Nuno Filipe Abalada do Val Bento

Bachelor of Science in Biomedical Engineering

ECG Biometrics using Deep Neural Networks

Dissertation submitted in partial fulfillment
of the requirements for the degree of

Master of Science in
Biomedical Engineering

Adviser: Doctor Hugo Filipe Silveira Gamboa, Assistant Professor,
NOVA University of Lisbon

Examination Committee

Chairperson: Doctor Célia Maria Reis Henriques
Raporteur: Doctor Pedro Manuel Cardoso Vieira
Member: Doctor Hugo Filipe Silveira Gamboa

FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

March, 2019

ECG Biometrics using Deep Neural Networks

This document was created using the (pdf)LaTeX processor, based in the "novathesis" template[1], developed at the Dep. Informática of FCT-NOVA [2].
[1] https://github.com/joaomlourenco/novathesis [2] http://www.di.fct.unl.pt

Acknowledgements

Throughout my academic path, I counted on admirable people, who always inspired me to pursue my goals.

I am grateful to my adviser, Hugo Gamboa, for his belief in my abilities and for helping me to overcome my weaknesses.

To David Belo, who, as my direct collaborator and the initial proposer of this thesis, has turned the last 6 months into a constant search for new ideas, methods and solutions.

To the Libphys Biosignals group (Cátia, Daniel, David, Hugo, João, Nafiseh and Ricardo), for the friendly working environment, for all their support and all the interesting discussions.

To all my colleagues and professors, for being a part of my life during these academic years and for shaping the person I am now.

To my family, for all the support they give me every day, from since I was born, until these days.

Abstract

Biometrics is a rapidly growing field, with applications in personal identification and security. The Electrocardiogram (ECG) has the potential to be used as a physiological signature for biometric systems. However, current methods still lack in performance across different recording sessions.

In this thesis, it is shown that Deep Learning can be applied successfully in the analysis of physiological signals for biometric purposes. This is accomplished in two different experiments by formulating novel approaches based on Convolutional Neural Networks and Recurrent Neural Networks, which may receive heartbeats, signal segments or spectrograms as input. These methods are compared in tasks implying the recognition of subjects from four public databases: Fantasia, ECG-ID, MIT-BIH and CYBHi. This work obtained state-of-the-art results for across-session authentication tasks on the CYBHi dataset, reaching Equal Error Rates of 10.57% and 10.01% for the best model, with corresponding identification accuracy rates of 55.58% and 58.91%. It also demonstrates that using spectrograms as input to the classifier is a promising approach for biometric identification, achieving accuracy values of 99.79% and 96.88%, respectively for Fantasia and ECG-ID databases. Further, it is shown empirically that for ECG biometric systems, the ability of a model to generalize is crucial, not only its capacity to relate and store information.

These contributions represent another step towards real-world application of ECG-based biometric systems, closing the gap between intra and inter-session performance and providing some guidelines that can be applied in future work.

Keywords: Biometrics, Deep Learning, Signal Processing, Electrocardiogram

Resumo

A Biometria é uma área em rápido desenvolvimento, com aplicações em identificação pessoal e segurança. O Eletrocardiograma (ECG) tem potencial para ser utilizado como uma assinatura fisiológica em sistemas biométricos. No entanto, os métodos atuais ainda têm um baixo desempenho em tarefas que implicam várias sessões de aquisição.

Nesta tese é mostrado que a Aprendizagem Profunda pode ser aplicada com sucesso na análise de sinais fisiológicos para fins biométricos. Isto é conseguido em duas experiências distintas através da formulação de novas técnicas baseadas em Redes Neuronais Convolucionais e Redes Neuronais Recorrentes, as quais podem processar batimentos cardíacos, segmentos de sinais ou espetrogramas. Estes métodos são comparados em tarefas que implicam o reconhecimento de indivíduos em quatro bases de dados públicas: Fantasia, ECG-ID, MIT-BIH and CYBHi. São obtidos resultados competitivos com o estado da arte em tarefas inter-sessão na base de dados CYBHi, atingindo taxas de erro igual de 10.57% e 10.01% para o melhor modelo, com valores de exatidão de 55.58% e 58.91% para as tarefas de identificação correspondentes. Também é demonstrado que o uso de espetrogramas à entrada do classificador constitui uma técnica promissora para identificação biométrica, alcançando valores de exatidão de 99.79% e 96.88% para as bases de dados Fantasia e ECG-ID, respetivamente. Além disso, é mostrado empiricamente que para sistemas biométricos com ECG, a habilidade de um modelo para generalizar é crucial, não só a sua capacidade de relacionar e armazenar informação.

Estas contribuições representam mais um passo para a aplicação de sistemas biométricos baseados em ECG no mundo real, aproximando o desempenho intra e inter-sessão e dando algumas diretrizes que poderão ser aplicadas em trabalhos futuros.

Palavras-chave: Biometria, Aprendizagem Profunda, Processamento de sinais, Eletrocardiograma

Contents

List of Figures

List of Tables

Abbreviations

BPTT Backpropagation Through Time.

BS Biometric system.

CNN Convolutional Neural Network.

DNN Deep Neural Network.

ECG Electrocardiogram.

EEG Electroencephalogram.

EMG Electromyogram.

GRU Gated Recurrent Unit.

LDA Linear Discriminant Analysis.

LSTM Long Short-Term Memory.

NN Neural Network.

PCA Principal Component Analysis.

RNN Recurrent Neural Network.

SVM Support Vector Machine.

TCNN Temporal Convolutional Neural Network.

Chapter 1

Introduction

1.1 Context

In the last decade, with the increase in concerns about security, more trustworthy authentication technologies are in demand by the military, government and health sectors [1, 2].

Human biometrics is a rapidly developing field, which uses inherent physical or behavioral data to identify or authenticate people. It was largely popularized by fingerprint-based biometric systems, for example, the touch ID in smartphones [3].

Presently, digital health innovation is growing at a fast rate, with reports showing a total investment of $3.5bn during the first half of 2017 in the USA alone [3]. This goes along with a considerable increase in the number of consumers, generating large amounts of data, which can potentially be used in medicine [4], as well as to improve biometric systems.

As the ECG is a signal originated internally and unique to each person, it has the potential to turn into a reliable source for biometrics [3, 5]. For this to occur, pattern recognition algorithms are applied to ECG signals, given the existence of intrinsic features based on their characteristic waveforms and segments (PQRST). This morphology enables biometric identification by matching ECG templates to the subjects who generate these signals.

Although efforts are being made to develop more robust ECG-based Biometric systems (BSs), there are still some factors hindering its reliance, namely noise and artifacts, which happen due to imperfections during signal aquisition. Consequently, computing features may be arduous as the electrode material, movement, instrumentation of the devices and power-line may interfere, therefore increasing the variability of ECG signals from the same source [6, 7].

In the last few years, there was an increase in research activity on Deep Neural Networks (DNNs) for biometric applications. This includes many successful approaches to the recognition of ECG signals, suggesting its usefulness for this task [8–11]. This type of model has been extensively employed in speech recognition, machine translation, video games and robotics [12]. The increase of interest in DNNs is a consequence of advances in computational power, parallel computing, software infrastructure and the availability of large datasets.

In this thesis, with the objective of improving current BSs, there are two proposed experiments: "ECG Biometrics using Spectrograms"and "ECG Biometrics using Signal Segments and Fusion". The first experiment uses Convolutional Neural Networks (CNNs) and spectrograms to identify subjects from Fantasia [13] and ECG-ID [14] databases. The second experiment compares several DNN architectures for biometric identification and authentication. ECG data for this experiment are taken from Fantasia, MIT-BIH [15] and CYBHi [16] databases.

Artificial Intelligence techniques may learn patterns that a researcher could not perceive as easily. For example, the DNN could identify the sign of the fourth derivative as a good feature, whereas it would not be trivial for the human mind to reach the same conclusion. Therefore, these algorithms have the ability to learn more complex features, allowing more information to be extracted from the same data.

1.2 Objectives

This thesis makes the following contributions to the field of ECG biometrics:

- Frameworks for personal identification and authentication using ECG signals and deep neural networks;
- A comparison between different signal processing methods for ECG biometrics;
- The implementation of different DNN architecture topologies, namely convolutional and recurrent neural networks;
- Application of these architectures in differing conditions, such as on-the-person and off-the-person ECG Biometrics, recording duration, sampling frequency, recording hardware, signal to noise ratio and moments of acquisition.

1.3 Concepts

1.3.1 Electrocardiogram

The ECG (Figure 1.1) is a recording of an electrical signal which represents the activation of the heart muscles during their contraction and relaxation. It consists in the propagation of electrical activity through the membrane of cardiac cells, resulting in voltage

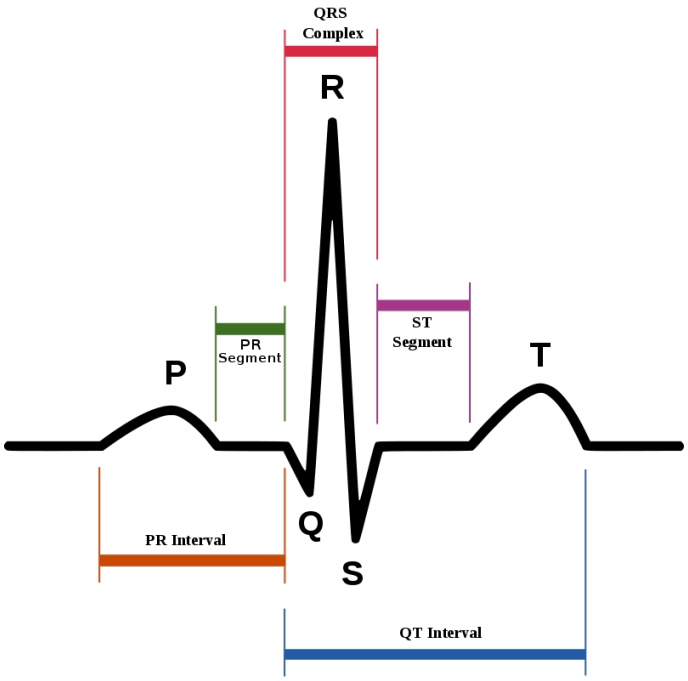


Figure 1.1: Electrocardiogram signal, waves and segments.

variation over time. These values may be measured by electrodes placed on the skin surface, linked to an electrocardiograph, the signal acquisition device. There are several different ways to place the electrodes, capturing different derivations. Each derivation allows to obtain voltage differences in a specific direction, leading to diverse, but highly correlated measurements.

The ECG is represented by waves and segments, each corresponding to a different stage of the cardiac cycle: (1) P wave: Contraction of the atria; (2) PQ segment: Time between the contraction of the atria and activation of the ventricles; (3) QRS complex: The combination of the Q, R and S waves, associated with the contraction of the ventricles; (4) ST segment: Stage in which the ventricles contract; (5) T wave: The repolarization of the ventricles [2].

This complex morphology is specific to each individual and, in addition to other qualities (described in Section 1.3.3.1), makes the ECG a potential data source for biometrics.

1.3.2 Signal Processing

Signal processing is an area that studies the characteristics and components of signals, as well as a wide range of methods that enable their analysis. In this thesis, it is used throughout all the experiments, with large importance in the preprocessing stage, allowing signals to become more interpretable to the classifiers. In this section, the applied signal processing methods are described in detail.

1.3.2.1 Convolution

Throughout the experiments, the convolution operation appears frequently. It allows a form of product between two different time series and is commonly used as a way to apply filters to signals or images.

The convolution operation consists in integrating a multiplication between the values of two functions f and g, one being successively translated by a constant τ. It can be defined as:

$$(f * g)(t) = \int_{-\infty}^{\infty} f(\tau)g(t-\tau)d\tau. \tag{1.1}$$

Computationally, as the signals are represented by samples (discrete values), the discrete version of convolution is utilized instead. The integral becomes a sum and the operation is given by:

$$(f * g)[n] = \sum_{m=-\infty}^{\infty} f[m]g[n-m]. \tag{1.2}$$

1.3.2.2 Fourier Transforms and Spectrograms

A Fourier Transform is a representation of a signal in the frequency domain. It acts by decomposing the input sequence into a sum of complex functions and is given by [17]:

$$X(\omega) = \int_{-\infty}^{\infty} x(t)e^{-j\omega t}dt, \quad \omega \in \mathbb{R}, \tag{1.3}$$

where ω is the angular frequency and x is the input signal.

Its discrete version, the Discrete-Time Fourier Transform (DTFT), is defined as:

$$X(e^{j\omega}) = \sum_{n\in\mathbb{Z}} x_n e^{-j\omega n}, \quad \omega \in \mathbb{R}, \tag{1.4}$$

where ω is the angular frequency and x is the input signal.

Depending on the employed functions, these transforms can have several types, such as Wavelet transforms, Z-transforms, Sine/Cosine transforms, among others. These functions produce coefficients with varying resolution in time and frequency domains. From just a few coefficients, the original signal can be reconstructed in great detail, making these algorithms popular for audio, image and video compression [18].

Spectrograms (Fig. 1.2) are visual representations of signals, expressing the occurrence of certain frequencies in a time window and their corresponding magnitudes. These magnitudes can be represented in a frequency/time plot as a third dimension, usually represented in false color. This kind of representation is generally achieved by applying Fourier Transforms to segments of the original signal, in order to obtain the existing frequencies and their corresponding magnitudes.

1.3.2.3 Filters

In signal processing, a filter is a function which has the purpose of removing unwanted characteristics of a signal and emphasizing the relevant ones. It may be used for data

transformation, acting as a tool to increase the signal-to-noise ratio or to detect different frequency components. It may also be used to obtain information from these inputs, for instance, detecting waveforms in a signal or shapes in an image. These filters can be of various types, depending on [17]: attenuated frequencies (low-pass, high-pass, band-pass, band-stop); linearity (linear or nonlinear); memorylessness, whether the output depends only on the input at the same instant; causality, being causal if the output depends only on the past input; shift invariance; and stability.

As an example, the moving average is a low-pass filter which calculates averages of successive signal segments. It can be defined as [17]:

$$y_n = \frac{1}{N} \sum_{k=-(N-1)/2}^{(N-1)/2} x_{n-k}, \quad n \in \mathbb{Z}, \tag{1.5}$$

where N is an odd positive integer and represents the filter size.

An ECG has a frequency band between 0.5 and 80 Hz, however, as it often suffers from added noise in the acquisition phase, it may be useful to process the signal with filters. These filters can be used to eliminate various types of noise, such as [19]:

- Baseline wandering, a low frequency noise which occurs due to small movements, respiration or offset voltages.

- Power-line noise (50/60 Hz), which affects the ECG on a higher frequency spectrum and throughout all the signal length (see Fig. 1.2).

