Biometrics on Mobile Devices Using the Heartbeat

Marcelo Edgar Ferreira dos Santos
Engenharia de Redes e Sistemas Informáticos
Departamento de Ciências de Computadores
2016

Orientador
Luís Filipe Coelho Antunes, Professor Associado,
Faculdade de Ciências da Universidade do Porto

Coorientador
Manuel Eduardo Carvalho Duarte Correia, Professor Auxiliar,
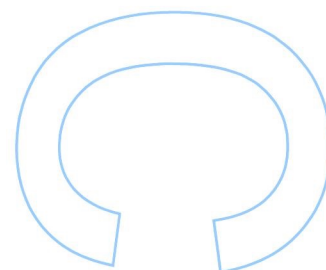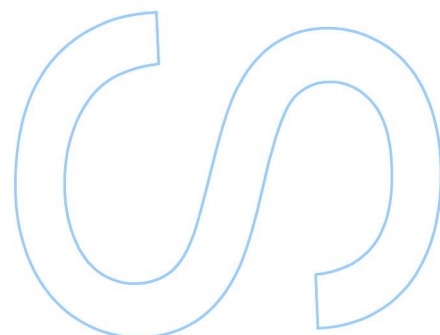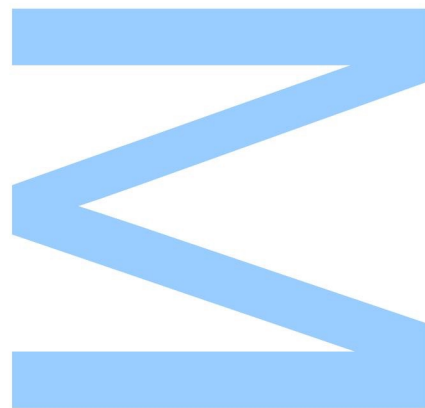Faculdade de Ciências da Universidade do Porto
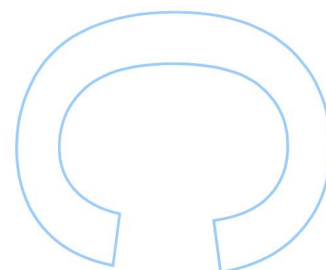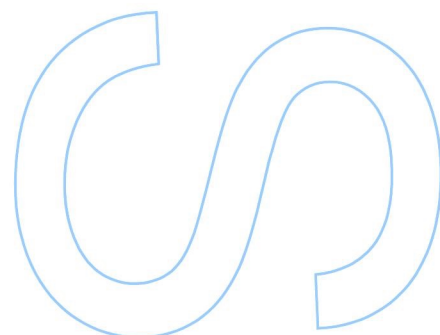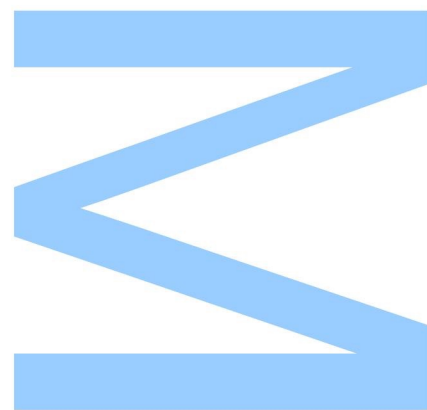
Todas as correções determinadas pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, _____/_____/_____

To my parents

Acknowledgments

I would like to dedicate this thesis to everyone that contributed to my success during my academic career.

To my supervisors Dr. Luis Antunes and Dr. Manuel Correia, I would like to express my gratitude for their guidance and support. They provided me all the necessary tools and complete freedom throughout my dissertation.

I would like to thank my family and friends, for the unconditional support provided and for encouraging me to always aim higher.

To my friends from the Computer Science department, thank you for the advices and all the cheerful moments. It was year of hard work but you definitely made it easier to bare.

Last but certainly not least, thank you Sandrina for all the love, motivation and support, and for always being by my side.

Abstract

As technology evolves, authentication systems tend to maximize its security, while simultaneously minimizing usability related issues. Security systems based on passwords or tokens can be easily compromised enabling the growth of biometric systems. However, since static biometrics, such as fingerprint or hand geometry, may be forged to circumvent systems security, different solutions must be studied.

