) individual intricacies of each subject. Following this reasoning, the produced similarity score ($S(p, i, w)$) is given by the following equation:

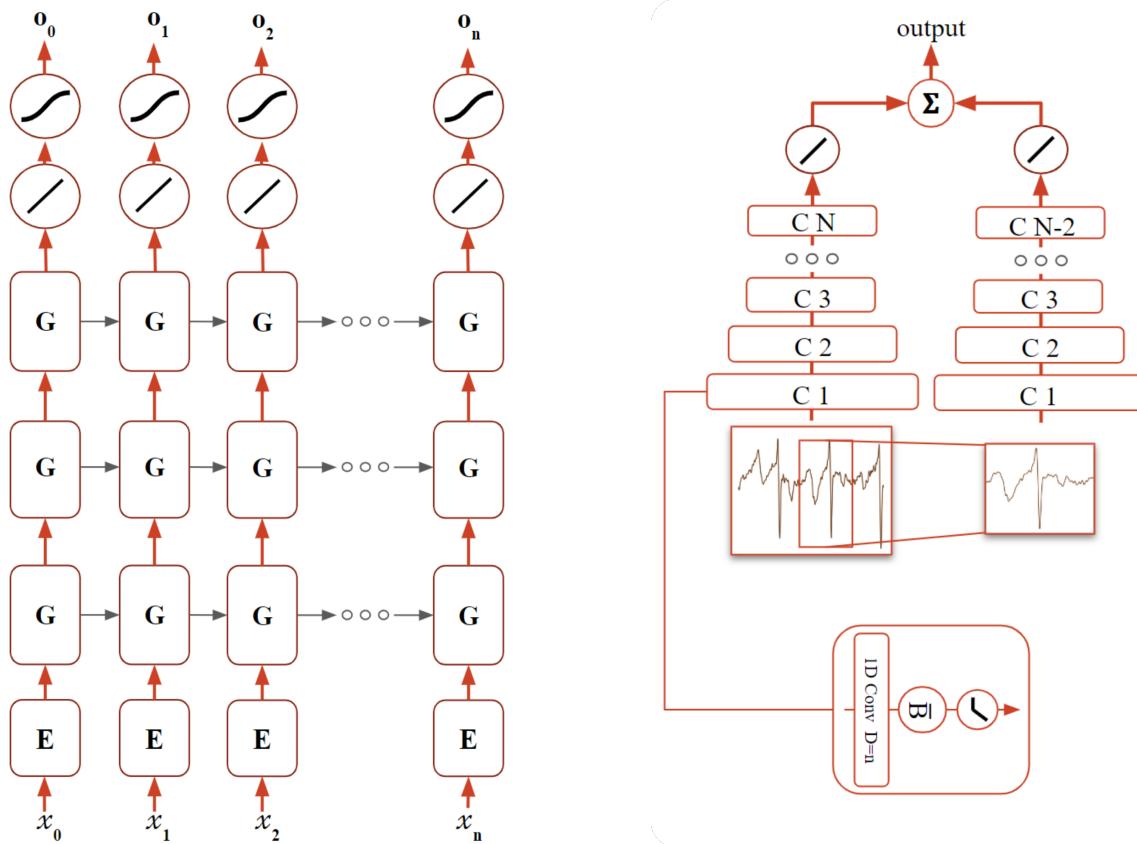
$$S(p, i, w) = \frac{e(p, i, w)}{\max(e(:, i, w))}, \quad (2)$$

where e is the prediction error made by each predictor— p —for the time-window w of the signal which belongs to the subject i . The expression $\max(e(:, i, w))$ represents the maximum for the same window w for all predictors. Each model is trained for each subject while predicting the next sample amplitude. These models were trained using mean-squared error using the RMSProp [35] optimization algorithm. The prediction error is provided by the loss function given by the following equation:

$$e_p = \frac{1}{W} \sum_{n=0}^{W-1} (\hat{y}_n - y_n)^2 , \quad (3)$$

where the W is the window size (number of samples in the time-window), \hat{y}_n the predicted n -th sample and y_n the real sample value. Note that the s_n is the raw signal, x_n is the quantized signal, while y_n is equal to x_n dephased by one sample since the RNN architecture predicts the next sample [34].

In summary, the parameters required for this architecture are the time-window size (W), the overlap ratio, the edges cropping ratio, the quantization dimension of the signal (S_D), and the number of hidden nodes in each GRU (H_D).



(a) Recurrent Neural Network Approach

(b) Temporal CNN Approach

Figure 3. Proposed architectures. (a) depicts the whereRNN approach, where each input (x_n) is fed into an embedded matrix (E) and, consequently, into a sequence of three GRU (G) cells. The output of this sequence is transformed through a fully connected network and a softmax function resulting in o_n . (b) depicts the TCNN approach, where a time window (TW) and the first complete cycle are input to two different sequences of CNN blocks (CN) each comprised of a one-dimensional convolutional layer with an N dilation factor, batch normalization, and a ReLU activation function. One of the networks has an N number of blocks while, for the other $N - 2$, they both end in a fully connected network,. Finally, both of the resultant vectors are summed producing the logit vector for classification.

3.4. Temporal Convolutional Neural Network

The second proposed approach is a two-stream [TCNN](#) [36], which uses one-dimensional convolutional layers with dilated convolutions to learn temporal patterns. It combines predictions from two different inputs: [ECG](#) non-fiducial window segment and a sub-segment of the same window containing only one full cycle centered in R peak of the QRS complex. Earlier configurations of this [CNN](#) were assembled and experimented, such as including spectrograms as input [37], and others including a one-dimensional [CNN](#) without the use of dilated convolutions. After testing using Fantasia and CYBHi datasets with try-and-error parameter hyper-tuning, the resultant architecture provided the best results in relation to the previous versions.

Before data are fed into the networks, the signals are pre-processed by removing the moving average, convolution with the *Hanning* window, normalized with the absolute maximum, and finally segmented, analogous with the [RNN](#) procedure.

Within this architecture, depicted in Figure 3b, each 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. The number of convolutional layers for the Fantasia and MIT-BIH databases is 6 of sizes 256, 128, 64, 16, 8, and 4 for the non-fiducial network and 4 of 64, 16, 8, and 4 for the fiducial network. The input window had 256 samples for these databases. For the CYBHi database, due to the increase of the sampling frequency, each network had an increase of two initial layers, 1024 and 512 added to the previous layers of the non-fiducial network, and 256 and 128 added to the previous layers of the fiducial network. The input window of this dataset had 1028 samples.

Each [CNN](#) was trained using the Adam optimizer [38] with the cross-entropy loss function (L), given by:

$$L = -\frac{1}{I} \sum_i^I [i \log (\hat{i}) + (1 - i) \log (i - \hat{i})] \quad , \quad (4)$$

where the I is the number of subjects, \hat{i} is the predicted subject index, while i the real subject index. Subsequently, after each network is trained independently, 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), according to this equation:

$$S(p, i, w) = 1 - \frac{o(p, i, w)}{\max(o(:, i, w))} \quad (5)$$

The parameters specific for this architecture are the time-window size (W), the overlap ratio, the size, and the number of filters in each of the convolutional layers, and the size of the fully connected layer before the logit layer. The number of layers depends on the size of the input window.