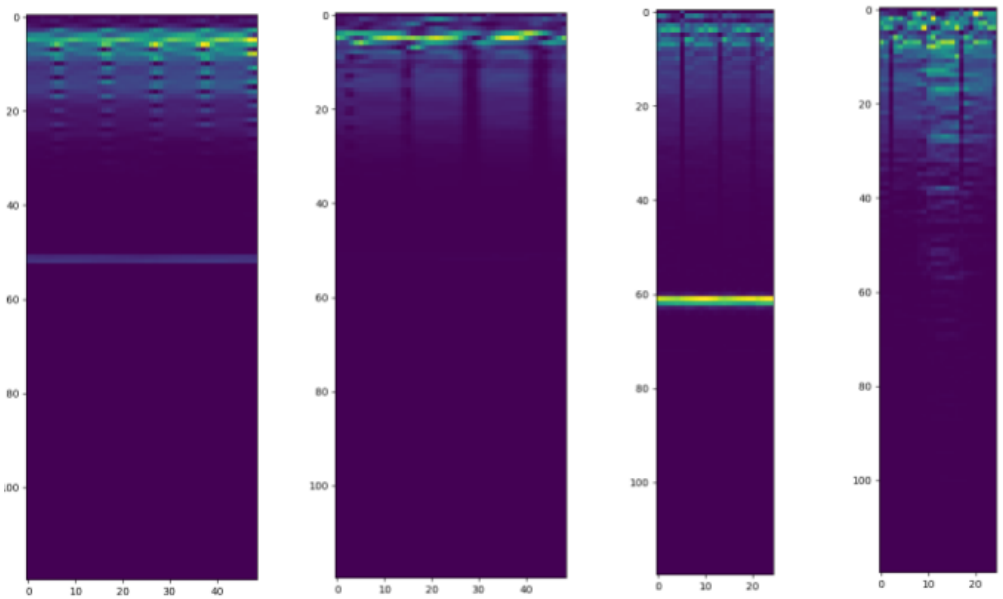


Figure 1.2: Two examples of spectrograms created from ECG-ID database (left) and two examples from Fantasia database (right). The frequencies seen around 50 and 60 Hz are due to power-line noise.

- Electromyogram (EMG) motion artifacts, which can be removed with low pass filters due to their high frequencies.

1.3.3 Biometrics

Biometrics are measurements of distinct physical, chemical or other characteristics with the objective of establishing a connection between subjects and their identity. They are present in our daily life either for personal identification (e.g. identification of a fingerprint in a crime scene) or authentication (e.g. passport verification at an airport).

The main advantage of these systems is the absence of pre-knowledge, of which passwords are the primary example. Unlike these traditional methods, most biometric attributes cannot be forgotten, transferred or stolen and, even though some could be reproduced, their forgery is difficult and requires the presence of the individual. Since the application of the elementary fingerprint recognition in 1883, efforts have been made to discover different and more reliable types of biometric data [20].

Biometric data can be divided into physical and behavioral, depending on their nature. Examples of these characteristics include [1]:

- Physical: fingerprints, face, iris, DNA, body odor or physiological signals (ECG, EMG, Electroencephalogram (EEG));
- Behavioral: handwriting, gait or human behavior in a web page (mouse and keystrokes) [20, 21];

Biometric technology has been used in a broad range of applications, reaching several fields, such as: forensic science, financial security, physical access checkpoints, information systems security, customs and immigration, national identification cards, driver licenses, among others [20].

1.3.3.1 Biometric Systems

A biometric system is a system which has the objective of performing a biometrics task. It may be seen as a computer program containing a pattern matching algorithm that allows the authentication of subjects from a database given one or more types of biometric data.

BSs have two temporal phases: the enrollment phase, when an input is fed into the algorithm and a template of a personal identity is generated; the authentication/identification phase, when feature extraction methods are applied and the model attributes a class to new inputs.

BSs can be fiducial or non-fiducial-based, depending on the nature of the features in use. As an example for ECG-based biometrics, the fiducial features are extracted between reference points in the cardiac cycle (e.g. temporal intervals). As for the non-fiducial features, the ECG is considered as a whole, considering characteristics as R-R intervals or relative heights between waves [2, 22].

There are several aspects that a BS should follow in order to be considered reliable. These include universality, uniqueness, permanence, measurability, performance, acceptability and circumvention. The ECG has the potential to be chosen as a source of information to a BS because it excels in some of these critical points, namely: universality, everyone possesses a heart; uniqueness, the physical and chemical structural differences of the heart provide differences in electrical conduction; and circumvention, as it is extremely difficult to counterfeit these signals [2].

Currently, the strongest issues hindering the mainstream use of ECG for commercial BSs are noise and variability over time [23]. Robustness to noise is important, due to the growing necessity of using cheaper and more practical acquisition devices. Handling variability over time in ECG signals is another non-trivial problem, since the features used to classify an ECG segment must be invariant over a wide temporal range, while still being able to distinguish all of the subjects. Modern datasets, including CYBHi, are made to test BSs for both these issues, but, even for state of the art systems, performance is still unsatisfactory [11].

1.3.3.2 Evaluation Metrics

The metrics used for the evaluation of BSs can be general classification metrics or performance measures. The former is used for identification, while the latter is used for authentication and is more common in biometrics.

Classification metrics are more common in the machine learning literature and in the medical field (e.g. diagnosis). The advantage of these metrics is the fact that they do not depend on threshold values. The following binary classification metrics are the most common:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}, \tag{1.6}$$

$$Sensitivity = \frac{TP}{TP + FN}, \tag{1.7}$$

$$Specificity = \frac{TN}{TN + FP}, \tag{1.8}$$

where TP, TN, FP and FN respectively stand for True Positives, True Negatives, False Positives and False Negatives.

When personal data is fed into a BS, scores are attributed based on the similarity between the input and an identity template. As this similarity is generally not perfect (100%), even if it is scoring a trait of the same person on the same day, a threshold is needed to determine if a particular score is within limits. Although this value is variable, it still allows a more robust evaluation. With the application of a threshold in biometric authentication systems, a given subject can be rejected if the provided input does not reach the lower limit for acceptance. On the other hand, if this limit is too high, the subject can be falsely rejected, i.e. have its access refused when it should be able to enter the system. After this selection is made, the following metrics can be calculated: False

Acceptance Rate (FAR), given by:

$$FAR = \frac{FP}{FP + TN}, \tag{1.9}$$

False Rejection Rate (FRR), given by:

$$FRR = \frac{FN}{FN + TP}, \tag{1.10}$$

where TP, TN, FP and FN respectively stand for True Positives, True Negatives, False Positives and False Negatives after being subject to the acceptance threshold.

The Receiver Operating Characteristic (ROC) curve is a plot of FRR against FAR and is determined after submitting the scores to different threshold values. Equal Error Rate (EER) corresponds to a value in the ROC curve at which FAR and FRR are equal.

As for this thesis, performance measures (thresholding) are considered the most important metric, due to a more direct indication of robustness in the system, which will only accept an input if its values satisfy a minimum degree of similarity with any template in the database. Nevertheless, as it can be noticed by analyzing their formulas, there is a high correlation between them and, in principle, if there are good identification results, the authentication results are most likely in accordance.

1.3.4 Neural Networks

A Neural Network (Fig. 1.3) is a machine learning algorithm inspired by the human brain. It is represented by a computational graph which consists in one or more layers of neurons. The synapses of a neuron, also known as weights, are represented by numerical matrices and are changed during the learning process, to minimize a cost function.

A Neural Network (NN) produces a result by propagating the inputs through each layer consecutively, while each neuron performs a weighted sum of its inputs. Optionally, the output of a neuron can pass through an activation function, a nonlinear function such as the hyperbolic tangent ($tanh$) or the rectified linear unit (ReLU), given by $max(x, 0)$, being x the input vector. This increases the ability of a NN to solve nonlinear problems.

Similarly to weights, biases are parameters that can be useful if there is a need for an offset value from the input to the output of a neuron. The term "parameters"is often used as a reference to the set of weights, biases and other variables present in a NN.

The first layer of a neural network is called a visible layer or an input layer. The subsequent layers are called hidden layers up to the last one, the output layer. A NN with two or more layers is called a Deep Neural Network (DNN).

1.3.4.1 Learning

A neural network learns by optimizing its weights based on an error signal. The algorithms employed in this phase can be diverse, but are generally nonlinear optimization algorithms.

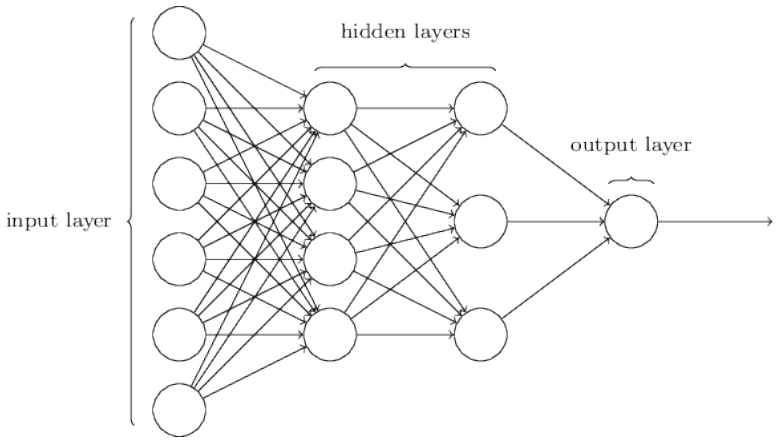


Figure 1.3: Neural Network [24].

For instance, in supervised learning, after the network produces results (forward propagation), a comparison with the ground truth labels takes place. This comparison originates an error signal, which is then propagated backwards, updating the weights of each neuron. This procedure is called backpropagation. A common technique used to implement it is gradient descent, which works by iteratively propagating the partial derivative of the error with respect to the parameters θ, until the cost function $J(\theta)$ reaches a minimum. It can be expressed by:

$$\theta = \theta - \eta \nabla_\theta J(\theta), \tag{1.11}$$

where η is the learning rate, a small constant which determines the magnitude of each update.

In supervised learning, the functions used to obtain the error information (cost functions) depend on the nature of the problem. The objective may be the attribution of a class to an input - classification; or the approximation of a continuous variable - regression. A popular metric used to calculate the classification cost function is the cross entropy loss (H) between the predicted labels ($\hat{y}_i$) and the ground truth labels (y_i). It is given by:

$$H(y, \hat{y}) = \sum_{i=1}^{C} y_i \log \frac{1}{\hat{y}_i} = -\sum_{i=1}^{C} y_i \log \hat{y}_i, \tag{1.12}$$

where C is the total number of classes.

As for a regression task, the standard cost function is the mean squared error, defined as:

$$MSE(y, \hat{y}) = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2, \tag{1.13}$$

where $\hat{y}_i$, y_i and N represent, respectively, the predicted values, the ground truth values and the total number of samples.

To facilitate the learning process, weights are generally initialized with small random values following a specific distribution (e.g. normal distribution).

Besides supervised learning, NNs are capable of unsupervised learning, in which case there is no ground truth information and the network only finds a representation to make sense of patterns in the data; semi-supervised learning, where a limited degree of ground truth is provided; self-supervised learning, where the network (e.g. an autoencoder) tries to learn a latent representation of its inputs, which can be useful for data compression or as a pretrained module; and reinforcement learning, where a positive reinforcement (reward) signal encourages the parameters to update in its direction and magnitude, while the opposite occurs for negative reinforcement signals. NNs are considered universal function approximators [25], as they are able to learn linear or nonlinear functions, when given inputs and a cost function. With that in mind, along with their known ability to generalize, this can be considered a very promising approach to create powerful applications for human biometrics.

1.3.4.2 Convolutional Neural Networks

A CNN (Fig. 1.4) is a type of neural network that works by applying convolution operations between kernels and an input tensor. This input tensor is an n-dimensional vector carrying the input information. It may be, for example, a 2D matrix representation of an image, a 3D sequence of images (video) or a sequence of voltage samples (1D), as it often is, in the case of ECG biometrics. This algorithm has recently been successful in the fields of machine translation and robotics, among others [12].

The kernels correspond to the weights in regular neural networks and can also be learned. They can be viewed as filters which have the capability of attending to the morphological characteristics of the input tensor. In the case of image recognition, these filters can detect shapes, edges and other patterns that may appear in an image. Any layer in a network with this ability is called a convolutional layer due to the fact that it is based on the convolution operation. Its output is a tensor of stacked n-dimensional matrices that indicate the neural activations originated by the input. These activation maps correspond to the features of the input tensor, being also called feature maps. After a convolution, a pooling layer may be used to reduce the dimensions of the feature maps. In most cases, it is taken the maximum value of small slices of a feature map, an operation called max pooling. This diminishes the need for computation and helps filters in the detection of shapes that suffer translations [12].

1.3.4.3 Recurrent Neural Networks

A Recurrent Neural Network (RNN) (Fig. 1.5) is a type of NN with recurrent connections and is used for sequential data processing. It generally consists of neurons connected to themselves with the objective of encoding memory into a state vector. A state vector is an embedding of the last n timesteps to the network and it is usually updated at each timestep. The representation of RNNs may also be in an unfolded form, resulting in a computation graph which can be easily implemented, where self connections turn into

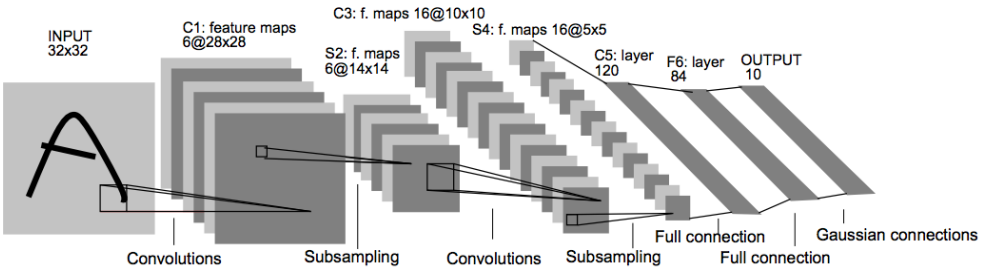


Figure 1.4: Convolutional Neural Network (LeNet architecture) [26].

sequential connections between neurons with shared parameters and the length becomes finite.

These models can operate in different modes, depending on the task at hand:

- One to many, which can generate a sequence based on only one input. An example application is audio generation.
- Many to many, which maps a sequence to another. Both sequences can have the same or different lengths. This mode is applied in machine translation.
- Many to one, in which a single value is obtain from a sequence. It is mostly applied in classification tasks, including in ECG biometrics.

The conventional algorithm for RNN training is called Backpropagation Through Time (BPTT). This algorithm is similar to regular backpropagation, yet the shared parameters of the network are learned with a recursive temporal propagation of the cost function's gradient. Usually, due to limitations in memory and computation, RNNs are trained by truncated BPTT. This method limits the backpropagation time window, leading to faster training, but also to restricted temporal dependencies [27].