In this thesis the aforementioned problem will be addressed using biological signals derived from the heartbeat, *i.e.* Electrocardiogram (ECG) and Photoplethysmogram (PPG) waveform signals.
Two main approaches regarding the aforementioned signals were identified, being either *fiducial dependent* or *fiducial independent*. Focusing on the latter and resourcing to a feature extraction method exposed in the literature as AC/DCT, several data sets containing both waveforms were generated from two public databases to evaluate the feasibility of using bio-signals in the context of an authentication scenario in mobile devices.

Using a wristband with a PPG sensor and a mobile device, an authentication system was simulated and evaluated. For comparison purposes, a feature extraction algorithm using time-domain features extracted from Heart Rate Variability (HRV) artifacts was also implemented. Using Random Forest (RF) as the classifier, the AC/DCT approach presented better results, having False Positive Rate (FPR) of 7% and False Negative Rate (FNR) of 11%, and correctly classifying 91.8% of the samples.

Resumo

Com a evolução tecnológica, os sistemas de autenticação tendem a maximizar a segurança e simultâneamente minizimar possíveis questões relacionadas com a usabilidade. Sistemas de segurança baseados em *passwords* e *tokens* podem ser facilmente comprometidos, permitindo o crescimento de sistemas biométricos. Contudo, pelo facto de biometrias estáticas, tais como a impressão digital ou geometria da mão, poderem ser forjadas para contornar os sistemas de segurança, é necessário estudar soluções alternativas.

Nesta tese, os problemas supramencionados serão abordados com recurso a sinais biológicos provenientes do batimento cardíaco, *i.e.* usando ondas de Electrocardiograma (ECG) e Fotopletismograma (PPG).
São identificadas duas abordagens distintas relativas à extracção de informação relativa aos sinais fisiológicos mencionados, sendo estas dependentes ou independentes de pontos fiduciais. Direccionando o foco para a segunda abordagem, e recorrendo ao método de extracção de atributos referenciado na literatura como AC/DCT, foram criados vários *data sets*, através de bases de dados públicas, contendo as ondas referidas, de forma a avaliar a viabilidade da utilização de bio-sinais no contexto de cenários de autenticação em dispositivos móveis.

Recorrendo a uma pulseira com um sensor de PPG e um dispositivo móvel, foi simulado um sistema de autenticação e consequentemente avalidado. Para fins de comparação, foi também implementado um algoritmo de extracção de atributos usando caracteristicas do domínio temporal, extraídos de artefactos relacionados com a variabilidade do batimento cardíaco. Usando o *Random Forest* como classificador, a abordagem que recorre ao método AC/DCT apresenta melhores resultados, verificando taxas de falsos positivos de 7% e falsos negativos de 11%, e classificando correctamente 91.8% das amostras.