3.5. Relative Score Threshold Classification

The [RSTC](#) is a simple method that classifies by choosing the lowest normalized similarity score ($\bar{S}(p, i, b)$) for each batch of windows (b); consequently, S is a three-dimensional tensor with dimensions $(I \times I \times B)$, where B is the size of the batch. This measurement is calculated after normalizing in relation to each predictor. When B is higher than one, then only the minimum of that batch of windows is considered, according to the following function:

$$\tilde{S}_{p,i,b} = \min \left[\left\{ S_{p,i,B \cdot b}, S_{p,i,B \cdot b+1}, \dots, S_{p,i,B \cdot b+B} \right\} \right] \quad , \quad (6)$$

where the notation $B \cdot b$ represents the multiplication of the batch number by the batch size, for index purposes. The following normalization for each b in relation to the p is determined by:

$$\tilde{S}_{p,i,b} = \frac{\tilde{S}_{p,i,b} - \min(\tilde{S}_{:,i,b})}{\max[\tilde{S}_{:,i,b} - \min(\tilde{S}_{:,i,b})]} , \quad (7)$$

where minimum and maximum values are calculated in respect to all predictors for each batch of windows. Considering that $\tilde{S}(p, i, b) \in [0, 1]$, the value 0 encodes to the most probable class, while 1 codes to the lower probable one. The predicted class for the identification paradigm (C^I) for each individual (i) and batch of windows (b) is given by:

$$C_{i,b}^I = \arg \min_p [\tilde{S}_{:,i,b}] \quad (8)$$

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 on the [Receiver Operating Characteristic \(ROC\)](#) curve, when [FNR](#) = [False Positive Rate \(FPR\)](#) for each binary classifier (C^A). The authentication paradigm is evaluated with this measurement because these systems prioritize the compromise between the minimization of the possible imposters to breach the system, with the maximization of the number of rightful individuals who can access the system. Each binary classifier is made by comparing the score of all predictors p for each individual i and all windows w with a changing threshold (δ). This system is fine-tuned for each individual and the output gives if that specific b of windows belongs to the claimed individual:

$$C_{p,i,b}^A = \begin{cases} 1, & \text{if } \tilde{S}(p, i, b) \leq \delta \\ 0, & \text{if } \tilde{S}(p, i, b) > \delta \end{cases} \quad (9)$$

When $C_{p,i,b}^A$ has the value 1, the classification for class A is positive, while 0 when negative. The [ROC](#) is calculated with an array of changing thresholds and the [EER](#) may be extracted for all the intersections with the diagonal where [FNR](#) = [FPR](#). Since each individual has its own fine-tuned [ROC](#), the mean and standard deviation must be calculated and used for evaluation. An example of how these thresholds are obtained is depicted in Figure 4.

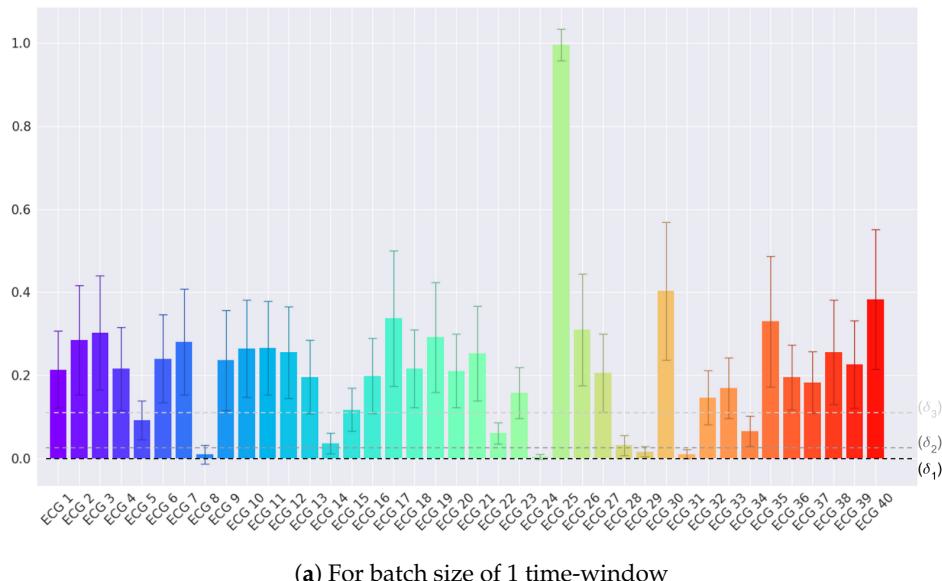
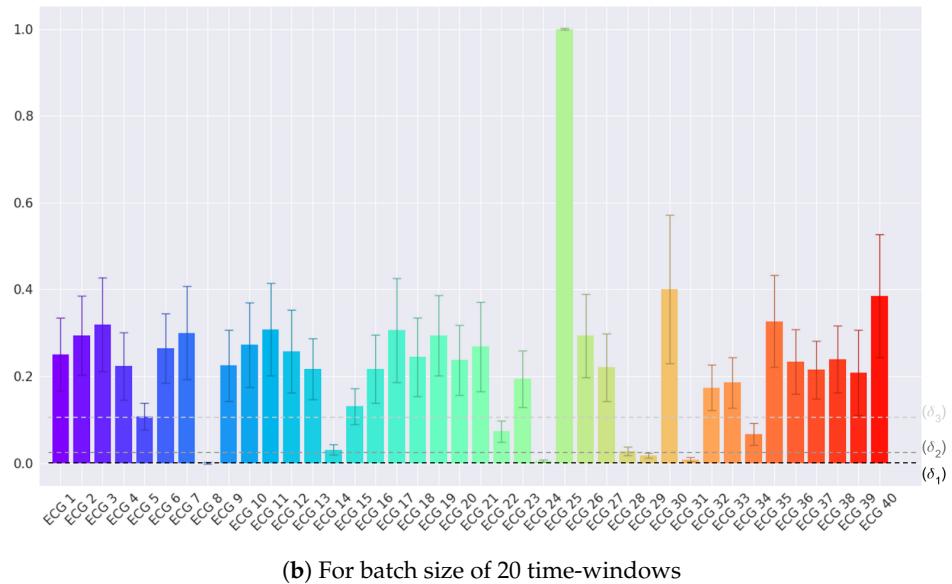


Figure 4. Cont.



(b) For batch size of 20 time-windows

Figure 4. Distribution of the score of the RNN network for the predictor ECG 8 for each of the ECG signals of the Fantasia Dataset. While (a) refers to a single window of 512 samples, (b) 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.