During the propagation of gradients through many timesteps, the successive differentiation leads to a decrease in the magnitudes of the gradients, causing a loss of long-term information. This is known as the vanishing gradient problem, where the last layers, being closer to the labels, receive high quality gradients, whereas the learning process becomes harder for earlier layers. To counteract this issue, more complex recurrent cells were developed, such as the Long Short-Term Memory (LSTM) [28] and the Gated Recurrent Unit (GRU) [29]. These computational memory cells possess gates with the ability to learn to selectively attend to the past, choosing what to remember and what to forget.

The GRU is known to perform similarly to the LSTM in terms of recognition ability, while being considerably faster, due to the absence of a forget gate. The remaining gates, reset and update, compose the GRU. The general equations for these gates, respectively, are:

$$r = \sigma(W_r x_n + U_r h_{n-1}) \tag{1.14}$$

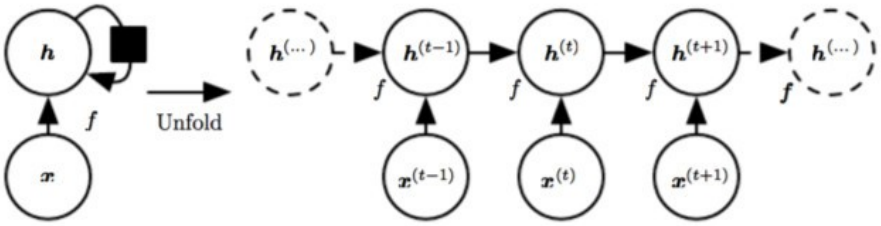


Figure 1.5: Recurrent Neural Network in folded (left) and unfolded (right) form (Adapted from [12]).

and

$$z = \sigma(W_z x_n + U_z h_{n-1}), \tag{1.15}$$

where σ is the logistic sigmoid function, h_{n-1} is the previous state and the weight matrices W_r, W_z, U_r and U_z are optimized through training.

As the data flows through the GRU, these gates change their values, consequently updating or forgetting the input data (by the reset gate) or computing the next state, which is nominated as candidate (by the update gate), ensuring that the essential information passes through. Therefore, the next state is a combination between the last state and a mapping of the input data. The candidate for the next state is computed by:

$$\widetilde{h} = \phi(W_h x_n + U_h(r_n \cdot h_{n-1})), \tag{1.16}$$

where $(\cdot)$ denotes element-wise multiplication and ϕ is an activation function, being *tanh* the most commonly used for this purpose.

The next state, i.e. the output of the current unit, is computed by:

$$h_n = z_n \cdot h_{n-1} + (1 - z_n)\widetilde{h}_n. \tag{1.17}$$

Overall, this ability to memorize is useful for signal analysis, as it can have a perspective of the global behavior of a signal, which is in itself a sequence of samples.

Chapter 2

State of the Art

This chapter includes a review of modern applications for ECG analysis, the most relevant work in ECG Biometrics, along with important research in deep learning. Finally, some of the works that join these areas and are most significant to the scope of this thesis are presented.

2.1 ECG

The latest research in the areas of ECG signal processing and classification has very diverse applications. Besides biometrics, research can be found in signal compression, blind source separation [30], denoising [31], artifact removal [6], medical diagnosis, signal synthesis, among others.

For ECG signal compression, standard techniques include discrete wavelet transforms [32] or block sparse Bayesian learning [33]. Deep learning has already been applied to this subject, as Yildirim et al. [34], using deep convolutional autoencoders, achieved a very high compression ratio (32.25), while maintaining good signal quality.

There are various disease-related events which can be detected by analyzing the ECG of a given patient. Park et al. [35] compared kernel density estimation, a NN and a Support Vector Machine (SVM) to detect myocardial ischemia using fiducial features. A more recent method used CNNs for a similar task [36]. Acharya et al. [37] were able to diagnose congestive heart failure using a one-dimensional CNN.

For arrhythmia detection, several works have been proposed, using very diverse classification methods. Polat and Güneş [38] used Principal Component Analysis (PCA) and an SVM to achieve 100% accuracy on a public arrhythmia dataset. Osowski et al. [39] employed an ensemble of neural networks on the MIT-BIH Arrhythmia database. More recently, Yildirim et al [40] applied a 1D CNN to the same task. Zihlmann et al. [41]

applied spectrograms and CNNs for a similar task, using a larger database.

Synthesizing ECG signals may be important to assess if a particular model understands the underlying characteristics of a signal. It can also improve the representation capability of that particular model, so that it can extract better features. Work in this area includes Belo et al. [8], which synthesized ECG, RESP and EMG signals using RNNs. Golany and Radinsky [42] used generative adversarial networks to synthesize ECG signals, augmenting the MIT-BIH dataset and improving the performance of an LSTM classifier.

It can be verified that, mainly due to their ability to effectively learn features from data, DNNs are becoming increasingly popular throughout many tasks that require the analysis of ECG signals.

2.2 ECG Biometrics

Since the field of ECG Biometrics was first introduced, a fair amount of progress has been made. Recently, this progress has been amplified by modern algorithms, including DNNs.

The typical biometric system with ECG comprises the following flow [43]:

1. The acquired data is submitted to a feature extraction algorithm;
2. The extracted features are fed to a classification module, which compares the inputs with ECG templates of a given subject and makes a decision;
3. The features and/or the classifier modules are stored in a database during the enrolment phase.

In fiducial-based ECG Biometrics research, the most common feature extraction methods are the extraction of characteristic points, such as intervals, amplitude, angles, areas, linear combinations of features, euclidean distances and slopes [5, 44]. Wavelet transforms [45], cosine transforms [46] or Fourier transforms [47] are also utilized for this purpose, as they can accurately capture frequencies and waveforms in the cardiac cycle. Fiducial features may be submitted to a dimensionality reduction technique, such as PCA [5], Linear Discriminant Analysis (LDA) [48]; or to dynamic time warping [49], which measures the similarity between sequences by aligning them through time. These methods may be also used in the classification module [50, 51], however, NNs [47] and decision trees [52] are also deployed for these tasks. Figure 2.1 contains an illustration of commonly used fiducial features.

For non-fiducial approaches, R-R intervals and PCA are commonly used for feature extraction [51]. Wavelet, cosine or Fourier transforms can extract this kind of features as well [44], as some of the generated coefficients can represent patterns that occur over time lengths superior to the duration of a cardiac cycle. Pinto et al. [53] compared various combinations of features and classifiers for ECG identification in signals acquired from a steering wheel. Overall, the best results were obtained with a combination of discrete

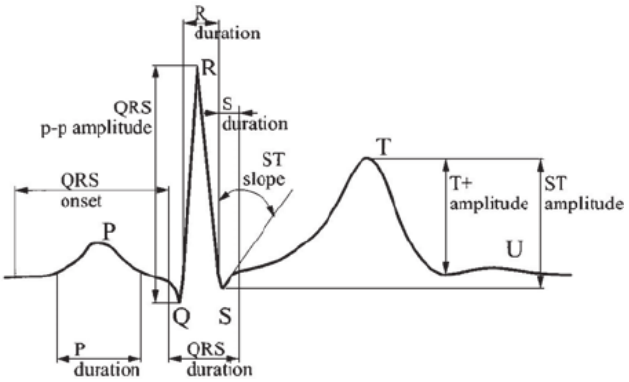


Figure 2.1: Examples of fiducial features. Reprinted from [5].

cosine transform and SVM. In their approach, clustering was used to remove artifacts that were present in the signals. Ferdinando et al. [54] used bivariate empirical mode decomposition, spectrogram analysis and a k-nearest neighbor classifier for biometric identification when emotions are considered. Besides achieving high classification rates, this work found a correlation between accuracy and two spectrogram parameters: overlap percentage and window size. Carvalho et al. [55] applied a similarity measure called Normalized Relative Compression to outperform all the previously tried methods on a locally collected dataset by a large margin, stating their intention to benchmark the proposed approach on CYBHi database, which was acquired off-the-person and contains two different recordings (see Section 3.4). Marques [56], using k-nearest neighbors and a similarity-based approach, achieved an EER of 3.97% for within-session tasks and 14.31% for across-session tasks on CYBHi database. In the latter kind of task, training and testing is performed on two separate recording sessions. This creates a more usable biometric system, as these systems must function through long periods of time.

Currently, DNNs have been employed in ECG biometrics to improve accuracy rates, since automatically extracted features are generally more effective, since they are explicitly optimized for a particular task. Page [57] used NNs for both QRS detection and classification, resulting in an identification accuracy of 99.96% for the ECG-ID dataset. It demonstrated the efficacy of DNNs in ECG signal recognition.

Eduardo et al. [9] implemented a deep autoencoder to learn lower dimensional feature representations, obtaining low identification errors on a private dataset. Zheng et al. [58] also used an autoencoder to achieve 98.1% accuracy using a self-collected dataset. In both these works, autoencoders are used for self-supervised feature representation, with the purpose of increasing the recognition rate capability of the classifiers.

For a more in-depth review of recent work in ECG Biometrics, as well as future research directions, see [23].

2.3 Deep Learning

Deep Learning is a subset of Machine Learning that has recently been popularized as a research field. It consists in the usage of DNNs on large datasets and can be applied to diverse areas of study, while requiring little domain knowledge in some cases, as the feature extraction step is done automatically.

In order to make sense of the recent breakthroughs in DNNs, here is presented some of the most relevant research that is applicable to the analysis of ECG signals. It covers two of the most effective NN types: CNNs and RNNs.

2.3.1 Convolutional Neural Networks

A Convolutional Neural Network is a type of DNN which uses the convolution operation as a way to extract complex hierarchical features. Traditionally, CNNs were created to handle 2-dimensional data, being used to recognize images [12].

One of the models with the highest recent impact in this area are Residual Networks (ResNets) [59], which have allowed the increase of CNN layers up to the thousands, by stacking feature maps formed in earlier layers to feature maps at higher levels - skip connections. This acts a way to counteract the vanishing gradient problem (see Section 1.3.4.3). DenseNet [60], another recently proposed CNN architecture, includes extra connections in groups of successive layers (dense blocks), potentiating the flow of information through layers and outperforming ResNets while using less parameters. These advances enabled CNNs to reach human-level performance on image recognition tasks.

This trend in image recognition can be helpful for ECG biometrics, especially when the signals are represented by spectrograms. Besides enhancing the signal representation by increasing its resolution in the frequency domain, spectrograms form 2D matrices, which can serve directly as input to a 2D CNNs. This method has previously been applied in sound recognition [61].

Although being mostly used for tasks related to computer vision or natural language processing, CNNs can also capture sequential patterns, having inclusively been applied to ECG signals in their 1-dimensional form [10, 11, 62–64]. Despite accomplishing positive results, recent work in 1D CNNs for processing time series indicates there can still be a margin for improvement.

One of these examples is Van Den Oord et al. [65], which used dilated causal 1D convolutional layers, obtaining near-human naturalness scores in audio generation. These layers used stacked convolutions that attend to the past with increasing spacing between samples (dilation), leading to a large receptive field. Bai et al. [66] compared a similar approach to RNNs and verify its superior potential for sequence modeling, as well as a much faster rate of convergence.

2.3.2 Recurrent Neural Networks

The main purpose of RNNs is to associate information gathered through a time period. In this type of NN, self-connections between neurons allow a long-term temporal integration of samples or features. In recent years, some improvements have been made to the standard RNN architecture besides LSTM and GRU cells.

Graves et al. [67] proposed a complex network which learns how to read and write to an external memory. This made it able to memorize complex dependencies and learn question answering tasks from scratch. Ba et al. [68] surpassed LSTM on associative memory tasks by a large margin. They used fast weights to update the network with information about the recent past. Jaderberg et al. [69] integrated a regression problem into truncated backpropagation through time, by trying to predict the gradients after a certain number of timesteps. This approach was able to improve the convergence rate of their recurrent model. Vaswani et al. [70] used a self-attention network to reach state of the art results in machine translation, an area previously dominated by RNNs. Wu et al. [71] recently proposed a simplified version of self-attention that includes convolutions and yields even better performance.

RNNs and CNNs may be joined into a single network. This method has been applied successfully in video recognition, since the abilities of vision and temporal association are present in the same model [72].

2.4 ECG Biometrics using DNNs

This thesis focuses on applying deep learning to ECG signal analysis, a very recent approach that has already produced interesting results.

Salloum & Kuo [73] proposed an aggregation of RNN architectures for analysis and classification after selecting QRS segments without further feature extraction methods. For the identification problem, it was reported nearly 100% classification accuracy on the ECG-ID database. Values of EER between 3.5% and 0% in an authentication task using the MIT-BIH database. This proved the efficacy of RNNs for this kind of problem. Zhang et al. [63] used a multiresolution one-dimensional CNN, where the feature extraction step comprises discrete wavelet transform, autocorrelation and component selection, to obtain 93.5% identification rate as the average result on 8 different datasets. Labati et al. [64] also applied a 1D CNN to identify patients using the IDEAL database as well as the PTB diagnostic ECG database, achieving 100% accuracy on the latter. Zhao et al. [74] obtained promising results in a recently published ECG dataset [4] by feeding a 2D CNN with ECG trajectories obtained by applying generalized S transforms to the signals. Wu et al. [75] reached state of the art performance on MIT-BIH Arrhythmia and ECG-ID by applying a two-stage approach, consisting of a 1D CNN, followed by an attention-based bidirectional LSTM.

The mostly related paper to the work presented in this thesis is Luz et al. [11]. It used

an architecture that fuses a 1D CNN, with raw ECG heartbeats as input, with another 2D CNN, fed with the corresponding spectrograms, demonstrating the potential of this type of classifiers on off-the-person ECG biometrics, namely on CYBHi and UofTDB databases. This work achieved EERs of 12.78% and 13.93% in across-session tasks. It also demonstrates that spectrograms can be a complement to ECG segments when using DNNs.

Chapter 3

Datasets

The data used for the first experiment is from two well-known public datasets from the Physionet database [76]: Fantasia and ECG-ID. The second experiment, besides also being tested on the Fantasia database, uses two other datasets: MIT-BIH and the CYBHi.

3.1 Fantasia

The Fantasia database consists on 40 subjects (20 young and 20 elderly) whose signals were recorded during 120min with a sampling frequency of 250 Hz, while watching the movie Fantasia from Disney [13]. The main advantages of this database are its popularity and the long duration of its recordings.

3.2 ECG-ID

ECG-ID consists of 310 recordings of 90 subjects aged from 13 to 75. While the original data ranges from 2 to 20 recordings per subject, only the first two were used, with the purpose of having the same number of samples for each subject and ensuring a balanced dataset. Each recording is 20 seconds long with a sampling frequency of 500 Hz [14].

3.3 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 [15]. In this experiment,

although the full database is used, resampling to 250Hz is performed, so that there is a more direct comparison with the results obtained on Fantasia database.