Contents

List of Tables

List of Figures

Acronyms

AC   Auto Correlation

AG   Augmentation Index

ANN   Artificial Neural Network

API   Application Programming Interface

AUC   Area Under Curve

Bagging Bootstrap Aggregating

BPM   Beats Per Minute

BVP   Blood Volume Pressure

CT   Crest Time

DBNN Decision Based Neural-Network

DCT   Discrete Cosine Transformation

DTW   Dynamic Time Warping

DWT   Discrete Wavelet Transform

ECG   Electrocardiogram

EEG   Electroencephalogram

EER   Equal Error Rate

EID   Error of Identification

EMG   Electromyogram

FAR   False Acceptance Rate

FFT   Fast Fourier Transformation

FIR   Finite Impulse Response

FMR   False Matching Rate

FN   False Negative

FNR   False Negative Rate

FNMR False Non-Matching Rate

FP   False Positive

FPR   False Positive Rate

FRR   False Rejection Rate

GATT Generic Attributes

HR   Heart Rate

HRV   Heart Rate Variability

IIR   Infinite Impulse Response

KNN   k-Nearest Neighbors

LDA   Linear Discriminant Analysis

LDR   Light Dependent Resistor

LED   Light Emitting Diode

MeanHR Mean Heart Rate

MCC   Matthews Correlation Coefficient

MIC   Maximal Information Coefficient

MIT-BIH MIT-BIH Arrhythmia Database

MLE   Maximum Likelihood Estimator

NB   Naive Bayes

NN   Nearest Neighbors

OS Operating System

PAM Piecewise-Aligned Morphology

PCG Phonocardiogram

PPI Peak to Peak Interval

PPG Photoplethysmogram

PPMCC Pearson Product-Moment Correlation Coefficient

PTB Physikalisch-Technische Bundesanstalt

RF Random Forest

RMSSD Root Mean Square of Successive Differences

ROC Receiver Operating Characteristic

RRI R-R Interval

SIMCA Soft Independent Modeling of Class Analogy

SDHR Standard Deviation of Heart Rate

STFT Short Term Fourier Transformation

SVM Support Vector Machine

TAR True Acceptance Rate

TBME Transactions on Biomedical Engineering

TN True Negative

TNR True Negative Rate

TP True Positive

TPR True Positive Rate

TRR True Rejection Rate

WT Wavelet Transform

Chapter 1

Introduction

In today's society there has been a remarkable growth regarding technology, whether it is personal computers, tablets, smartphones, smartwatches, among others. In order to fulfill the consumer demands, enterprises need to find new ways to constantly transform the user experience into something more pleasant and convenient whilst keeping in mind users security and confidentiality. Technology is widely distributed where each user uses it in order to suit his needs. However, it is up to each one of us to properly use what we are offered, although sometimes such purposes are not the most "noble". Still, it is also up to the people who create, manage and distribute to know the intrinsic hazards.

In the context of convenience, new alternatives are emerging to improve the user experience without compromising other equally relevant aspects which tend to get more exposed and consequently debilitated as a result of the content globalization, namely security issues.
It was established that, in order to attempt to maintain information protected when inserted in a globalized context as the Internet, where information travels freely, the owner must use something (presumably) only known by him, a key - *Password* - as a way to guarantee exclusive access to something he claims to own. The process of proving to be who you claimed to be, by resorting to something as proof of identity is acknowledged as *Authentication*. However, in a highly "connected" society and consequently exposed, it is proven that the credentials system composed by the pair (User, Password) tend to become further obsolete concerning both, usability and intrinsic security factors.

The human being tend to become displeased with routine, protracted and highly repetitive tasks, as the process of authentication using passwords is. He is also limited from the cognitive point of view which impose severe boundaries on the passwords in both, length or type. People tend to associate meanings to passwords as a way to ease its memorization. Due to physical limitations, they also tend to reuse them which introduce further debilities to the data they try to protect.

Such fact led the concept of *Token*. It consists of a physical object used for the purpose of authentication, whose possession provides evidence for the owner to prove his identity. It is,

then, a verification based on possession - "something you own". It is a widely used system and allows multiple usages. It can be used exclusively, for instance a physical key that grants access to restrict places, or together with a knowledge based modality, as the case of passwords, so that, in spite of the object possession, the user must also authenticate with a key, as the case of credit/debit cards.

This new approach regarding authentication introduced a certain level of security, however, it is also liable of being compromised, for instance if the token is stolen or lost.

As a way to suppress those fragilities on systems based on possession or knowledge, the concept of *Biometrics* emerged.

A biometric is "something you are", *i.e.* an innate characteristic, intrinsic to each individual that identifies him, that is, which distinguish him from each other. It is, thus, a personal and non-transmissible characteristic, unlike tokens and passwords. Biometric systems are, hence, systems that use such singular characteristics in order to authenticate them, being able to replace the usage of passwords whilst introducing a new paradigm regarding security in information systems. People have several unique traits that differentiate them, as the case of fingerprint, handwriting, voice, iris or the heartbeat [7, 12, 34, 41]. However, as exploited in the aforementioned systems, some of them may still be forged, for instance using high definition photos do delude facial recognition systems, replicate fingerprints using physical models built in latex or even by creating contact lens that replicate the iris [19, 33, 37, 45].

In this sense, biometric systems that resource to heartbeat are more reliable due to its intrinsic properties desirable in authentication scenarios, such as uniqueness, robustness to attacks, liveness detection and universality. Specifically,

Universality consists of a characteristic that lies in all the population, meaning the system would be usable by the general population and not only by a minor group.