4. Results

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

4.1. Fantasia

All data were segmented in windows of 512 samples (approximately 2 s) with an overlap of 67%. After the rejection algorithm, the selected windows were on average 1464 windows of a total of 3787 per subject.

The score distribution per subject when the test set is fed into the RNN model trained with ECG,8 is depicted in Figure 4. The scores distribution, which is presented in Figure 4a, considers all batches of windows with $B = 1$ and $B = 20$ in Figure 4b. When analyzing this example, it can be observed that, with the increase in the number of windows contained in each batch, both average and standard deviation are reduced. The three thresholds presented in Figure 4 will have a different binary classification for the predicted ECG,8 when fed to different values of B . In the case of (a), the δ_1 threshold will give a positive value for most of the ECG 23 batches, which is false. When both δ_2 and δ_3 thresholds are considered, most of the ECG,8 batches will increase, at the cost of increasing also the number of imposters.

When the B increases to 20, the variance and average value of ECG 8 will decrease significantly. In this case, δ_1 will now classify correctly most, or even all, of the ECG 8 batches, while reducing the number of ECG 23 classified incorrectly, ensuring lower values for the FNR and FPR. Both δ_2 and δ_3 thresholds will produce more imposters while maintaining the same number of correct classifications. The ROC curve is generated for each predictor and each batch size, producing different values for EER that can be interpreted for authentication systems.

The identification rate, i.e., accuracy for the identification paradigm, for both algorithms is depicted in Figure 5. 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 one minute of signal per batch. While the RNN algorithm increases until reaching close

to 100% at approximately 112 s, the **TCNN** algorithm reaches 99.1% after approximately 90 s, but its curve displays a steady behavior. The best results are presented in the confusion matrices of Figure 5b for **RNN** and Figure 5c **TCNN** approaches.

The results for the authentication mode are depicted in Figure 6a,b, which show the evolution of the **EER** per time contained 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 and sooner than the **RNN** counterpart.

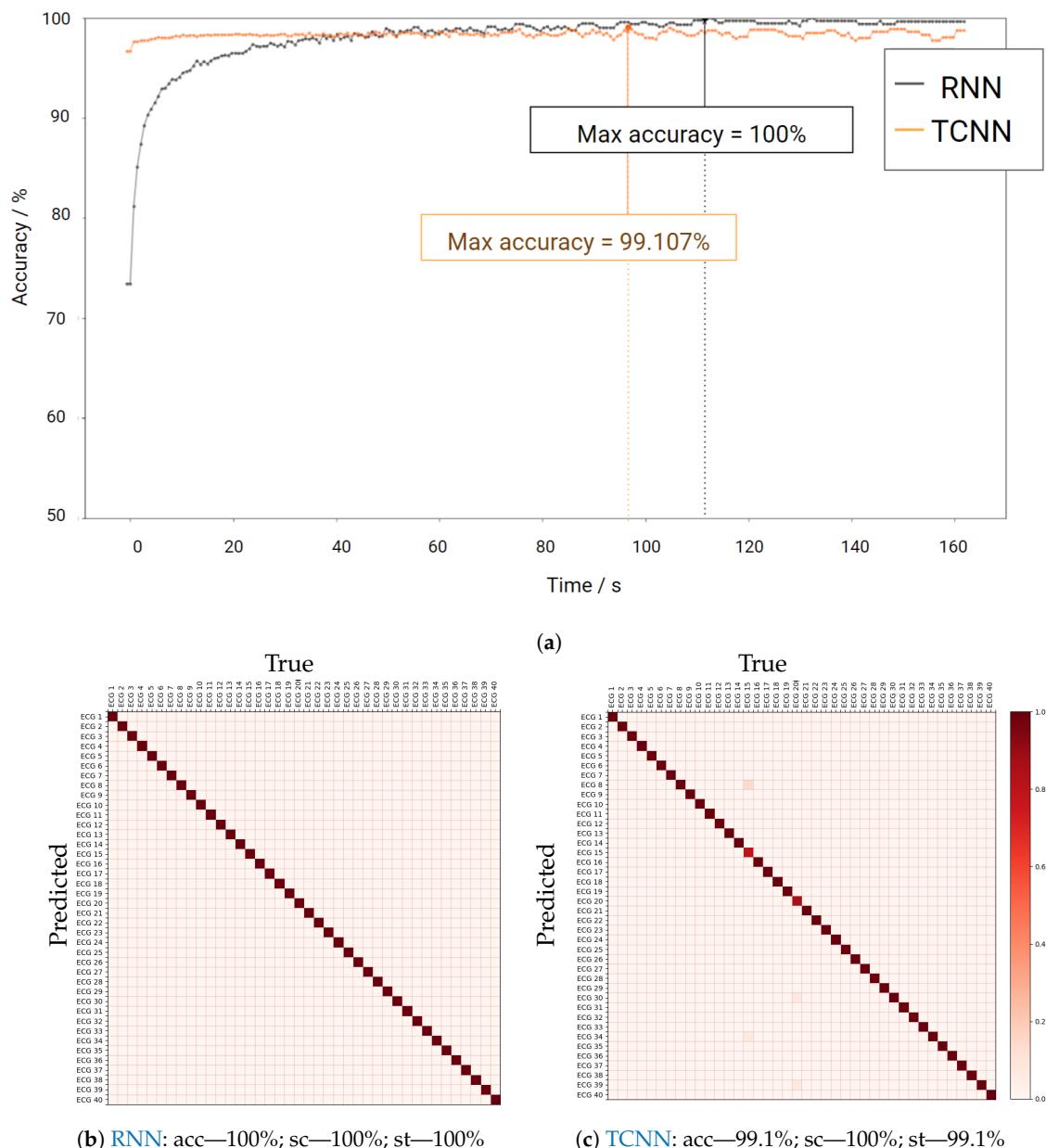


Figure 5. Results for the identification paradigm related to the Fantasia dataset. (a) displays the evolution of accuracy for the Fantasia Dataset with increasing batch size. (b,c) depict the confusion matrices for the best identification performance for both architecture approaches (ac is accuracy, sc is specificity, and st is sensitivity).