3.4 CYBHi

ECG signals are also from the Check Your Biosignals Here initiative (CYBHi) database [16]. For this dataset, signals were acquired through the subjects' hand palms and fingers, while they were sitting in a resting position and watching videos. Signals are 2 minutes long and the sampling frequency is 1000 Hz. In this experiment, only the long-term portion of the signals is taken into account. This subset contains data from 63 subjects aged between 18 and 24, from which 14 are male and 49 are female. It consists of two sessions, recorded approximately 3 months apart.

This dataset is regarded as a good way to test for the robustness of the algorithms in use. The reasons for this include: a relatively high number of subjects; the fact that there are recordings took at two moments distant in time; its low signal-to-noise ratio. This database was considered to be one of the best for biometric studies [77].

Chapter 4

Signal Processing Methods

In this chapter, the signal processing methods applied in the preprocessing step are explained in detail for both experiments.

4.1 Overview

Here is presented an overview of the preprocessing pipeline for each experiment.

4.1.1 ECG Biometrics using Spectrograms

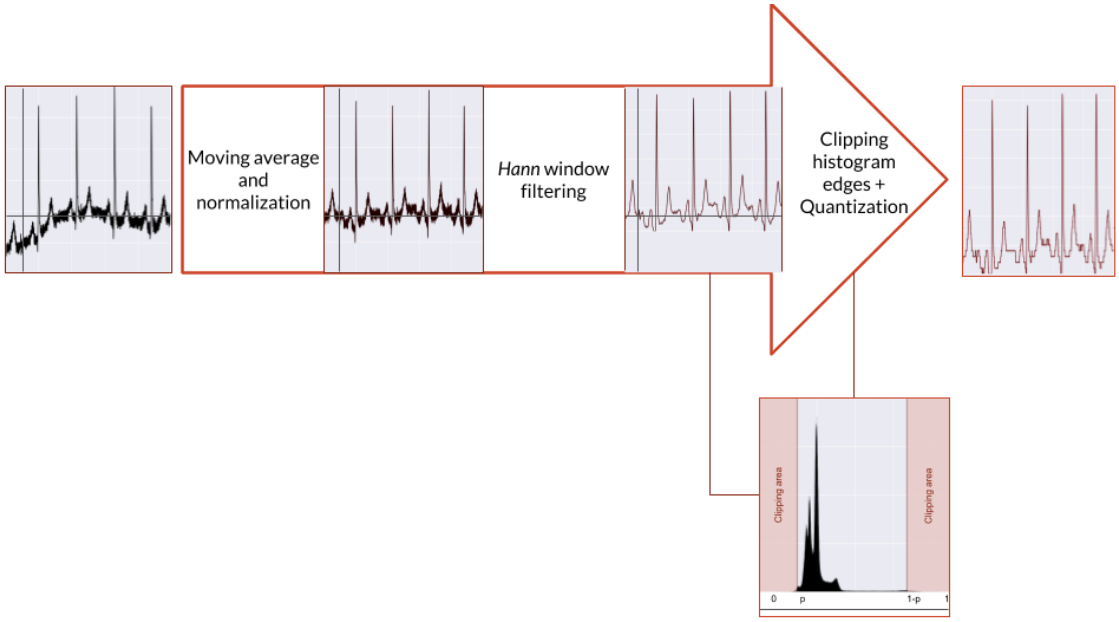
The full preprocessing step for this experiment is composed by:

- Normalization
- Filtering
- Segmentation and Spectrogram Generation

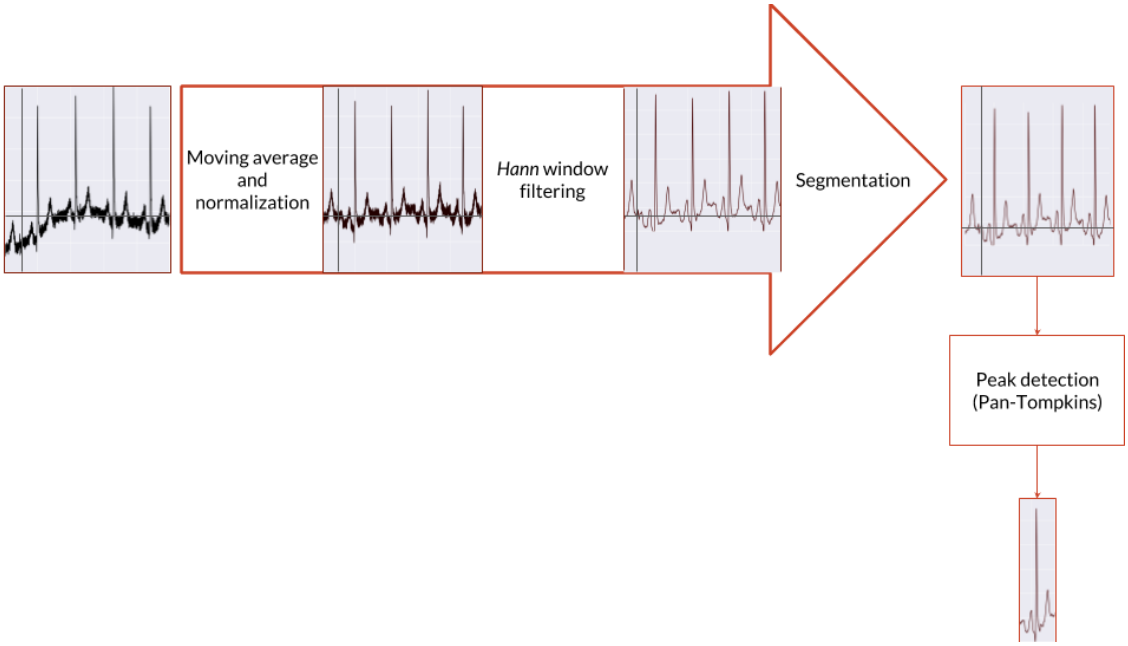
4.1.2 ECG Biometrics using Signal Segments and Fusion

The full preprocessing step for this experiment is composed by (Fig. 4.1):

- Normalization
- Filtering
- Quantization (Personalized RNN only)
- Segmentation + Peak Detection (CNN only)
- Segment Elimination



(a) Preprocessing for the personalized model.



(b) Preprocessing for the non-personalized models.

Figure 4.1: Preprocessing flow for Fantasia and MIT-BIH. For CYBHi, filters are replaced by a single Butterworth bandpass filter.

4.2 Normalization

Normalization works by rescaling the data. It is applied twice for each of the performed experiments, during different phases and using different formulas.

Usually, the main reason why normalization is applied in machine learning is that it makes the classifier perform better. In the particular case of this work, the training process becomes easier for neural networks. This happens because most gradient-based optimization algorithms have faster convergence when normalization is applied [78].

Normalization is the first preprocessing step for both experiments, being directly applied to the raw signals, so that the filters have a more stable behavior.

For the first experiment, each signal starts by being normalized by subtracting its mean and dividing the result by the difference between its maximum and its minimum:

$$x' = \frac{x - \overline{x}}{\max(x) - \min(x)}, \tag{4.1}$$

where x denotes the signal. After spectrograms are generated, they are normalized by subtracting the mean and dividing by the standard deviation. After this step, ECG data is ready to be used as input to the CNN.

For the second experiment, normalization is performed before filtering and after peak detection occurs. Before filtering, the following formula is used:

$$x' = \frac{x - \overline{x}}{\max(|x - \overline{x}|)}, \tag{4.2}$$

while after the signal is segmented, the normalization formula in use is:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}, \tag{4.3}$$

known as min-max normalization. This second normalization stage is crucial for the generalization ability of the algorithm. It is demonstrated empirically in Section 6.

4.3 Filtering

For the first experiment, a Hann window filter is applied, so that higher frequencies, which are associated with noise, are attenuated. The Hann window is given by:

$$w(n) = \sin^2\left(\frac{\pi x}{N-1}\right), \tag{4.4}$$

where N denotes the window size. Afterwards, baseline wandering is removed by applying a sliding window which subtracts to the signal its moving average.

For the second experiment, for the CYBHi database, similarly to [11], a Butterworth bandpass filter with cutoff frequencies of 0.5Hz and 40Hz is applied to the raw ECG data, with the purpose of removing undesired frequencies and smoothen the signal. This type

of filter is popular in signal processing due to having almost no ripple, i.e. oscillations in its pass band. It has been used for ECG denoising and, when compared with other filters, has a reported accuracy close to wavelet-based filters at a fraction of the computational cost [79].

For MIT-BIH and Fantasia, a moving average filter is applied, followed by a Hann window filter (Fig. 4.1).

4.4 Quantization

For the personalized model, a large RNN, quantization is used to simplify the signal, in order to make the network converge faster.

First, signals are clipped at their peaks by the use of an amplitude histogram with a confidence threshold. this allows for better resolution outside the QRS complex. The resulting signal, x, undergoes the same quantization process presented in [8]. This process consists of discretizing its range of values, transforming it into a vector with S_D=256 possible integer values. The following equation is used:

$$x_n = \text{round}\left(\frac{s_n - \min(s)}{\max(s)}(S_D - 1)\right) \tag{4.5}$$

where x_n is the n-th sample of the input vector.

4.5 Segmentation

During the segmentation procedure, filtered signals are separated into chunks with a determined size and overlap percentage. For the first experiment, each segment that is taken from the signal becomes a spectrogram. As for the second experiment, signal segments are directly processed by the classifier.

4.5.1 Spectrogram Generation

Since databases used in the first experiment have a significant difference in recording duration (20 seconds vs 120 minutes), the need for a minimum amount of training and validation data requires a change in the parameters used for spectrogram generation in both databases.

For the Fantasia database, signals are segmented into chunks with a length of 1536 samples (approximate duration of 6 seconds) and the fast Fourier transform window has a size of 256 samples with an overlap of 87.5%.

For the ECG-ID database, segments consist in 2048 samples (approximately 4 seconds) and a fast Fourier transform window size of 512, with an overlap of 93.75%. The shorter duration of the segments is due to the need of increasing the number of generated windows for training and validation.

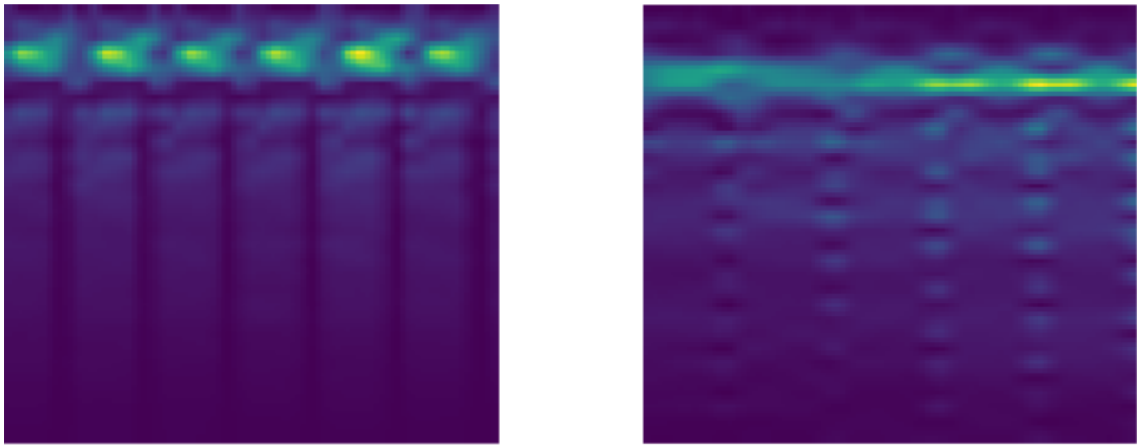


Figure 4.2: Examples of resized spectrograms for each of the used databases: Fantasia (left) and ECG-ID (right).

Once the spectrograms are generated, frequencies above 120 Hz are removed. The resulting matrix is then resized to 80×80 and normalized by subtracting the mean and dividing by the standard deviation. Examples of these processed spectrograms are shown in Figure 4.2.

The reason behind the use of high overlap percentages for spectrogram generation is that it provides more data for the same length of a signal. In addition to that, DNNs generally have better performance in larger datasets [12]. Also, it has previously been demonstrated that overlap percentage increases the performance of other classifiers, such as k-nearest neighbors with statistical features extracted from spectrograms [54]. During the validation procedure, several values of overlap were tested and this correlation was verified.

4.5.2 Peak Detection

For the second experiment, segments containing at least two heartbeats are cut from the filtered signals. From these segments, QRS peaks are detected using the Pan-Tompkins algorithm [80], which encompasses the following steps:

1. Apply a bandpass filter (e.g. Butterworth) to the signal;
2. Differentiate the signal, to obtain the slopes;
3. Square the signal, so that it becomes positive;
4. Apply moving-window integration, to eliminate noisy peaks and emphasize the QRS complex;
5. Capture the QRS peaks by selecting thresholds.

This algorithm is used to isolate single heartbeats, which are obtained by selecting the same number of samples before and after the QRS peak. This is done so that two different

kinds of input can be compared: segments and heartbeats. The possibility of joining both these two methods is then assessed.

For the CYBHi database, with sampling a frequency of 1000Hz, segments are 2048 samples long (approximately 2 seconds) and are cut with an overlap of 89%. Heartbeats contain 400 samples to each side of the detected QRS peaks, adding to 800 samples.

For both MIT-BIH and Fantasia, with the sampling frequency being 250Hz, samples are still approximately 2 seconds long, but only contain 512 samples and are cut using an overlap of 66%. Heartbeats have a length of 220 samples. For Fantasia, the total number of generated segments was 145435, while for MIT-BIH, this number was 171002.

The use of heartbeats as input is only tried on CNN-based models.

4.6 Segment Elimination

Segment elimination is tested on the CYBHi database as a way to slightly improve results, although it can be used to reject data in any ECG biometrics task.

As the data were obtained through ECG measurements in suboptimal settings (e.g. hand palms and fingers on CYBHi), some segments may contain substantial amounts of noise and motion artifacts when compared to others. These unwanted events generally harm the performance of the model in use and their explicit elimination may be needed. This elimination can be simply done by thresholding in terms of a given metric, such as similarity [11]. A more robust approach can in principle be achieved by clustering algorithms, which are unsupervised and can use features extracted from the data [6], as well as cross-correlation [53].

Following the possibility that clustering can robustify the system, a k-means clustering algorithm is proposed for tasks involving the CYBHi database and compared with thresholding. This unsupervised learning technique is performed before segmentation and used to separate the signals in k=2 clusters - motion artifact or normal ECG. The only feature in use is a moving standard deviation with a window size of 1536. This method may allow the generation of cleaner segments, avoiding an explicit rejection of data points for classification. The k-means clustering algorithm can be described by:

1. Initialize k initial cluster centroids as points in an N dimensional space, where N is the number of features;

2. Create clusters by assigning the datapoints to the closest centroid, with respect to a distance metric (e.g. Euclidean);

3. Compute k new centroids, which are given by the average position of all the datapoints in each cluster;

4. Return to 2. until convergence.

Figure 4.3: Detected ECG artifacts (yellow) using k-means clustering.