Robustness to attacks lies on the fact that the heart related waveforms appearance, such as Electrocardiogram (ECG) or Photoplethysmogram (PPG), among others, derives from physiological events thus making it harder to mimic or replicate [48].

Uniqueness resides on the influence of its physiological origin. Despite the identical pattern between individuals, there is a vast variability among their biological signals.

Liveness detection is addressed naturally since the signal is obtained directly from the cardiac cycles, enabling a trivial liveness analysis of the subject.

Furthermore, as opposed to static biometrics such as fingerprint or iris, these biological signals enables the use of more dynamic modalities, such as continuous authentication. Also, being firstly used for medical purposes, they can easily be inserted in the context of data minimization, where the data could, not only be used for authentication/identification, but also inserted in a different context like health or condition monitoring, in hospitals or field missions, respectively [1].

Considering the facts presented above, the objective of this thesis is to assess the feasibility of using biological signals related with the heartbeat, specifically the ECG and PPG waveform signals, as a distinguishing factor in a biometric system for mobile devices. Resourcing to the AC/DCT method, presented in [38], to obtain attributes (features) from the signals, different data sets were tested using binary classification where records could be classified as "Owner" or "Intruder" . Regarding the decision stage of the system, *i.e.* the classification of records, Random Forest (RF) algorithm was used, as it outperformed the remaining 4 classifiers considered, namely Linear Discriminant Analysis (LDA). k-Nearest Neighbors (KNN), Support Vector Machine (SVM) and Naive Bayes (NB).

The remainder of this thesis is organized as follows.

- Chapter 2 discloses the concepts regarding the signals used, the pre-processing performed, machine learning algorithms chosen and evaluation criteria followed.

- Chapter 3 covers the prior art in ECG and PPG biometric recognition. Differences between fiducial dependent and independent approaches, major advantages and disadvantages among them and, respective results and conclusions are exposed.

- Chapter 4 exposes in detail the methodology adopted, *i.e.* the properties of all the ECG and PPG databases, including the collected records to perform the experiment, the development environments used, and all the stages of the approach. Signal pre-processing, features extracted and data sets generated, machine learning models variations and, attempted approaches and major difficulties encountered, are also referred.

- Chapter 5 describes the system architecture, hardware specificities, and all the stages and major procedures implemented. Diagrams related with the system operations and the major phases of the implementation are also presented.

- In Chapter 6 all the data sets and respective results are exposed and evaluated. A comparison between fiducial dependent and fiducial independent approaches is performed and a general analysis regarding all the work performed while exposing some pros and cons is presented. The feasibility of the proposed system is also assessed and relevant aspects are highlighted.

- Chapter 7 contains all the conclusions obtained and a brief description of the future work.

Chapter 2

Definition of Concepts

Before presenting the study of the prior art, some concepts will be presented to allow a better understanding of the content of this thesis. The main constraints regarding the work presented in the subsequent chapters are mostly associated data acquisition, data pre-processing, machine learning techniques and evaluation criteria, exposed in the following sections. Section 2.1 describe the waveform signals used, Section 2.2 exposes the differences between two types of filters and Section 2.3 discloses the machine learning models used to classify instances and the evaluation criteria used to assess the performance of the classifications.

2.1 Waveform Signals

Biological signals have been captured and measured in the last centuries for medical purposes, but the possibility of using such signals as unique identifiers introduced a new paradigm. As addressed in the literature, signals like Electroencephalogram (EEG) and Phonocardiogram (PCG) can contain relevant information when considering individuals distinction. Nevertheless, the majority of studies converge to the Electrocardiogram (ECG) and Photoplethysmogram (PPG) waves, thus making it the focus of this thesis.