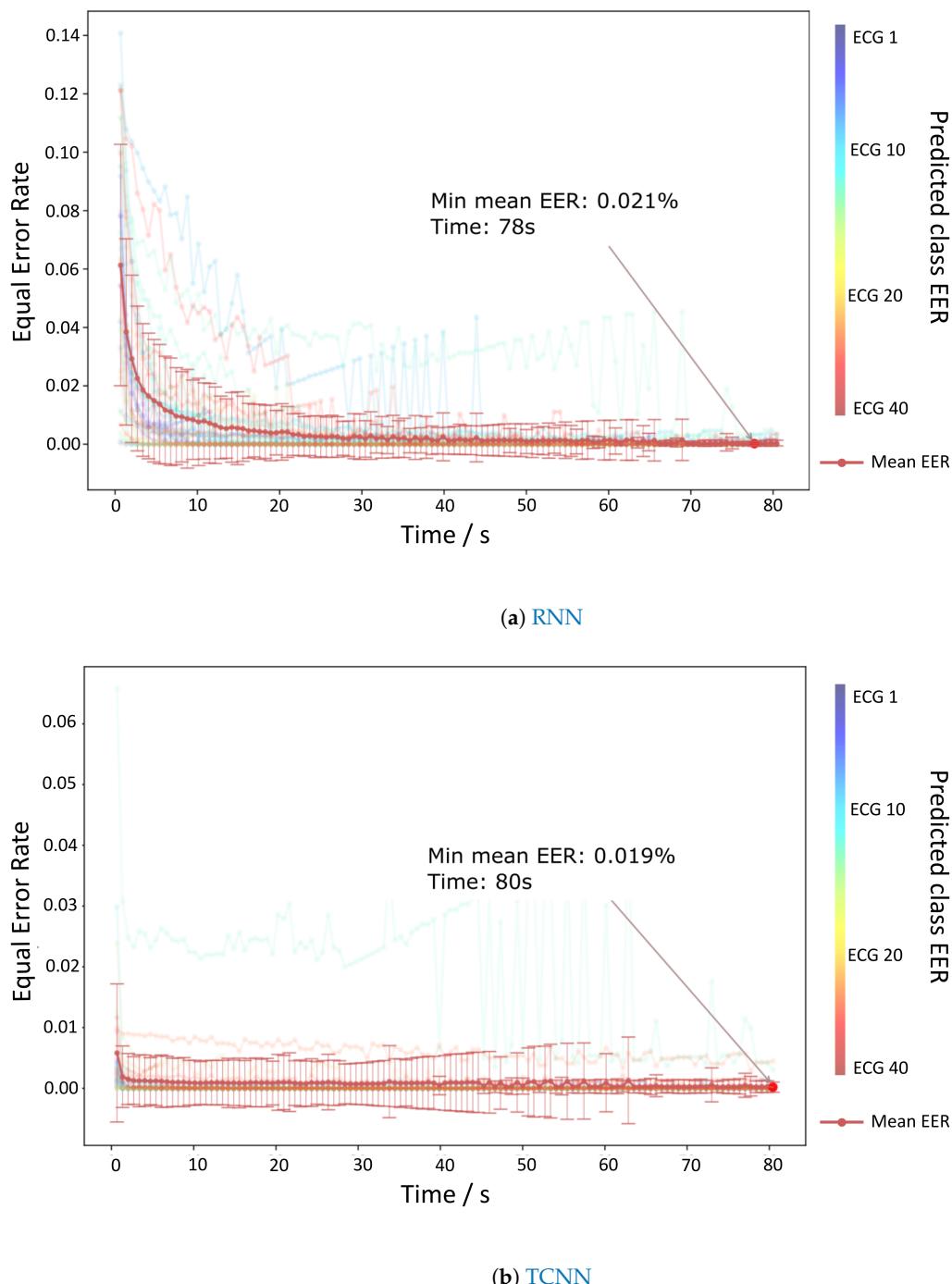


Figure 6. Results for the authentication paradigm related to the Fantasia dataset. (6a) and (6b) depict the evolution of the EER values over time for both approaches.

4.2. 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 of the noise removal algorithm was on average 42% for each subject.

The conducted identification test utilizing the MIT-BIH database can be observed in Figure 7, showing 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 (Figure 7). The RNN approach does not go above 93% for all the MIT-BIH but reached almost 100% accuracy for Normal Sinus Rhythm. When analyzing Figure 7b, it is possible to see the

diagonal clearly, but some trained models generalize the ECG signals instead of the individual traits and, consequently, classifying incorrectly. Some possible limitations include the loss of information during resampling, sensibility to noise, and symptomatic behavior of the arrhythmia events. The last option is supported with the results close to 100% accuracy for the Normal Sinus Rhythm subset.

Concerning the **TCNN** approach, 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 stability and robustness to noise. The confusion matrix depicted in Figure 7c shows that only two predictors display a faulty behavior, probably due to the presence of arrhythmia events. When examining the individuals from the Normal Sinus Rhythm, the accuracy reached almost 100% after approximately 10 s (Figure 7a).