Table 4.1: Total generated segments and rejected segments for each session of CYBHi.

Session	Total	Rejected (% of total)
1	35115	97 (0.28)
2	33217	118 (0.36)

This clustering method removed 10±30% of the total signal length for every subject of the CYBHi database. This high variance is due to the fact that some signals have no apparent artifacts, while others may have many. Figure 4.3 shows an example of an ECG with detected artifacts.

By performing segment elimination by thresholding, segments outside adjusted limits of standard deviation are rejected. After validation, it was verified that the optimal percentages of segments to reject were very low. Nonetheless, this procedure was still employed (Table 4.1).

Chapter 5

Architectures

The choice of a model for classification is a key part of a biometric system. In this chapter, models used for the tasks of identification and authentication are explained in detail. The presented architectures are optimized by cross-validation on small subsets of the training data. Training is done by backpropagation, using the Adam optimizer [81] with $softmax$ activation and cross-entropy loss.

5.1 ECG Biometrics using Spectrograms

In this experiment, CNNs are fed with spectrograms as input to identify subjects from two different databases, Fantasia and ECG-ID. With that in mind, the differences between applying a small CNN (baseline) and DenseNet [60] are assessed. The same architectures are maintained for each database.

DenseNet is a state of the art CNN architecture which has a very large capacity. It can be seen as an evolution of ResNet [59], in the sense that it maintains its residual connections, with a simple modification. In this type of architecture, the inputs of a layer contained in a block are a concatenation of all the previous feature maps (Fig. 5.1). Besides allowing a more efficient use of the network parameters, this type of architecture achieved state of the art results on popular image recognition datasets. This version of DenseNet has 19 layers, including 3 dense blocks with a growth rate of 10, starting with 20 kernels on the first convolutional layer.

A more conventional CNN (Fig. 5.2) is also implemented. This simpler model consists of 4 convolutional layers with rectified linear unit (ReLU) activations. It is used as a baseline to verify how much of an improvement is given by having a larger architecture. This model also has the objective of optimizing parameters such as spectrogram dimensions and overlap percentage.

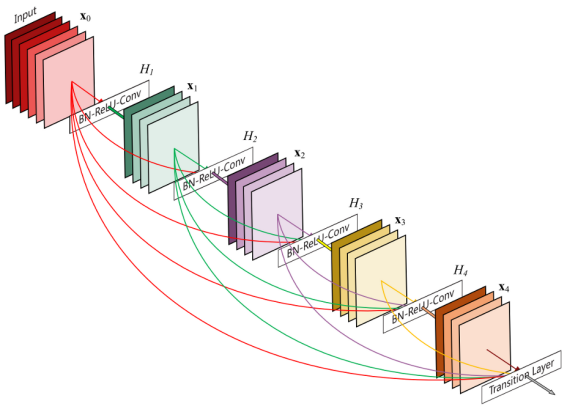


Figure 5.1: "A 5-layer dense block with a growth rate of k = 4", reprinted from [60].

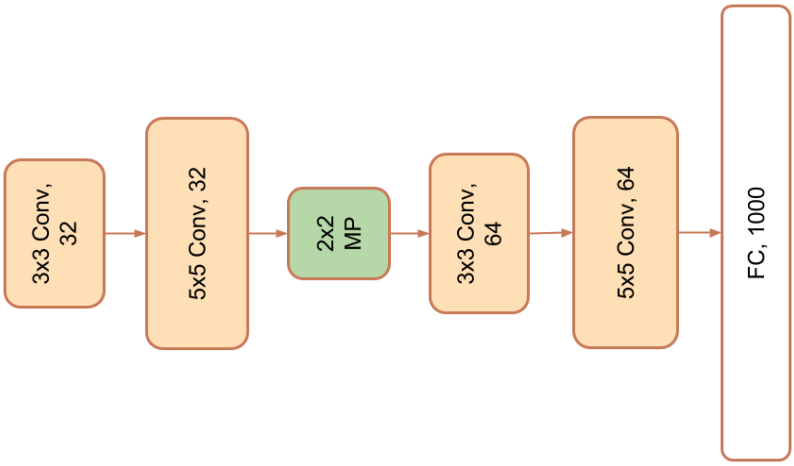


Figure 5.2: Conventional CNN architecture used as a baseline. "Conv", "MP"and "FC"respectively stand for convolutional, max pooling and fully connected layers.

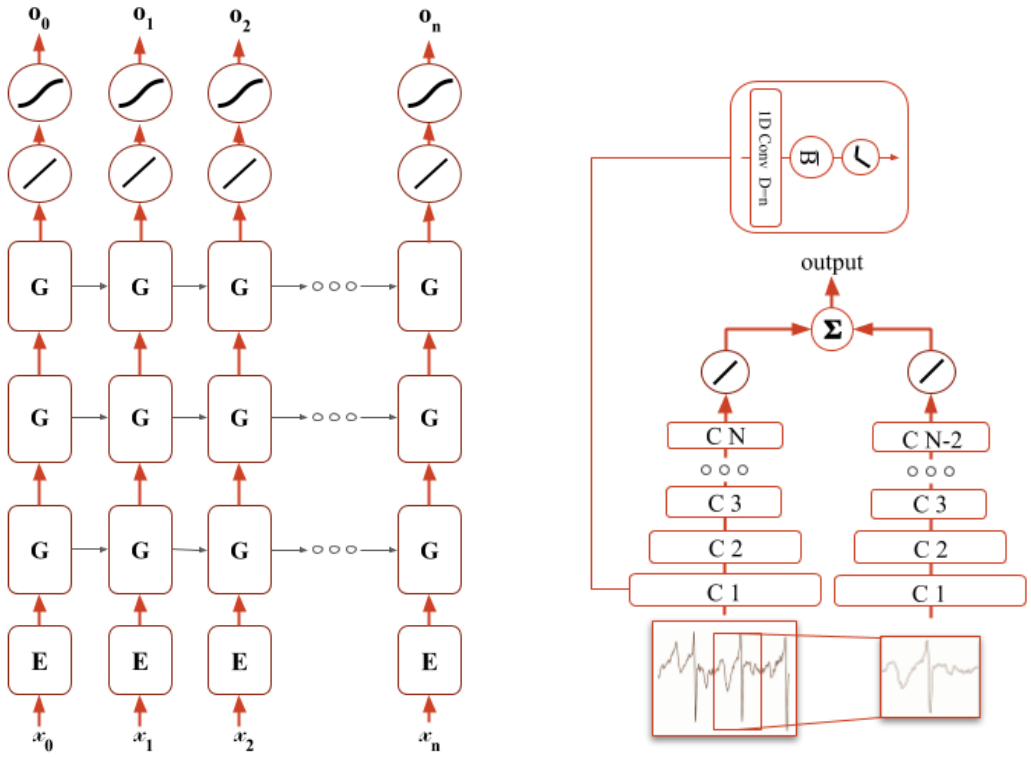
As DenseNet as a much higher number of weights, it is expected to outperform the baseline by a significant margin.

5.2 ECG Biometrics using Signal Segments and Fusion

In this second experiment, different DNN architectures are evaluated for the classification of ECG signals, directly using signal samples after being filtered and segmented.

5.2.1 Recurrent Neural Networks

The architecture used for the experiments involving personalized models, represented in Figure 5.3a, is a RNN based on GRUs (see Section 1.3.4.3). It is comprised by an embedding matrix E, three sequential GRUs G and a fully connected layer with a softmax activation. This kind of activation function, generally used for multi-class problems,



(a) Recurrent Neural Network Architecture

(b) Temporal CNN Architecture

Figure 5.3: Proposed architectures. "G"stands for GRU, "E"for embedding matrix, "C N"for n-th CNN layer and "B"for batch normalization

turns the outputs of the last layer into normalized probabilities. It is given by:

$$\text{softmax}(x_j) = \frac{\exp(x_j)}{\sum_i \exp(x_i)}. \tag{5.1}$$

The embedding matrix is popular in RNNs used for text processing. It works as a translation mechanism between the input and the first GRU layer. This matrix starts with random values and is optimized during the course of training. Each of the following GRU layers are fed by the output (temporal state) of the previous one. Each GRU layer has gates of size 256.

A non-personalized version of the RNN is also implemented and tested on CYBHi. This version is used to compare the performance of CNNs and RNNs more directly, as it uses the same preprocessing method (no quantization) and is much smaller (only 1 GRU layer with 32 neurons and no embedding matrix). The final layer has 128 neurons.

5.2.2 Convolutional Neural Networks

The second proposed architecture (Fig. 5.3b) for this experiment is a two-stream Temporal Convolutional Neural Network [66], which uses 1-dimensional convolutional layers to learn temporal patterns. It combines predictions from two different inputs: ECG window segments and extracted heartbeats.

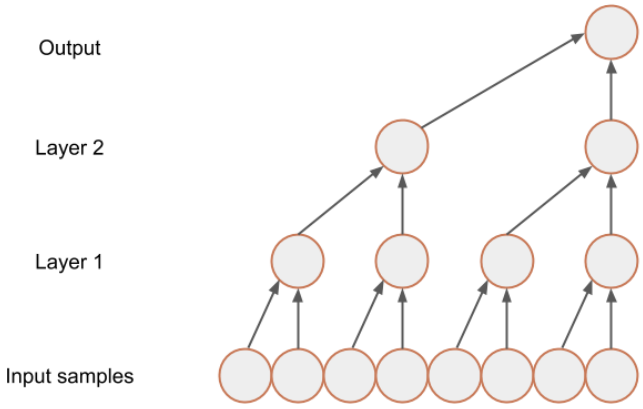


Figure 5.4: Temporal Convolutional Neural Network with three convolutional layers (L), kernel size of 2 and dilation rate of 2^{L-1}.

A Temporal Convolutional Neural Network (TCNN) (Fig. 5.4) is a 1-dimensional CNN with dilated convolutions. This type of network performs the convolution operation between an input signal and a moving vector (1D) with a fixed size, the kernel. These kernels are learned over the course of training. Each kernel will be activated by a particular type of waveform, enabling the NN to distinguish between different the sequences of waveforms that compose the signals.

To perform dilated convolutions, each kernel, while keeping its size constant, steps over several input samples. This provides a larger receptive field to the kernel, while reducing its resolution. The size of each step is controlled by a dilation rate which, in this case, is a power of 2. Successive layers of increasingly dilated convolutions become an effective way to obtain hierarchical relationships between samples, while maintaining computational efficiency [65, 66].

A high-level scheme of this architecture is presented in Figure 5.3b. Each neural network is composed by several convolutional layers with 24 kernels of size 4, followed by batch normalization and a ReLU (Rectified Linear Unit) activation function, given by $\text{ReLU}(x) = \max(0, x)$. Then, a fully connected layer of size 256 is applied to the output of the final convolutional layer. As there is a difference in length between both data streams (segments and heartbeats), a different number of layers is used for each network. This must happen due to the exponential factor on the dilation rate, making each additional layer drastically increase the total receptive field.

With the purpose of having the same recording time per segment, the significant difference in sampling frequency (250 to 1000 Hz) between datasets causes segments to differ in length, consequently having to change the architecture. This change is done simply by varying the number of layers. For Fantasia and MIT-BIH, the number of convolutional layers is 6 for the network with segments as input and 4 for the heartbeat network. These values are increased by 2 (8 and 6) for the CYBHi database.

Each network is trained independently and the sum rule is used to combine the raw

predictions (logits) produced by the output layers.

As an attempt to further improve generalization across different sessions on CYBHi, an autoencoder is used as an auxiliary task, jointly trained with the model. This technique is inspired by Le et al. [82], which showed significant improvements in performance for several datasets. In this variant, the base TCNN, excluding the output layer, is used as the encoder. For the decoder, a simple hidden layer of size 512 maps the final TCNN layer to a reconstruction $\hat{x}$ of the original input x with size D. Everything else is maintained, with either segments or heartbeats used as input to each of the two NNs which have their predictions combined after training. Mean squared error was used as the loss function for the autoencoder and no weighting coefficients are multiplied by either parcel of the loss function, given by:

$$L(y, \hat{y}, x, \hat{x}) = -\sum_{i=1}^{C} y_i \log \hat{y}_i + \frac{1}{D} \sum_{j=1}^{D} (x_j - \hat{x}_j)^2, \tag{5.2}$$

where $\hat{y}_i$, y_i and C represent, respectively, the predicted labels, the ground truth labels and the number of classes.

Chapter 6

Experimental Results

This section presents the results obtained for each experiment. All training and testing was performed on two NVIDIA GeForce GTX 1080 Ti graphics cards using TensorFlow [83], a numerical computation library that can run on GPU.

6.1 ECG Biometrics using Spectrograms

This experiment takes advantage of developments in machine learning to conceive a biometric identification system based on ECG spectrograms and CNNs.

The proposed task is to classify ECG spectrograms generated from signal segments into as many classes as subjects in the system. While doing so, two CNN architectures are tested: the baseline CNN and a DenseNet. The approach followed by this experiment is displayed in Figure 6.1.

As the subset of ECG-ID in use contains two sessions per subject, two different tasks are proposed: within-session and across-session. In within-session (or intra-session) tasks, both sessions are used for training and testing, being the split by time. In this modality, approximately the last 7 seconds of each session are used for testing. The across-session

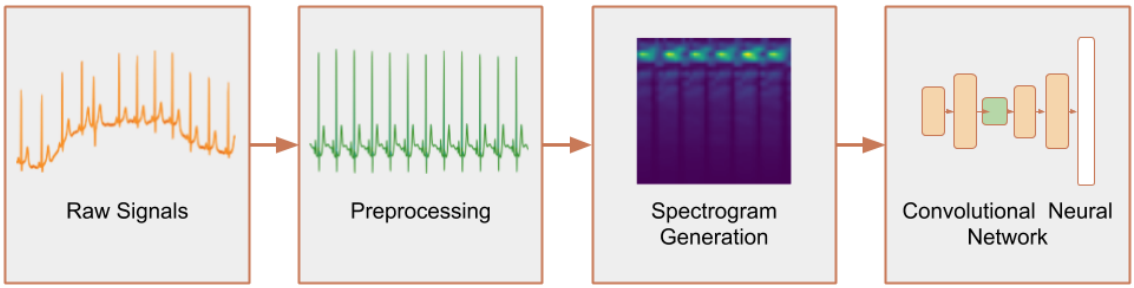


Figure 6.1: Flowchart of the spectrogram-based subject identification method.

Table 6.1: Within-session classification performance (%).