2.1.1 Electrocardiographic (ECG) Wave

The ECG signal measures the variation of the electrical activity of the heart over time, caused by periodic depolarization and repolarization of atria and ventricle of the heart. An ECG waveform represents several heartbeats. A heartbeat consist of three major segments: P wave, QRS complex and T wave, illustrated in Figure 2.1, which corresponds to different stages of the cardiac cycle, *i.e.* atrial depolarization, ventricular depolarization and ventricular repolarization, respectively.
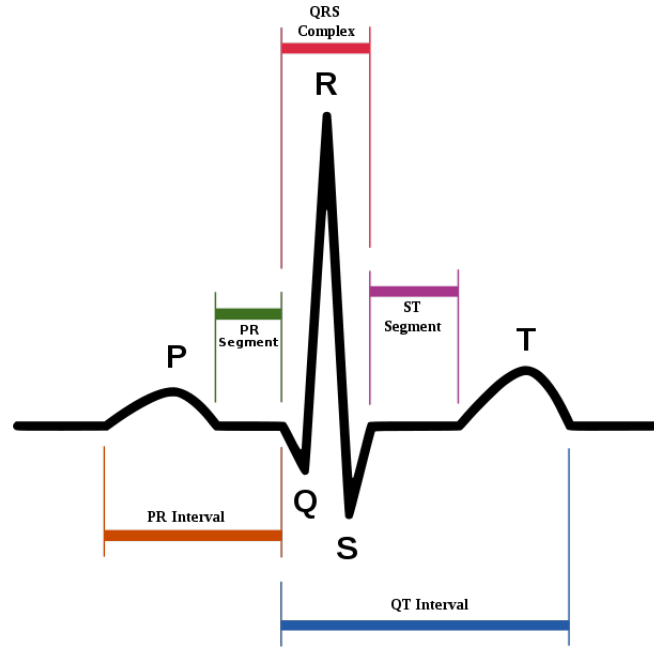
Figure 2.1: Illustration of the ECG waveform traits.[51]

2.1.2   Photoplethysmographic (PPG) Wave

PPG, results from a plethysmogram, which measure the volume changes of the whole body or a specific organ, obtained with an optical sensor. A PPG sensor consists of a light transmitter source, usually a Light Emitting Diode (LED), and a photo resistor (Light Dependent Resistor (LDR)). The sensor is placed directly over the skin and emits a light that penetrates the skin into the blood vessels to be measured by the photocell. As the blood density in the capillaries varies, the light reaching the photocell varies as well, producing resistance changes in the LDR, thereby generating a waveform signal, as illustrated in Fig. 2.2.
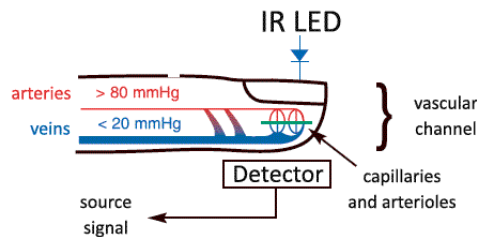
Figure 2.2: Schematic representation of a PPG sensor. [3]

The PPG consists of three major traits: the *systolic peak*, *dicrotic notch* and *diastolic peak* (Fig. 2.3).
During the cardiac cycle there are two main phases related with the PPG traits: the systole and diastole. The prior regards the contraction of the ventricles, when the blood is ejected from the lower chambers to the organs, while the latter consists of the dilation phase, *i.e.* when the

ventricles are relaxing and filling with blood.
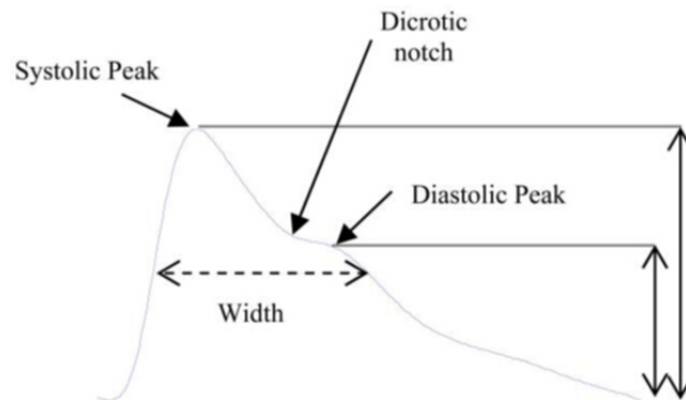
Figure 2.3: Illustration of the PPG waveform major traits. [14]

2.2 Data Pre-Processing

The pre-processing phase can be related to the disposal of unnecessary data that can compromise the correct identification of individuals, or simply the normalization of the data to ensure it fits the same scale and provide a sharper analysis.