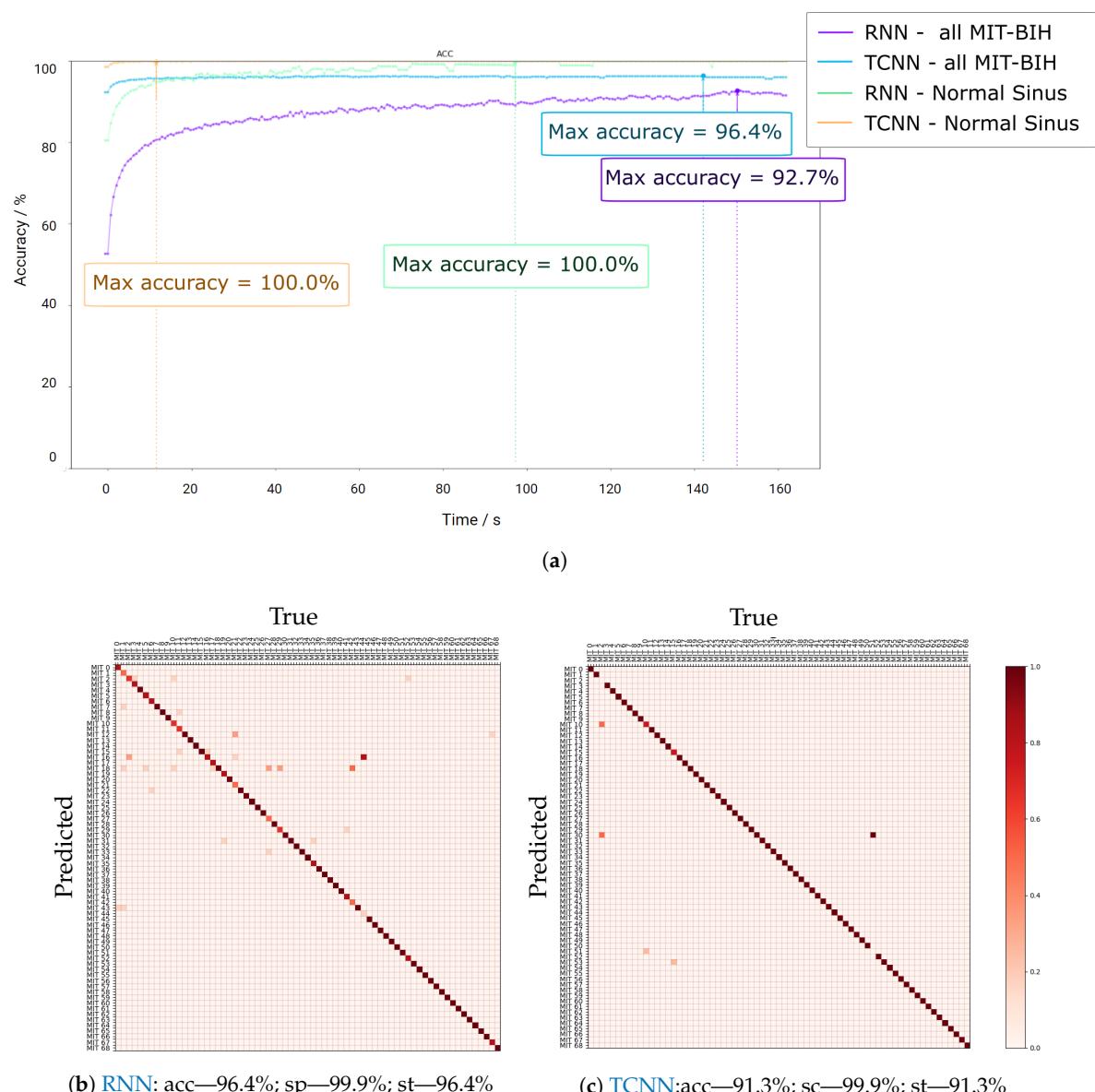


Figure 7. Results for the identification paradigm related to the Fantasia dataset. (a) displays the evolution of accuracy for Fantasia Dataset with increasing batch size. (b,c) depict the confusion matrices for the best identification performance for both architecture approaches (ac is accuracy, sc is specificity, and st is sensitivity).

The authentication evaluation of all the MIT-BIH individuals is displayed in Figure 8a,b. These figures show that the **TCNN** approach reaches **EER** values close to 0% with 72 s in each

batch and approximately 1.5% for RNN. Even though the best result for TCNN is after 1 min, the values are quite low from the start.

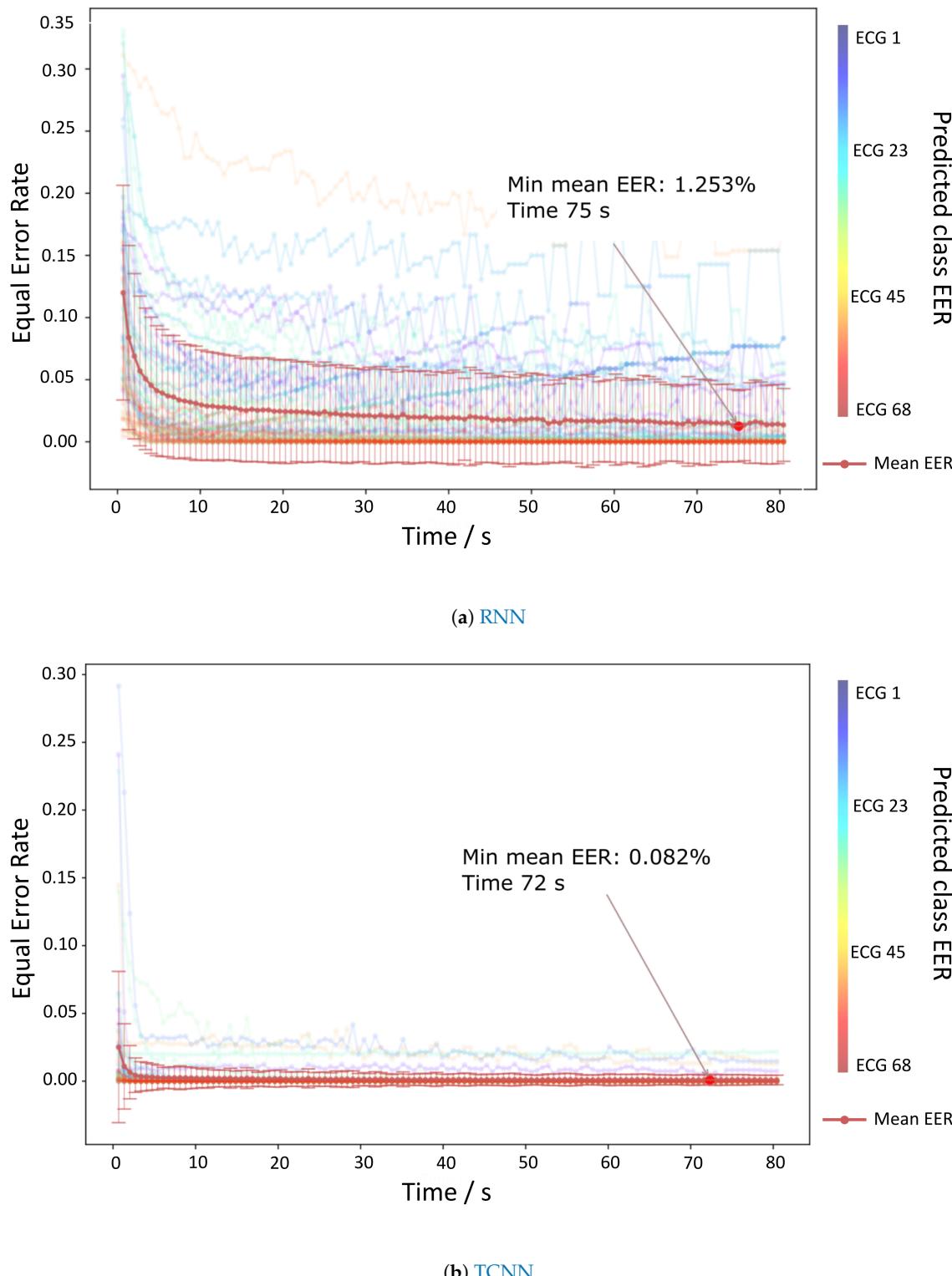


Figure 8. Results for the authentication paradigm related to the MIT-BIH dataset. (a,b) depict the evolution of the EER values over time for both approaches.

4.2.1. CYBHi

As mentioned before, the CYBHi database comprises two sessions with a time distance of three months. Therefore, both algorithms are evaluated with all the combinations of both sessions (M1 and M2) for 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”.

The segmentation of the signal 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. As stated before, when the training and testing session was the same (“M1 vs. M1”, “M2 vs. M2”), the data were separated by 50% and 100% of each session across sessions (“M1 vs. M2” and “M2 vs. M1”).

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. Figure 9a shows that the **TCNN** algorithm achieves 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 sessions was significant as expected, revealing accuracies only above 75 % for both “M1 vs. M2” and “M2 vs. M1”. The low accuracy presented by the **RNN** approach reflects the sensitivity to noise. Figure 9b shows that the **EER** achieved 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%, probably due to the temporal proximity of the training and testing sets. As for the evaluation across different sessions, the **EER** for the **RNN** approach is 18.3% and 18.8% for “M1 vs. M2” and “M2 vs. M1”, respectively, and 2.2% and 4.1% for the **TCNN** in the same order. These results outperform the current state-of-the-art for off-person measurements.