Model	Database	Accuracy	Sensitivity	Specificity
Simple CNN	Fantasia	99.42	99.42	99.98
DenseNet	Fantasia	99.79	99.78	99.99
Simple CNN	ECG-ID	94.23	94.26	99.94
DenseNet	ECG-ID	96.88	96.89	99.96

Table 6.2: Across-session classification performance for ECG-ID (%).

Model	Accuracy	Sensitivity	Specificity
Simple CNN	73.54	72.72	99.70
DenseNet	73.28	72.46	99.70

(or inter-session) task is formulated as a way to test for generalization performance. In this setting, one session from each subject is used for training, while the other is used for testing.

The CNNs were trained on 67% of the data and tested on the remaining 33%. For the Fantasia database, spectrograms were randomly chosen throughout the entire length of every signal, as the recording time is much longer when compared to ECG-ID, of which all the generated spectrograms were used.

The results presented on Table 6.4 suggest that the combination of deep learning and spectrograms can be used effectively for human identification, only needing a recording of 6 seconds to be able to accurately classify a subject on a universe of 40.

In the across-session experiments 6.2, both models reach similar results. This may have to do with the fact that these models were overfitting to a small training set (10s) with a very high overlap percentage (93.75%), which generated very similar spectrograms, not enabling the models to learn some of the features required to better distinguish between subjects. In this case, using a larger model does not bring any benefits in performance.

In the within-session modality, the performance of these models reaches state of the art on both Fantasia and ECG-ID databases and the values of sensitivity and specificity presented on Table 6.1 are in a reasonable range. In this setting, as hypothesized, DenseNet outperforms the baseline CNN.

The presented approach is not as accurate on ECG-ID database (Table 6.3), when compared with the results for Fantasia. This happens due to the large number of subjects, along with the short recording time, only allowing the generation of 366 spectrograms per subject, even with a considerable overlap percentage (93.75%). While the accuracy values were not higher than most of the related works, we believe them to be at the same level, having in mind that only a subset of the database is required. In addition

Table 6.3: Within-session accuracy comparison for ECG-ID Database. PCA, LDA and (W)NM stand, respectively, for Principal Component Analysis, Linear Discriminant Analysis and (Weighted) Nearest Mean.

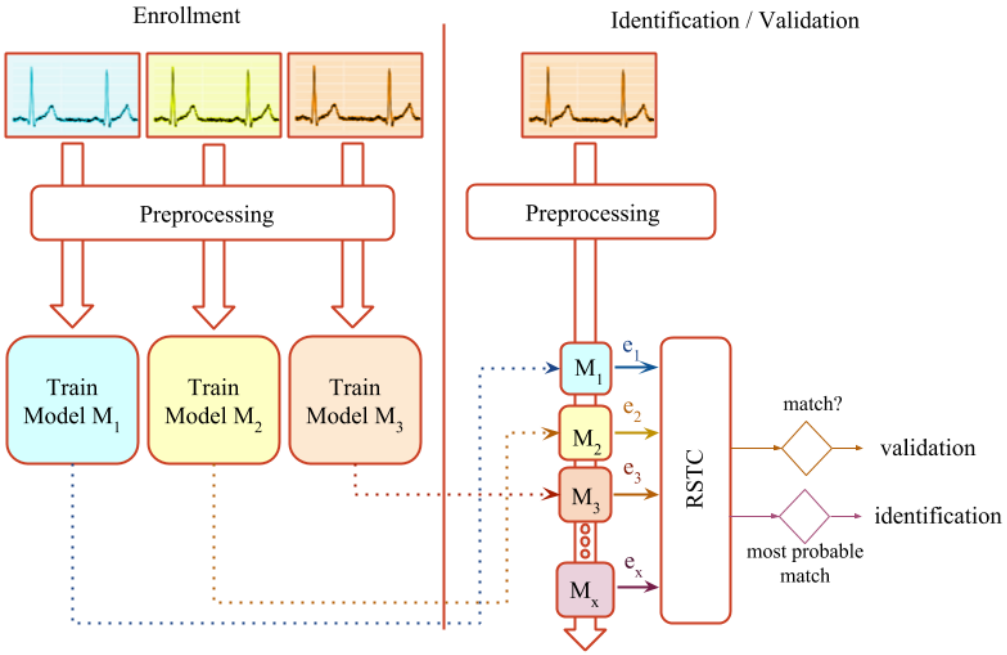
Work	Method	Accuracy (%)
Lugovaya [14]	Wavelets, PCA and LDA + (W)NM ensemble	96
Tan & Perkowski [52]	Wavelets + Probabilistic RF	98.79
[57]	NNs for QRS detection and classification	99.96
Salloum & Kuo [73]	LSTM (3 beats as input)	98.2
Salloum & Kuo [73]	LSTM (9 beats as input)	100
Proposed	Spectrograms + Small CNN	94.23
Proposed	Spectrograms + DenseNet	96.88

to this, past experiments on ECG-ID database have had very diverse setups: Lugovaya [14], Tan & Perkowski [52] and Page et al. [57], although having used all the available data, which varies from 2 to 20 sessions per subject, the splitting between training and test sets diverged considerably; Salloum & Kuo [73] used only one session per subject, however, the inputs were fed in sequence lengths of either 3 or 9 heartbeats out of 18 total; while our method uses 2 sessions per subject and 4 second segments. As there are no widely followed standard procedures for ECG biometrics tasks, it is hard to fairly compare results across different works.

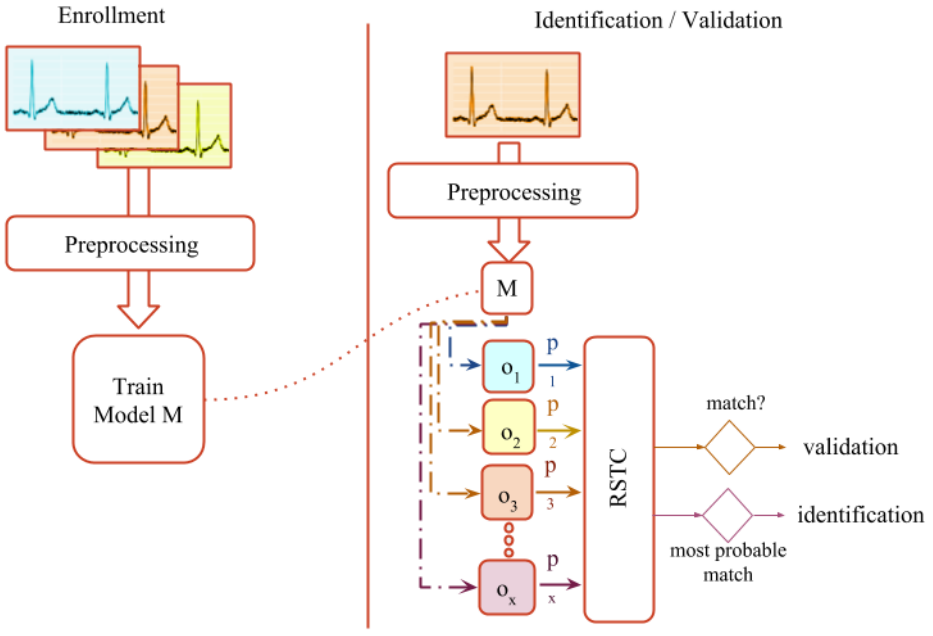
An advantage of these algorithms is their robustness to variations in the moment of signal acquisition, as spectrograms with small offsets in time are correctly identified without the need for QRS detection. This experiment demonstrates there is potential in the use of spectrograms for biometric recognition applications. Using data acquired across different sessions has shown that this method is not very reliable. The second experiment goes deeper into this issue, trying to understand which models generalize better over time.

Table 6.4: Accuracy comparison for Fantasia Database. RBF and RF stand, respectively, for Radial Basis Function and Random Forest. PRNN stands for personalized RNN.

Work	Method	Accuracy (%)
Tantawi et al.[45]	Wavelets + RBF NN	95.89
Zhang et al. [84]	Hand-crafted features + RF	98
Zhang et al.[63]	Wavelets + 1D-CNN	97.2
Exp 1	Spectrograms + Small CNN	99.42
Exp 1	Spectrograms + DenseNet	99.79
Exp 2	Segments + PRNN	100
Exp 2	Segments and Heartbeats + TCNN	97.27



(a) Personalized setting



(b) Non-personalized setting

Figure 6.2: Biometric authentication setting. RSTC stands for Relative Score Threshold Classifier, which attributes each template to a subject by comparing the normalized scores given by the outputs of each model.

6.2 ECG Biometrics using Signal Segments and Fusion

In this experiment, segments or heartbeats are extracted from the filtered signals and fed directly into DNNs with the ability of processing one-dimensional data. Then, the model attributes a class to each segment, according to the subject from whom these signals were generated. Comparisons are made between RNN and CNN architectures, using personalized or non-personalized models. The models were trained on 70% of the data and tested on the remaining 30%.

6.2.1 Personalized models

In this case (Fig. 6.2a), a new model is trained for each subject. While it may be slower than training a single model for every subject, it could in principle achieve better results. This is based on the hypothesis that having a NN focused solely on a single subject can converge faster, due to a lower data complexity. A GRU, a type of RNN, is used for this task.

Tables 6.5 and 6.4 show the results for this model on MIT-BIH and Fantasia. While it underperforms for segments of 2 seconds, it reaches positive results if averaging predictions over a longer time window, as it can be seen in Figure 6.3. All the results shown for the personalized RNN were obtained in this longer setting. Although results can be competitive, requiring each subject to record almost 2 minutes of ECG data for every access is impractical in real world systems. For MIT-BIH, final EER was reached after 71 seconds, while for Fantasia it took 78 seconds to achieve an EER of 0.021%. [73] used a similar way to improve results, by using 9 out of 18 total heartbeats as input to achieve perfect scores.

Table 6.5: Performance comparison for the MIT-BIH database (%). PRNN stands for personalized RNN.

Work	Method	Accuracy	EER
Wang et al. [46]	DCT + Autocorrelation	97.8	-
Karegar et al. [85]	Nonlinear features + SVM	96.07	3.938
Salloum & Kuo [73]	RNN (3 HBs)	98.6	-
Salloum & Kuo [73]	RNN (9 HBs)	100	0
Proposed	PRNN	79.80	1.714
Proposed	TCNN	95.62	1.265

6.2.2 Non-personalized models

In this setting, a single model is trained to assign an ECG segment to one of many subjects registered in the system (Fig. 6.2b).

The TCNNs used in this experiment can receive either segments or heartbeats. Two or more heartbeats are isolated from each segment and randomly sampled to be used as input to a CNN. As each model has a different view of the same segment, we consider the hypothesis that a fusion of both CNNs can perform better than the best single model. This hypothesis is based on the fact that at least two heartbeats are required to obtain non-fiducial features, while, if the neural network focuses on a single heartbeat, it could be able extract better fiducial features.

A smaller version of the RNN model used for the personalized setting is compared with the proposed CNN models. This model only receives ECG segments as input.

Model architectures were only optimized for CYBHi, using a subset of the M1-M1 training data. No hyperparameter optimization was performed for either Fantasia or MIT-BIH. To improve the ability to generalize across sessions, only 10 epochs of training were used for all the experiments with CNNs, as during validation, loss values started to stabilize and it is well known that early stopping can improve generalization [86]. Further training would likely improve results for within-session tasks, however, creating similar training conditions provides a more direct comparison between both modalities. As it is not trivial to constrain the models to only learn features that generalize across sessions, a good way to infer the degree of overfitting to the session is the variation between intra and inter-session results.

For Fantasia and MIT-BIH, comparisons are made in Tables 6.4 and 6.5, respectively. Even without explicit optimization, results are on par with the state of the art.

For CYBHi, in the presented figures, tasks are encoded by the sessions used for training and testing. For example, "M1-M2"means training on the first session and testing on

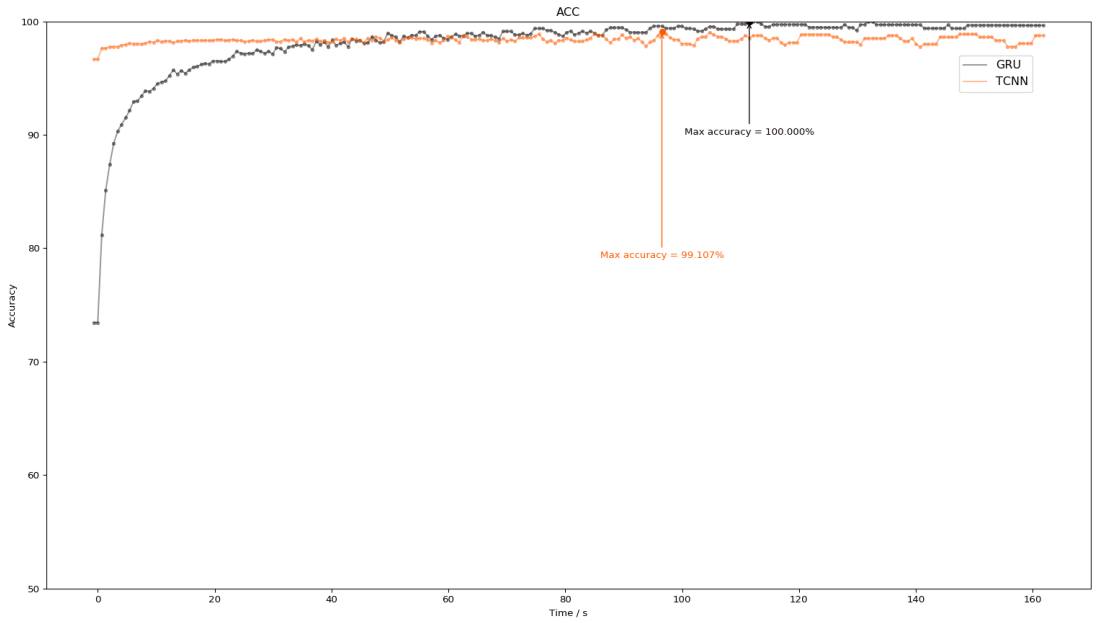


Figure 6.3: Evolution of accuracy values with an increasing prediction time window for the Fantasia database.

the second. The full results for the TCNN using thresholding for segment elimination are shown in Table 6.6. Table 6.8 shows more extensive evaluation, as the main results of this work. Results are presented for the RNN and for all three TCNN variants. TCNN indicates the base model decribed in Section 5 using thresholding for segment elimination, TCNN+Clustering denotes the model using k-means clustering before segmentation to remove motion artifacts and TCNN+AE refers to the variant using a supervised autoencoder. In all the tasks, CNNs achieve better results than the RNN. Even having a simpler architecture, this RNN, requiring 368 seconds per training epoch, still takes much longer to run than the CNN models, at 15 seconds per epoch. This model comparison can be seen as unfair, considering that the RNN takes advantage from much more computation.