When using heartbeats as a biometric feature, the first major step, after obtaining the data in raw format, whether ECG or PPG waveforms, is to ensure that the collected data is mostly associated with the subject and contains few noise attached.

There are different approaches to minimize the impact that noise may have on the waveform, being the main procedures filtering the signal based on the frequency of the noise itself.

Despite varying on the type and the range at which they are applied, filters are essentially focused on attenuating noise artifacts in waves, that may be present in different frequencies, ranging from low to high. Specifically,

High-pass Filters are known to mitigate the interference that low frequency events may introduce in signals, by defining a cutoff frequency, *i.e.* a frequency boundary that allows signals higher than the boundary and attenuates the signal lower than the cutoff value. In Figure 2.4 it is visible the result of applying a high-pass filter to a raw signal (2.4a), which led to a much stable signal (2.4b).

Low-pass Filters are used to remove high frequency artifacts, by also resourcing to a frequency boundary defined by the cutoff frequency. Figure 2.5 illustrates the usage of a low-pass filter to remove noise artifacts. In image 2.5a is visible a PPG signal containing high frequency noise while in image 2.5b the noise was removes using a low-pass Finite Impulse Response (FIR) filter, accomplishing a much smoother signal.

(a) Signal containing high frequency noise and baseline drift.

(b) Signal with baseline drift removed using high-pass filter.
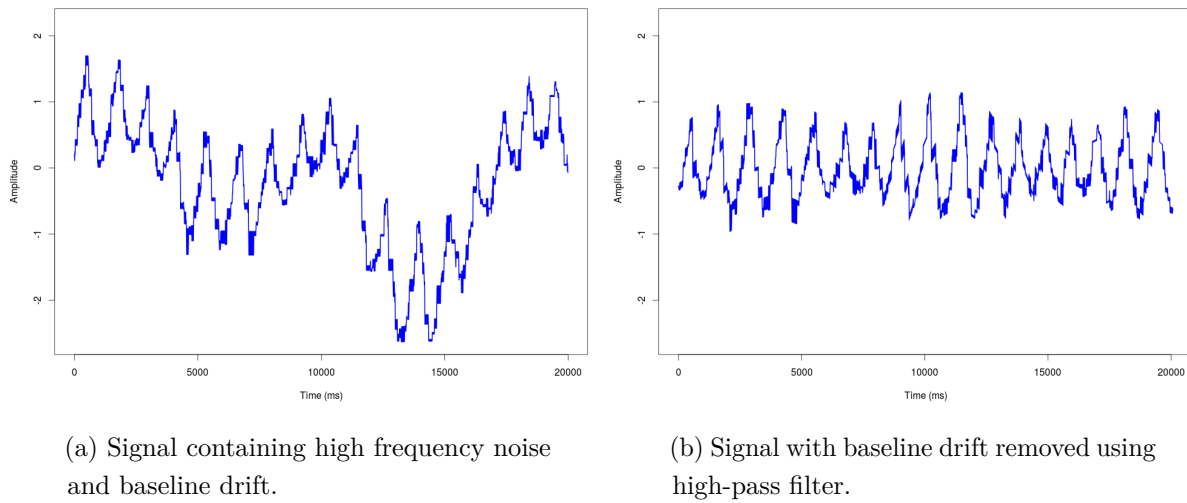
Figure 2.4: Comparison between a raw signal, containing both high frequency noise and baseline drift, and a filtered signal. a) Signal containing high frequency noise. b) Signal filtered using a high-pass filter.

(a) Signal containing high frequency noise.