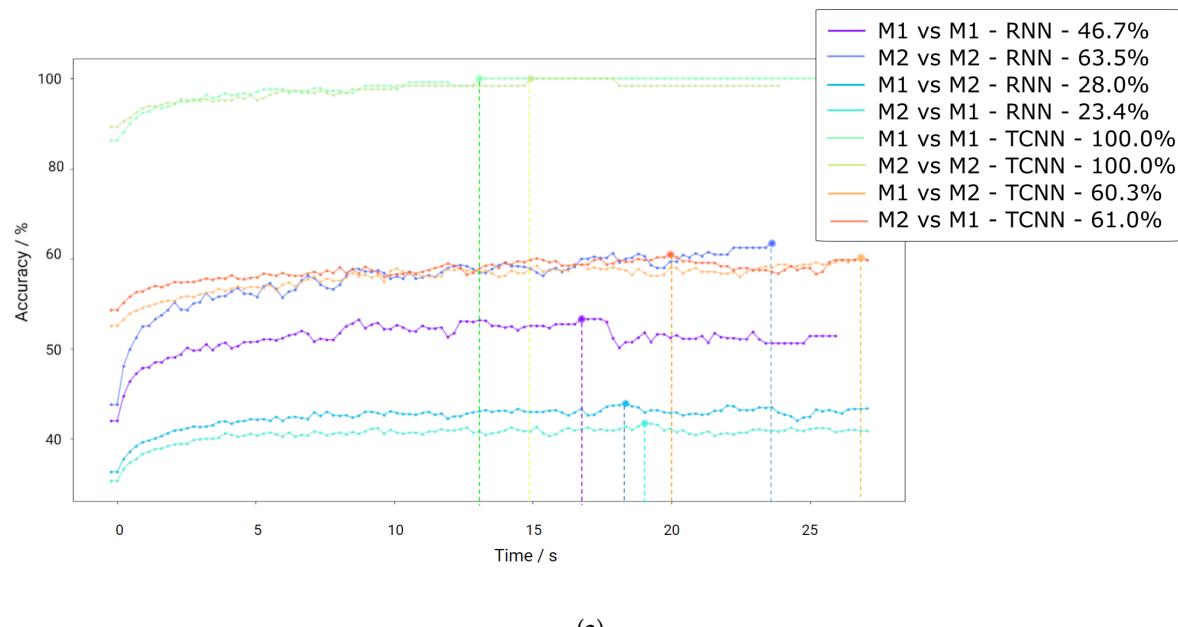


Figure 9. Cont.

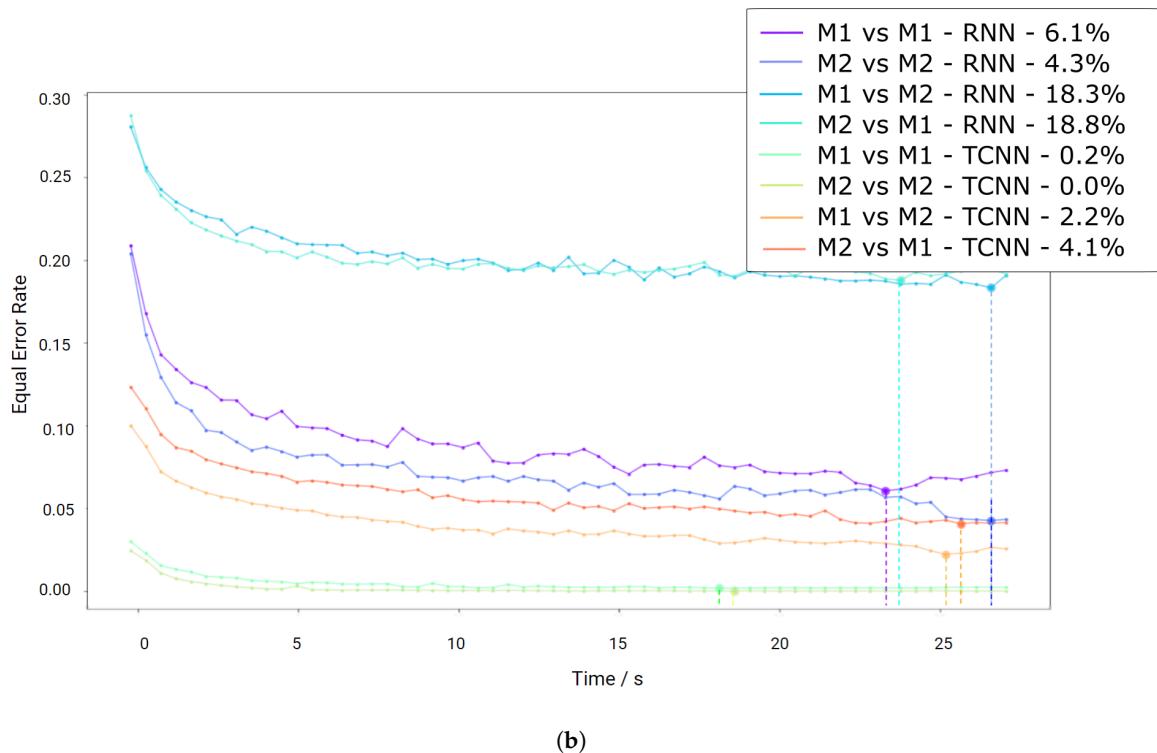


Figure 9. Results for the CYBHi dataset for each approach. M1 represents the first session while M2 represents the second session. (a) depicts the results for the identification paradigm, displaying the evolution of accuracy with increasing batch size. (b) displays the evolution of EER with increasing batch size for the authentication paradigm.

5. Discussion

Table 1 summarizes the results per database concerning the consulted literature. This report suggests 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. The only exception was for the "CYBHi M1 vs. M2" results for the identification modality, as [Lourenço et al. \(2012\)](#) [5] has better results, which can be explained mainly due to two reasons: (1) only 32, as opposed to 68, people were used and (2) the procedure was completely fiducially-based. The high accuracy that both approaches obtained comes with a cost of training time. For example, the **RNN** training time, using the CYBHi database and a computer with an Nvidia GTX 1080Ti GPU, had an average of 36 h per subject, totaling approximately 2300 h for this database, but only takes a few seconds to print the testing results. For the implementation of this system at an industrial level, this limitation should be mitigated as retraining should be faster, especially when the number of individuals increases.

It is possible to observe that the results for the CYBHi dataset regarding inter-sessions are lower due not only to the increase of the noise and artifacts displayed in the signal but also to the changes in the mental and emotional states of each user. This statement emphasizes the challenge that the biometrics systems face due to the intra-variability of the subjects. Hence, more acquisition setups should be made, such as expanding the number of people, extending the acquisition time, and increasing the number of acquisition sessions. Added to this, the amount of cleaner data would impact significantly the training of the models and, consequently, these measures would not only provide higher accuracy and **EER** but also more confidence in the results shown to claim that the systems were capable of a successful generalization.

Table 1. Fantasia Dataset.