All TCNN variants outperformed the previous state of the art for across-session tasks, the most indicative task for performance in real biometric systems. Applying k-means clustering before segmentation and using an autoencoder as a regularizer seem to improve results. More model runs could give values of standard deviation, however, there was no substantial variability between random seeds. Overall, the most natural choice for a commercial system would be TCNN+AE, due to lower EERs and higher accuracies for across-session tasks. It also presents low variation in the EERs between tasks M1-M2 and M2-M1, which may indicate more stability.

The notion of overfitting is further explored with the RNN model. From these results (Table 6.7), it can be seen that while the within-session EER has a relative 50% decrease from 20 to 30 epochs, the across-session performance decreases from 20% to 19% on M1-M2 and has a slight increase on M2-M1. Further training only aggravates this issue, leading either to a bigger difference between intra and inter-session tasks or to unstable training, which was verified in later epochs. This indicates that DNNs can overfit and that early stopping can be a good way to avoid it, possibly improving generalization.

As it can be seen in Figure 6.4, more epochs of training can lead to improved results. However, these results were obtained on test data, so this knowledge cannot be propagated to the experiments, as test data should remain unseen. The initial choice of 10 training epochs was made to avoid overfitting. This figure shows that this choice might have been a good one, as across-session accuracy ceases to increase steadily after that value. Due to gradient instability in later epochs, values of learning rate (0.001) and exponential decay (0.95) used in the experiments had to be reduced to 0.0008 and 0.7, respectively. It can still be noticed that later epochs can generate unstable gradients. A clear example is epoch 15 of task M2-M2, in which the accuracy dropped considerably, right before it settled at its maximum. Having more recording sessions for training would enable similar knowledge to be used by the model, most likely improving its ability to generalize.

A clear situation of overfitting is demonstrated in Table 6.9. In this table, a trade-off between intra and inter-session performance can be verified, as well as very poor temporal generalization without segment normalization. The most likely cause for this phenomenon is that the model learns undesired features which only occur within the

Table 6.6: Performance for the TCNN model on CYBHi (%).

Task	Input	Accuracy	Sensitivity	Specificity	EER
M1-M1	Segments	83.23	83.80	99.73	3.847
	Heartbeats	78.96	79.24	99.66	6.174
	Fusion	88.05	88.37	99.81	**3.013**
M2-M2	Segments	85.54	85.47	99.77	3.441
	Heartbeats	79.74	79.77	99.67	6.489
	Fusion	89.96	89.99	99.84	**2.632**
M1-M2	Segments	47.86	47.95	99.16	11.11
	Heartbeats	48.89	48.44	99.18	14.41
	Fusion	54.98	54.61	99.27	**10.25**
M2-M1	Segments	45.31	45.25	99.12	**12.88**
	Heartbeats	47.79	47.92	99.16	17.12
	Fusion	53.61	53.63	99.25	13.44

Table 6.7: Performance for the non-personalized RNN model on CYBHi (%).

Task	Epochs	Accuracy	Sensitivity	Specificity	EER
M1-M1	20	58.10	58.59	99.32	17.63
	30	67.42	67.96	99.47	**11.09**
M2-M2	20	57.21	57.29	99.31	14.84
	30	67.94	68.15	99.48	**10.09**
M1-M2	20	28.45	28.13	98.85	20.20
	30	32.61	32.56	99.91	**19.00**
M2-M1	20	30.32	30.44	98.88	**17.85**
	30	30.04	30.18	99.87	**17.89**

same session. These features can be viewed as confounders [87], which are correlated with the label, but are not its true cause. One way to correct for this issue would be by training the model with data acquired in various settings. This could be done, for example, by varying the subject's heart rate or the amplifier gain. In the presented experiment, as a partial correction for this issue, normalizing each individual segment led to much better generalization across sessions. This can be seen as a way to simulate diverse values of amplifier gain. Another technique that can be applied for this purpose is the data augmentation method presented in [11], which scales either the full heartbeat or a particular component. This augmentation is also reported to be crucial in their work.

When comparing personalized with non-personalized approaches, it can be seen that non-personalized models reach lower error rates. This may be due to the use of RNNs, as their weaker performance may be masking an otherwise promising approach.

Table 6.8: Performance comparison for CYBHi (%).

Task	Work	Accuracy	EER
M1-M1	Marques [56]	-	3.97±0.22
	Luz et al. [11]	-	**1.33**
	RNN	67.42	11.09
	TCNN	88.05	3.847
	TCNN+Clustering	**92.59**	2.073
	TCNN+AE	89.54	3.150
M2-M2	RNN	67.94	10.09
	TCNN	89.96	2.632
	TCNN+Clustering	**94.47**	**1.654**
	TCNN+AE	90.56	2.709
M1-M2	Marques [56]	-	14.31
	Luz et al. [11]	-	12.78
	RNN	32.61	19.00
	TCNN	54.98	**10.25**
	TCNN+Clustering	47.26	11.38
	TCNN+AE	**55.58**	10.57
M2-M1	Luz et al. [11]	-	13.93
	RNN	30.32	17.85
	TCNN	53.61	12.88
	TCNN+Clustering	49.70	**9.972**
	TCNN+AE	**58.91**	**10.01**

Table 6.9: Performance for the TCNN+fusion model on CYBHi with vs without segment normalization (%).

Task	Normalization	Accuracy	EER
M1-M1	No	92.67	**2.874**
	Yes	88.05	3.013
M2-M2	No	92.58	**2.404**
	Yes	89.96	2.632
M1-M2	No	11.60	23.56
	Yes	54.98	**10.25**
M2-M1	No	16.65	37.59
	Yes	53.61	**13.44**

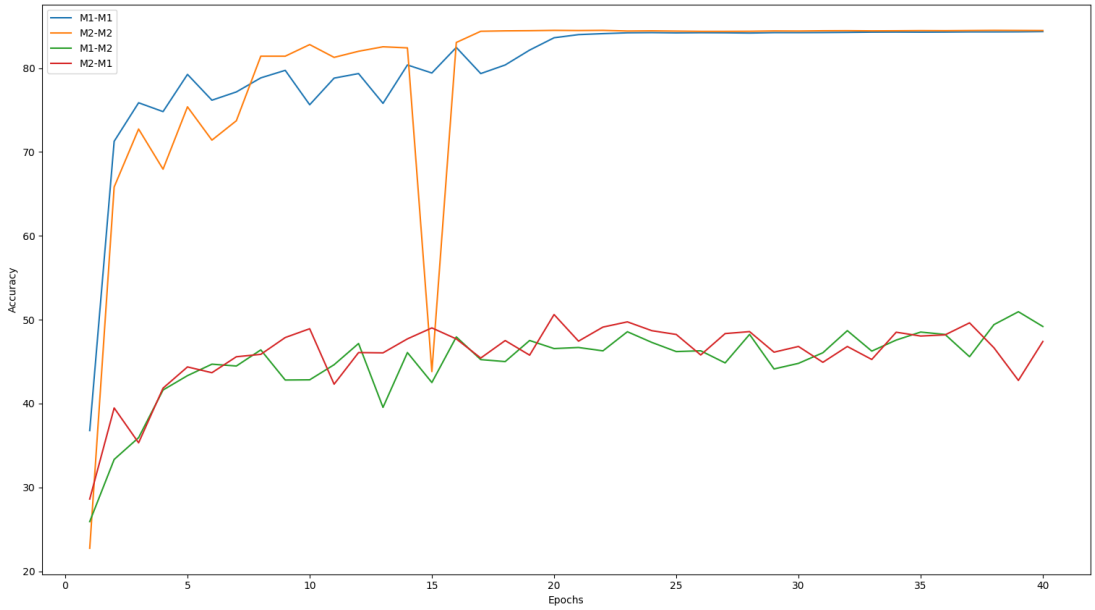


Figure 6.4: Evolution of accuracy with increasing training epochs on CYBHi, using TCNN with signal segments as input. Maximum accuracy values for M1-M1, M2-M2, M1-M2 and M2-M1 are, respectively, 84.35, 50.95, 84.50 and 50.61.

The personalized model is also much slower than all the others, requiring several days to complete training. This happens not only because it uses a large RNN, but also because each new subject requires a unique model to be trained from scratch. In future work, parts of these models can be shared between subjects as feature extractors, leaving only a smaller and faster classifier as the final layer. This way, the personalized modality can become more scalable than its non-personalized counterpart, as this kind of approach requires a much larger output layer, with as many classes as subjects in the system.

Segment elimination did not provide significant improvements when applied to both training and testing. However, it could lead to better results if only applied to remove testing data, since samples containing artifacts lose useful information and become harder to attribute. This is similar to the procedure of adding a rejection class to the model or a threshold of model uncertainty, which automatically rejects samples that are not well recognized. Future work on this subject would likely improve biometric systems.

Chapter 7

Conclusion and Future work

In the experiments performed for this thesis, small extracted segments were confidently attributed to the subject from whom they were generated. This was possible through the recognition of temporal and frequential information existent in ECG signals by DNNs.

This work reaches state of the art performance in important benchmark datasets by using models that are not only able to extract features automatically, but can also generalize to unseen data. This represents another step towards more capable and robust biometric identification and authentication systems.

Using segments instead of heartbeats for authentication provides a much higher number of templates for the same signal length, potentially increasing the performance of learning algorithms, such as DNNs. It can also avoid performing QRS detection, which is not fail-safe, to obtain templates. Nonetheless, this work shows that performing a fusion of both methods generally brings improvements.

Having the obtained results into account, it can be verified that, for these tasks, CNNs largely outperform RNNs. By using dilated convolutions, an effective mechanism to relate temporal data and propagate the error, all the TCNN variants were able to learn long-term dependencies in a more efficient way. In contrast, RNNs have a much slower convergence, possibly due to the vanishing gradient problem.

Although the achieved results show good performance rates, especially within the same session, there is a considerable margin for improvement. Performance in across-session tasks is still insufficient. This is the most relevant modality, as biometric systems must function for an indefinite amount of times and throughout several years. Although datasets such as CYBHi and ECG-ID are helping to improve it, there is a very large gap to within-session results.

A way to improve this lack in robustness is to take advantage of the fast growing trend in digital health, as the extensive amount of ECG data measured by wearable devices can

be used to improve current systems. A similar idea has been applied by Clifford et al. [4], which used a large open-sourced ECG database collected from one of these devices for an international challenge with the objective of detecting atrial fibrillation events.

Another direction could be the use of several datasets for transfer learning, as it may enable models to learn a better representation of the signal, especially in low data regimes, as is the case of ECG-ID database. This would likely improve its ability to differentiate between more subjects and hopefully needing less training data for each one.

Further development could also be made regarding signal preprocessing methods, as the effect of noise and artifacts present in the signal is a cause for recognition error. This could be achieved by explicitly learning to denoise ECG signals. Using models that are inherently robust to noise is another promising approach to deal with the same issue.

Bibliography

[1] J. A. Unar, W. C. Seng, and A. Abbasi. "A review of biometric technology along with trends and prospects." In: *Pattern Recognition* 47.8 (2014), pp. 2673–2688. ISSN: 00313203. DOI: 10.1016/j.patcog.2014.01.016. URL: http://dx.doi.org/10.1016/j.patcog.2014.01.016.

[2] A. Nait-Ali. "Hidden biometrics: Towards using biosignals and biomedical images for security applications." In: *Systems, Signal Processing and their Applications (WOSSPA), 2011 7th International Workshop on*. IEEE. 2011, pp. 352–356.

[3] A. Condon and G. Willatt. "ECG biometrics: the heart of data-driven disruption?" In: *Biometric Technology Today* 2018.1 (2018), pp. 7–9.

[4] G. D. Clifford, C. Liu, B. Moody, L.-w. H. Lehman, I. Silva, Q. Li, A. Johnson, and R. G. Mark. "AF classification from a short single lead ECG recording: The Physionet Computing in Cardiology Challenge 2017." In: *Proceedings of Computing in Cardiology* 44 (2017), p. 1.

[5] L. Biel, O. Pettersson, L. Philipson, and P. Wide. "ECG analysis: a new approach in human identification." In: *IEEE Transactions on Instrumentation and Measurement* 50.3 (2001), pp. 808–812.

[6] J. Rodrigues, D. Belo, and H. Gamboa. "Noise detection on ECG based on agglomerative clustering of morphological features." In: *Computers in biology and medicine* 87 (2017), pp. 322–334.

[7] S. C. Fang and H. L. Chan. "QRS detection-free electrocardiogram biometrics in the reconstructed phase space." In: *Pattern Recognition Letters* 34.5 (2013), pp. 595–602. ISSN: 01678655. DOI: 10.1016/j.patrec.2012.11.005. URL: http://dx.doi.org/10.1016/j.patrec.2012.11.005.

[8] D. Belo, J. Rodrigues, J. R. Vaz, P. Pezarat-Correia, and H. Gamboa. "Biosignals learning and synthesis using deep neural networks." In: *BioMedical Engineering OnLine* 16.1 (2017), p. 115. ISSN: 1475-925X. DOI: 10.1186/s12938-017-0405-0. URL: https://doi.org/10.1186/s12938-017-0405-0.

[9] A. Eduardo, H. Aidos, and A. Fred. "ECG-based Biometrics using a Deep Autoencoder for Feature Learning." In: (2017).

[10] S. Kiranyaz, T. Ince, and M. Gabbouj. "Real-time patient-specific ECG classification by 1-D convolutional neural networks." In: *IEEE Transactions on Biomedical Engineering* 63.3 (2016), pp. 664–675.

[11] E. Luz, G. Moreira, L. S. Oliveira, W. R. Schwartz, and D. Menotti. "Learning Deep Off-The-Person Heart Biometrics Representations." In: *IEEE Transactions on Information Forensics and Security* (2017).

[12] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. http://www.deeplearningbook.org. MIT Press, 2016.

[13] N. Iyengar, C. Peng, R. Morin, A. L. Goldberger, and L. A. Lipsitz. "Age-related alterations in the fractal scaling of cardiac interbeat interval dynamics." In: *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* 271.4 (1996), R1078–R1084.

[14] T. Lugovaya. "Biometric human identification based on electrocardiogram." In: *LETI* (2005).

[15] G. B. Moody and R. G. Mark. "The impact of the MIT-BIH arrhythmia database." In: *IEEE Engineering in Medicine and Biology Magazine* 20.3 (2001), pp. 45–50.

[16] H. P. Da Silva, A. Lourenço, A. Fred, N. Raposo, and M. Aires-de Sousa. "Check Your Biosignals Here: A new dataset for off-the-person ECG biometrics." In: *Computer methods and programs in biomedicine* 113.2 (2014), pp. 503–514.

[17] M. Vetterli, J. Kovačević, and V. K. Goyal. *Foundations of signal processing*. Cambridge University Press, 2014.

[18] K. El-Shennawy. *Communication theory and signal processing for transform coding*. Bentham Science Publishers, 2014.

[19] S. Thalkar and D. Upasani. "Various techniques for removal of power line interference from ECG signal." In: *Int. J. Sci. Eng. Res* 4.12 (2013), pp. 12–23.

[20] H. Gamboa. "Multi-Modal Behavioral Biometrics Based on HCI and Electrophysiology." Doctoral dissertation. PhD thesis, Universidade Técnica de Lisboa, Instituto Superior Técnico, 2008.

[21] C. Cepeda, J. Rodrigues, M. C. Dias, D. Oliveira, D. Rindlisbacher, M. Cheetham, and H. Gamboa. "Mouse Tracking Measures and Movement Patterns with Application for Online Surveys." In: *International Cross-Domain Conference for Machine Learning and Knowledge Extraction*. Springer. 2018, pp. 28–42.

[22] F. Gargiulo, A. Fratini, M. Sansone, and C. Sansone. "Subject identification via ECG fiducial-based systems: Influence of the type of QT interval correction." In: *Computer Methods and Programs in Biomedicine* 121.3 (2015), pp. 127–136. ISSN: 18727565. DOI: 10.1016/j.cmpb.2015.05.012. URL: http://dx.doi.org/10.1016/j.cmpb.2015.05.012.

[23] A Lourenco, J. Cardoso, and J. T. Pinto. "Evolution, Current Challenges, and Future Possibilities in ECG Biometrics." In: (2018).

[24] M. A. Nielsen. *Neural networks and deep learning*. Determination Press, 2015.

[25] K. Hornik, M. Stinchcombe, and H. White. "Multilayer feedforward networks are universal approximators." In: *Neural networks* 2.5 (1989), pp. 359–366.

[26] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. "Gradient-based learning applied to document recognition." In: *Proceedings of the IEEE* 86.11 (1998), pp. 2278–2324.

[27] I. Sutskever. "TRAINING RECURRENT NEURAL NETWORKS." Doctoral dissertation. University of Toronto, 2013.

[28] S. Hochreiter and J. Schmidhuber. "Long short-term memory." In: *Neural computation* 9.8 (1997), pp. 1735–1780.

[29] K. Cho, B. Van Merriënboer, D. Bahdanau, and Y. Bengio. "On the properties of neural machine translation: Encoder-decoder approaches." In: *arXiv preprint arXiv:1409.1259* (2014).

[30] J. H.D. M. Goulart, P. M. R. de Oliveira, R. C. Farias, V. Zarzoso, and P. Comon. "Alternating group lasso for block-term tensor decomposition with application to ECG source separation." In: (2018).

[31] M. AlMahamdy and H. B. Riley. "Performance study of different denoising methods for ECG signals." In: *Procedia Computer Science* 37 (2014), pp. 325–332.

[32] H. Mamaghanian, N. Khaled, D. Atienza, and P. Vandergheynst. "Compressed sensing for real-time energy-efficient ECG compression on wireless body sensor nodes." In: *IEEE Transactions on Biomedical Engineering* 58.9 (2011), pp. 2456–2466.

[33] Z. Zhang, T.-P. Jung, S. Makeig, and B. D. Rao. "Compressed sensing for energy-efficient wireless telemonitoring of noninvasive fetal ECG via block sparse Bayesian learning." In: *IEEE Transactions on Biomedical Engineering* 60.2 (2013), pp. 300–309.

[34] O. Yildirim, R. San Tan, and U. R. Acharya. "An efficient compression of ECG signals using deep convolutional autoencoders." In: *Cognitive Systems Research* 52 (2018), pp. 198–211.

[35] J. Park, W. Pedrycz, and M. Jeon. "Ischemia episode detection in ECG using kernel density estimation, support vector machine and feature selection." In: *Biomedical engineering online* 11.1 (2012), p. 30.

[36] R. Xiao, Y. Xu, M. M. Pelter, D. W. Mortara, and X. Hu. "A Deep Learning Approach to Examine Ischemic ST Changes in Ambulatory ECG Recordings." In: *AMIA Summits on Translational Science Proceedings* 2017 (2018), p. 256.

[37] U. R. Acharya, H. Fujita, S. L. Oh, Y. Hagiwara, J. H. Tan, M. Adam, and R. S. Tan. "Deep convolutional neural network for the automated diagnosis of congestive heart failure using ECG signals." In: *Applied Intelligence* (2018), pp. 1–12.

[38] K. Polat and S. Güneş. "Detection of ECG Arrhythmia using a differential expert system approach based on principal component analysis and least square support vector machine." In: *Applied Mathematics and Computation* 186.1 (2007), pp. 898–906.

[39] S Osowski, T Markiewicz, and L. T. Hoai. "Recognition and classification system of arrhythmia using ensemble of neural networks." In: *Measurement* 41.6 (2008), pp. 610–617.

[40] Ö. Yıldırım, P. Pławiak, R.-S. Tan, and U. R. Acharya. "Arrhythmia detection using deep convolutional neural network with long duration ECG signals." In: *Computers in biology and medicine* 102 (2018), pp. 411–420.

[41] M. Zihlmann, D. Perekrestenko, and M. Tschannen. "Convolutional recurrent neural networks for electrocardiogram classification." In: *Computing* 44 (2017), p. 1.

[42] T. Golany and K. Radinsky. "PGANs: Personalized Generative Adversarial Networks for ECG Synthesis to Improve Patient-Specific Deep ECG Classification." In: (2019).

[43] E. L. Van Den Broek. "Beyond biometrics." In: *Procedia Computer Science* 1.1 (2010), pp. 2511–2519.

[44] F. G. S. Teodoro, S. M. Peres, and C. A. Lima. "Feature selection for biometric recognition based on electrocardiogram signals." In: *Neural Networks (IJCNN), 2017 International Joint Conference on*. IEEE. 2017, pp. 2911–2920.

[45] M. Tantawi, K. Revett, A.-B. Salem, and M. F. Tolba. "ECG based biometric recognition using wavelets and RBF neural network." In: *Proc. 7th Eur. Comput. Conf.(ECC)*. 2013, pp. 100–105.

[46] Y. Wang, F. Agrafioti, D. Hatzinakos, and K. N. Plataniotis. "Analysis of human electrocardiogram for biometric recognition." In: *EURASIP journal on Advances in Signal Processing* 2008.1 (2007), p. 148658.

[47] S. Saechia, J. Koseeyaporn, and P. Wardkein. "Human identification system based ECG signal." In: *TENCON 2005 2005 IEEE Region 10*. IEEE. 2005, pp. 1–4.

[48] M. M. Tantawi, K. Revett, A. Salem, and M. F. Tolba. "Fiducial feature reduction analysis for electrocardiogram (ECG) based biometric recognition." In: *Journal of Intelligent Information Systems* 40.1 (2013), pp. 17–39. ISSN: 09259902. DOI: 10.1007/s10844-012-0214-7.

[49] J. Shen, S.-D. Bao, L.-C. Yang, and Y. Li. "The PLR-DTW method for ECG based biometric identification." In: *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*. IEEE. 2011, pp. 5248–5251.

[50] S. A. Israel, J. M. Irvine, A. Cheng, M. D. Wiederhold, and B. K. Wiederhold. "ECG to identify individuals." In: *Pattern recognition* 38.1 (2005), pp. 133–142.

[51] I. Chamatidis, A. Katsika, and G. Spathoulas. "Using deep learning neural networks for ECG based authentication." In: *2017 International Carnahan Conference on Security Technology (ICCST)*. 2017, pp. 1–6. DOI: 10.1109/CCST.2017.8167816.

[52] R. Tan and M. Perkowski. "ECG Biometric Identification Using Wavelet Analysis Coupled with Probabilistic Random Forest." In: *Machine Learning and Applications (ICMLA), 2016 15th IEEE International Conference on*. IEEE. 2016, pp. 182–187.

[53] J. R. Pinto, J. S. Cardoso, A. Lourenço, and C. Carreiras. "Towards a continuous biometric system based on ECG signals acquired on the steering wheel." In: *Sensors* 17.10 (2017), p. 2228.

[54] H. Ferdinando, T. Seppänen, and E. Alasaarela. "Bivariate empirical mode decomposition for ECG-based biometric identification with emotional data." In: *Engineering in Medicine and Biology Society (EMBC), 2017 39th Annual International Conference of the IEEE*. IEEE. 2017, pp. 450–453.

[55] J. M. Carvalho, S. Brás, and A. J. Pinho. "Compression-Based ECG Biometric Identification Using a Non-fiducial Approach." In: *arXiv preprint arXiv:1804.00959* (2018).

[56] F. Marques. "ECG Biometrics: A Dissimilarity Representation Approach." In: ().

[57] A. Page, A. Kulkarni, and T. Mohsenin. "Utilizing deep neural nets for an embedded ECG-based biometric authentication system." In: *Biomedical Circuits and Systems Conference (BioCAS), 2015 IEEE*. IEEE. 2015, pp. 1–4.

[58] G. Zheng, S. Ji, M. Dai, and Y. Sun. "ECG Based Identification by Deep Learning." In: *Biometric Recognition*. Ed. by J. Zhou, Y. Wang, Z. Sun, Y. Xu, L. Shen, J. Feng, S. Shan, Y. Qiao, Z. Guo, and S. Yu. Cham: Springer International Publishing, 2017, pp. 503–510. ISBN: 978-3-319-69923-3.

[59] K. He, X. Zhang, S. Ren, and J. Sun. "Identity mappings in deep residual networks." In: *European conference on computer vision*. Springer. 2016, pp. 630–645.

[60] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger. "Densely Connected Convolutional Networks." In: *CVPR*. Vol. 1. 2. 2017, p. 3.

[61] H. Zhang, I. McLoughlin, and Y. Song. "Robust sound event recognition using convolutional neural networks." In: *2015 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE. 2015, pp. 559–563.

[62] S. Kiranyaz, T. Ince, R. Hamila, and M. Gabbouj. "Convolutional Neural Networks for patient-specific ECG classification." In: *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*. IEEE. 2015, pp. 2608–2611.

[63] Q. Zhang, D. Zhou, and X. Zeng. "HeartID: A Multiresolution Convolutional Neural Network for ECG-based Biometric Human Identification in Smart Health Applications." In: *IEEE Access* (2017).

[64] R. D. Labati, E. Muñoz, V. Piuri, R. Sassi, and F. Scotti. "Deep-ECG: Convolutional Neural Networks for ECG biometric recognition." In: *Pattern Recognition Letters* (2018).

[65] A. Van Den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu. "Wavenet: A generative model for raw audio." In: *CoRR abs/1609.03499* (2016).

[66] S. Bai, J. Z. Kolter, and V. Koltun. "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling." In: *arXiv preprint arXiv:1803.01271* (2018).

[67] A. Graves, G. Wayne, M. Reynolds, T. Harley, I. Danihelka, A. Grabska-Barwińska, S. G. Colmenarejo, E. Grefenstette, T. Ramalho, J. Agapiou, et al. "Hybrid computing using a neural network with dynamic external memory." In: *Nature* 538.7626 (2016), p. 471.

[68] J. Ba, G. E. Hinton, V. Mnih, J. Z. Leibo, and C. Ionescu. "Using fast weights to attend to the recent past." In: *Advances in Neural Information Processing Systems*. 2016, pp. 4331–4339.

[69] M. Jaderberg, W. M. Czarnecki, S. Osindero, O. Vinyals, A. Graves, D. Silver, and K. Kavukcuoglu. "Decoupled neural interfaces using synthetic gradients." In: *arXiv preprint arXiv:1608.05343* (2016).

[70] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. "Attention is all you need." In: *Advances in Neural Information Processing Systems*. 2017, pp. 5998–6008.

[71] F. Wu, A. Fan, A. Baevski, Y. Dauphin, and M. Auli. "Pay Less Attention with Lightweight and Dynamic Convolutions." In: *CoRR* abs/1901.10430 (2018).

[72] J. Donahue, L. Anne Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, and T. Darrell. "Long-term recurrent convolutional networks for visual recognition and description." In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 2625–2634.

[73] R. Salloum and C.-C. J. Kuo. "ECG-based biometrics using recurrent neural networks." In: *Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on*. IEEE. 2017, pp. 2062–2066.

[74] Z. Zhao, Y. Zhang, Y. Deng, and X. Zhang. "ECG authentication system design incorporating a convolutional neural network and generalized S-Transformation." In: *Computers in biology and medicine* 102 (2018), pp. 168–179.

[75] S. Wu, T. Zhang, B. Wu, C. Liu, and J. Xiao. "Single-Pixel Camera in the Visible Band With Fiber Signal Collection." In: *IEEE Access* 6 (2018), pp. 17768–17775. DOI: 10.1109/ACCESS.2018.2819358. URL: https://doi.org/10.1109/ACCESS.2018.2819358.

[76] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley. "Physiobank, physiotoolkit, and physionet." In: *Circulation* 101.23 (2000), e215–e220.

[77] M. Merone, P. Soda, M. Sansone, and C. Sansone. "ECG databases for biometric systems: A systematic review." In: *Expert Systems with Applications* 67 (2017), pp. 189–202. ISSN: 09574174. DOI: 10.1016/j.eswa.2016.09.030. URL: http://dx.doi.org/10.1016/j.eswa.2016.09.030.

[78] G. Montavon, G. B. Orr, and K.-R. Müller. "Tricks of the Trade." In: (1998).

[79] G. Lenis, N. Pilia, A. Loewe, W. H. Schulze, and O. Dössel. "Comparison of baseline wander removal techniques considering the preservation of ST changes in the ischemic ECG: a simulation study." In: *Computational and mathematical methods in medicine* 2017 (2017).

[80] J. Pan and W. J. Tompkins. "A real-time QRS detection algorithm." In: *IEEE Trans. Biomed. Eng* 32.3 (1985), pp. 230–236.

[81] D. Kingma and J. Ba. "Adam: A method for stochastic optimization." In: *arXiv preprint arXiv:1412.6980* (2014).

[82] L. Le, A. Patterson, and M. White. "Supervised autoencoders: Improving generalization performance with unsupervised regularizers." In: *Advances in Neural Information Processing Systems*. 2018, pp. 107–117.

[83] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015. URL: https://www.tensorflow.org/.

[84] Q. Zhang, D. Zhou, and X. Zeng. "Machine learning-empowered biometric methods for biomedicine applications." In: *AIMS Med Sci* 4 (2017), pp. 274–290.

[85] F. P. Karegar, A. Fallah, and S. Rashidi. "Using recurrence quantification analysis and generalized Hurst exponents of ECG for human authentication." In: *2017 2nd Conference on Swarm Intelligence and Evolutionary Computation (CSIEC)*. IEEE. 2017, pp. 66–71.

[86] R. Caruana, S. Lawrence, and C. L. Giles. "Overfitting in neural nets: Backpropagation, conjugate gradient, and early stopping." In: *Advances in neural information processing systems*. 2001, pp. 402–408.

[87] J. Pearl and D. Mackenzie. *The Book of Why: The New Science of Cause and Effect*. Basic Books, 2018.