Study	Database	Acc	EER
Tantawi et al. (2013) [28]	Fantasia	90.0%	—
Tantawi et al. (2013) [29]		95.9%	—
Gargiulo et al. (2015) [15]		99.0%	—
RNN		100%	0.02%
TCNN		99.1%	0.02%
Wang et al. (2008) [39]		98.1%	—
Fatemian and Hatzinakos (2009) [40]	MIT-BIH Sinus	99.6%	—
Wang et al. (2008) [39]		94.5%	—
Rabhi and Lachiri (2013) [31]		99.0%	—
Sidek et al. (2014) [32]		99.1%	—
RNN		100%	0.6%
TCNN		100%	0.0%
Lynn et al. (2019) [25]	MIT-BIH Arr	98.6%	—
RNN	MIT-BIH Arr, Sin, Long	92.7%	1.5%
TCNN		96.3%	0.1%
da Silva et al. (2014) [19]	CYBHi M1 vs. M1	94.4%	—
Lourenço et al. (2012) [5]		95.2%	—
da Silva Luz et al. (2018) [26]		—	1.3%
RNN		46.7%	6.1%
TCNN		100%	0.2%
Lourenço et al. (2012) [5]	CYBHi M1 vs. M2	90.2%	—
da Silva Luz et al. (2018) [26]		—	12.8%
RNN		28.0%	18.3%
TCNN		60.3%	2.2%
da Silva Luz et al. (2018) [26]		—	14.0%
RNN	CYBHi M2 vs. M1	23.5%	18.8%
TCNN		61.0%	4.1%
RNN	CYBHi	63.5%	4.3%
TCNN	M2 vs. M2	100%	0.0%

The implementation of a biometric system including the **TCNN** approach in a real-life scenario recognizes several challenges: (1) the high computational time, which increases significantly with the number of samples per window and the number of individuals; (2) when a new person is added to the model, all training must be repeated; (3) this approach is viable for a suitable number of individuals, but to use on a scale of millions, the training and testing process will be costly or even impossible with state-of-the-art computational power and memory; and (4) 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.

For tackling these concerns, the use of transfer learning could prove useful. If the training of a network introduced a first learning stage with a different database for recognition of the basic structures and inner mechanisms of the **ECG** morphology, the second stage of learning, it would fine-tune the appropriate filters for source recognition and optimize faster. Data augmentation would also be a solution to improve the training procedure, but with the intent to avoid over-fitting of the training data, novel solutions should be addressed. The current **ECG** synthesis methods need to encompass the variability of the data, but to keep certain aspects of the frequency due to the non-fiducial aspects of these algorithms. The study conducted by [Belo et al. \(2017\)](#) [34] shows great promise for this endeavor,

but experiments have shown that the frequency of the synthetic ECG was very machine-like in terms of frequency, as it did not embed the semi-randomness of the cyclic behavior of the signal.

Even though the RNN achieves state-of-the-art results for the Fantasia and Sinus Normal Rhythm of the MIT-BIH, the sensibility to changes in the signal is higher than the aforementioned architecture. The main advantages of the RNN architecture are that all the models are independent, so there is no need to train them all again when adding a new person to the biometric system, the fact that it is non-fiducial, this system is viable for good quality acquisitions, and it is possible to generate the signal, in order to track if the model is well trained.

6. Conclusions

This study has 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 challenges presented by the intra-variability issue impacted the results of these architectures, it is possible to state that they improved comparing with the literature increasing one step towards the true abstraction of each subject's ECG signal.

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 to mitigate the loss of information while increasing the training speed. The next steps to implement in a real-life scenario would be the introduction of transfer learning in the training process, the acquisition of extra time windows per person, the addition of more individuals, and the extension of the number of sessions through time. The study of the impact of the noise and the scalability of these systems would also give an added value. Finally, the training/retraining/inference time of these systems should be analyzed and enhanced for real-life applications for delivering close to real-time solutions.

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.

Author Contributions: D.B. and H.G. conceived the presented idea. D.B. and N.B. developed the theory and performed the computations. H.S. and A.F. verified the analytical methods. H.G. supervised the findings of this work. All authors discussed the results and contributed to the final manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: Funding was provided by FCT—Fundação para a Ciência e Tecnologia, under the project PD/BDE/130216/2017.

Conflicts of Interest: The authors declare there are no conflict of interest.

References

1. Unar, J.A.; Seng, W.C.; Abbasi, A. A review of biometric technology along with trends and prospects. *Pattern Recognit.* **2014**, *47*, 2673–2688. doi:10.1016/j.patcog.2014.01.016.
2. Caldwell, T. Market report: Border biometrics. *Biom. Technol.* **2015**, *2015*, 5–11.
3. Van Den Broek, E.L. Beyond biometrics. *Procedia Comput. Sci.* **2010**, *1*, 2511–2519.
4. Nait-Ali, A. Hidden biometrics: Towards using biosignals and biomedical images for security applications. In Proceedings of 7th International Workshop on Systems, Signal Processing and their Applications (WoSSPA) Tipaza, Algeria, 9–11 May 2011; pp. 352–356. doi:10.1109/WOSSPA.2011.5931509.
5. Lourenço, A.; Silva, H.; Fred, A. ECG-based biometrics: A real time classification approach. In Proceedings of the International Workshop on Machine Learning for Signal Processing (MLSP) Santander, Spain, 23–26 September 2012; pp. 1–6. doi:10.1109/MLSP.2012.6349735.

6. Sun, Y.; Zhang, M.; Sun, Z.; Tan, T. Demographic Analysis from Biometric Data: Achievements, Challenges, and New Frontiers. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 332–351. doi:10.1109/TPAMI.2017.2669035.
7. Snelick, R.; Uludag, U.; Mink, A.; Indovina, M.; Jain, A. Large-scale evaluation of multimodal biometric authentication using state-of-the-art systems. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 450–455. doi:10.1109/TPAMI.2005.57.
8. Wu, Z.; Yamagishi, J.; Kinnunen, T.; Hanilçi, C.; Sahidullah, M.; Sizov, A.; Evans, N.; Todisco, M.; Delgado, H. ASVspoof: The automatic speaker verification spoofing and countermeasures challenge. *IEEE J. Sel. Top. Signal Process.* **2017**, *11*, 588–604. doi:10.1109/JSTSP.2017.2671435.
9. Palaniappan, R.; Mandic, D.P. Biometrics from Brain Electrical Activity: A Machine Learning Approach. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 738–742. doi:10.1109/TPAMI.2007.1013.
10. Nait-Ali, A. Beyond classical biometrics: When using hidden biometrics to identify individuals. In Proceedings of the 3rd European Workshop on Visual Information Processing, Paris, France, 4–6 July 2011; pp. 241–246.
11. Rodrigues, J.; Belo, D.; Gamboa, H. Noise detection on ECG based on agglomerative clustering of morphological features. *Comput. Biol. Med.* **2017**, *87*, 322–334.
12. Fang, S.C.; Chan, H.L. QRS detection-free electrocardiogram biometrics in the reconstructed phase space. *Pattern Recognit. Lett.* **2013**, *34*, 595–602. doi:10.1016/j.patrec.2012.11.005.
13. Merone, M.; Soda, P.; Sansone, M.; Sansone, C. ECG databases for biometric systems: A systematic review. *Expert Syst. Appl.* **2017**, *67*, 189–202. doi:10.1016/j.eswa.2016.09.030.
14. Wübbeler, G.; Stavridis, M.; Kreiseler, D.; Bousseljot, R.D.; Elster, C. Verification of humans using the electrocardiogram. *Pattern Recognit. Lett.* **2007**, *28*, 1172–1175. doi:10.1016/j.patrec.2007.01.014.
15. Gargiulo, F.; Fratini, A.; Sansone, M.; Sansone, C. Subject identification via ECG fiducial-based systems: Influence of the type of QT interval correction. *Comput. Methods Programs Biomed.* **2015**, *121*, 127–136. doi:10.1016/j.cmpb.2015.05.012.
16. Da Silva, H.P.; Fred, A.; Lourenço, A.; Jain, A.K. Finger ECG signal for user authentication: Usability and performance. In Proceedings of the Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS), Washington, DC, USA, 29 September–2 October 2013; pp. 1–8.
17. Iyengar, N.; Peng, C.; Morin, R.; Goldberger, A.L.; Lipsitz, L.A. Age-related alterations in the fractal scaling of cardiac interbeat interval dynamics. *Am. J. -Physiol.-Regul. Integr. Comp. Physiol.* **1996**, *271*, R1078–R1084.
18. Moody, G.B.; Mark, R.G. The impact of the MIT-BIH arrhythmia database. *IEEE Eng. Med. Biol. Mag.* **2001**, *20*, 45–50.
19. 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. *Comput. Methods Programs Biomed.* **2014**, *113*, 503–514. doi:10.1016/j.cmpb.2013.11.017.
20. Chamatidis, I.; Katsika, A.; Spathoulas, G. Using deep learning neural networks for ECG based authentication. In Proceedings of the International Carnahan Conference on Security Technology (ICCST), Madrid, Spain, 23–26 October 2017; pp. 1–6.
21. Wieclaw, L.; Khoma, Y.; Fałat, P.; Sabodashko, D.; Herasymenko, V. Biometrie identification from raw ECG signal using deep learning techniques. In Proceedings of the 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), Bucharest, Romania, 21–23 September 2017; Volume 1, pp. 129–133.
22. Labati, R.D.; Muñoz, E.; Piuri, V.; Sassi, R.; Scotti, F. Deep-ECG: Convolutional Neural Networks for ECG biometric recognition. *Pattern Recognit. Lett.* **2018**. doi:10.1016/j.patrec.2018.03.028.
23. Pourbabae, B.; Howe-Patterson, M.; Reiher, E.; Benard, F. Deep convolutional neural network for ECG-based human identification. In Proceedings of the CMBES Proceedings, Charlottetown, QC, Canada, 8–11 May 2018; Volume 41.
24. Zhang, Q.; Zhou, D.; Zeng, X. HeartID: A Multiresolution Convolutional Neural Network for ECG-Based Biometric Human Identification in Smart Health Applications. *IEEE Access.* **2017**, *5*, 11805–11816. doi:10.1109/ACCESS.2017.2707460.
25. Lynn, H.M.; Pan, S.B.; Kim, P. A Deep Bidirectional GRU Network Model for Biometric Electrocardiogram Classification Based on Recurrent Neural Networks. *IEEE Access.* **2019**, *7*, 145395–145405.

26. 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 Trans. Inf. Forensics Secur.* **2018**, *13*, 1258–1270. doi:10.1109/TIFS.2017.2784362.
27. 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* **2000**, *101*, e215–e220.
28. Tantawi, M.; Revett, K.; Salem, A.b. ECG based Biometric Recognition using Wavelets and RBF Neural Network 1 Introduction 2 Related Work. In Proceedings of the 7th European Computing Conference, Dubrovnik, Croatia, 25–27 June 2013; pp. 100–104.
29. Tantawi, M.M.; Revett, K.; Salem, A.; Tolba, M.F. Fiducial feature reduction analysis for electrocardiogram (ECG) based biometric recognition. *J. Intell. Inf. Syst.* **2013**, *40*, 17–39. doi:10.1007/s10844-012-0214-7.
30. Shen, J.; Bao, S.D.; Yang, L.C.; Li, Y. The PLR-DTW method for ECG based biometric identification. In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Boston, MA, USA, 30 August–3 September 2011; pp. 5248–5251. doi:10.1109/IEMBS.2011.6091298.
31. Rabhi, E.; Lachiri, Z. Biometric personal identification system using the ECG signal. In Proceedings of the Computing in Cardiology Conference (CinC), Zaragoza, Spain, 22–25 September 2013; pp. 507–510.
32. Sidek, K.A.; Mai, V.; Khalil, I. Data mining in mobile ECG based biometric identification. *J. Netw. Comput. Appl.* **2014**, *44*, 83–91. doi:10.1016/j.jnca.2014.04.008.
33. Ye, C.; Kumar, B.V.; Coimbra, M.T. Heartbeat classification using morphological and dynamic features of ECG signals. *IEEE Trans. Biomed. Eng.* **2012**, *59*, 2930–2941.
34. Belo, D.; Rodrigues, J.; Vaz, J.R.; Pezarat-Correia, P.; Gamboa, H. Biosignals learning and synthesis using deep neural networks. *Biomed. Eng. Online* **2017**, *16*, 115. doi:10.1186/s12938-017-0405-0.
35. Tieleman, T.; Hinton, G. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *Coursera Neural Netw. Mach. Learn.* **2012**, *4*, 15–31.
36. Bai, S.; Kolter, J.Z.; Koltun, V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv* **2018**, arXiv:1803.01271.
37. Bento, N.; Belo, D.; Gamboa, H. ECG Biometrics Using Spectrograms and Deep Neural Networks. *Int. J. Mach. Learn. Comput.* **2019**, *10*, 259–264.
38. Kingma, D.; Ba, J. Adam: A method for stochastic optimization. In Proceedings of the 3rd International Conference for Learning Representations, San Diego, CA, San Diego, USA, 7–9 May 2015.
39. Wang, Y.; Agrafioti, F.; Hatzinakos, D.; Plataniotis, K.N. Analysis of Human Electrocardiogram for Biometric Recognition. *EURASIP J. Adv. Signal Process.* **2008**, *2008*, 148658. doi:10.1155/2008/148658.
40. Fatemian, S.Z.; Hatzinakos, D. A new ECG feature extractor for biometric recognition. In Proceedings of the DSP 2009: 16th International Conference on Digital Signal Processing, Santorini, Greece, 5–7 July 2009; pp. 1–6. doi:10.1109/ICDSP.2009.5201143.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).