(b) Signal filtered using a low-pass filter.
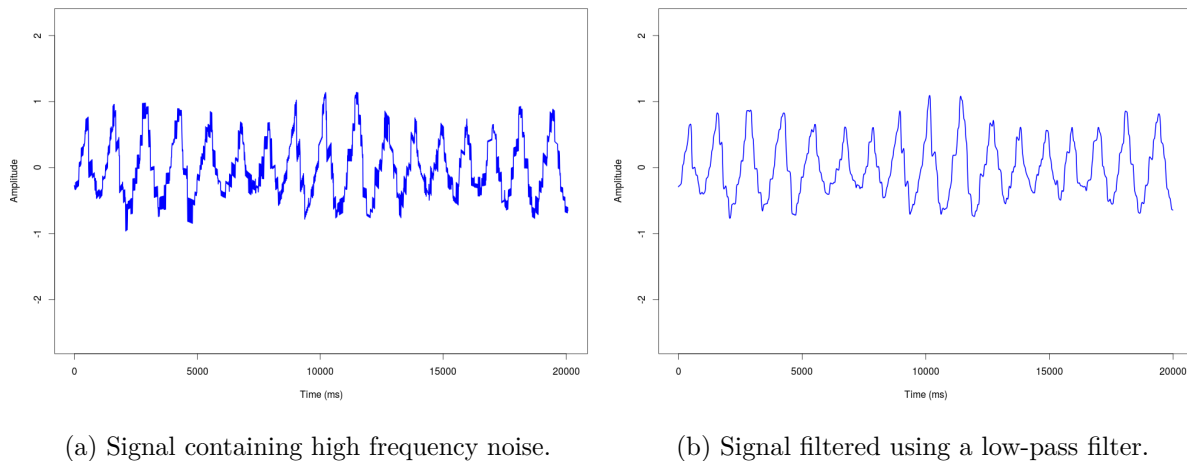
Figure 2.5: Comparison between a raw and filtered signal. (a) Raw signal containing high frequency noise. (b) Signal filtered using a high-pass filter.

Moreover, there are several filters which combine the aforementioned filters by defining an interval of frequencies through which the signal values are submitted and may be kept or discarded, depending if whether it is a band-pass or a band-stop filter.

Resourcing to these filters types, one can mitigate different noise interference such as powerline interference or misplacement of the sensor [1].
When analyzing waveform signals there is also the need to consider baseline drift also known as baseline wander, which consists of fluctuations of the base axis (X-axis) of a signal, typically caused by lower frequencies artifacts, like deep breathing or muscular movement, as illustrated in Figure 2.4a. Different approaches, refereed in the literature, were used to remove such effect,

using a several procedures, such as a wavelet transformation [50], high-pass filters [5], polynomial wave fitting [17] or designing novel filters based on a succession of filtering stages [10].

2.3 Machine Learning

Unlike knowledge and possession based systems, where the process of authentication depends exclusively on a direct match, to a key or token, respectively, in biometric systems there is the need for classifying the biological data to properly verify ones identity, as the biometric traits are continuously changing.
In order to fulfill such requirements the concept of machine learning was inserted in this context.

Machine Learning can be defined as the technique to provide a learning capability to a computer, or a device with computational capabilities, without explicitly programming it, when considering the same type of information, *i.e.* developing algorithms that resources to data input to group information into clusters or infer predictions over that basis, being the prior inserted in the context of *unsupervised learning*, known as *clustering*, and the latter related to *supervised learning*, which will be explored in this thesis. In order to properly differentiate both, we will subsequently described them.

Supervised learning consists of providing the model labeled data as a training factor (*training set*) so it can predict an outcome for future data related with the provided data used as test (*test set*), *i.e.* determine a decision rule to classify, optimally, future unclassified observations by approximating multivariate functions $y = f(x_1, x_2, \ldots, x_n)$ based on a sample $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$.

Unsupervised learning or *clustering* designates the discovery of relationships in data sets in order to generate clusters of information based on the patterns or trends found.

Among several modalities explored in the literature, the models used were Linear Discriminant Analysis (LDA), Support Vector Machine (SVM), k-Nearest Neighbors (KNN), Random Forest (RF) and Naive Bayes (NB), and will be subsequently exposed.

2.3.1 Linear Discriminant Analysis (LDA)

LDA belongs to the most basic type of classifier: linear combination of input variables and it is known for being both a classifier and a dimensionality reduction tool.
It is a statistical classifiers based on Bayes Theorem with the objective of determining the posterior probability of a sample X being classified in a class w_i, $P(w_i|X)$. It is assumed that

data follows a normal distribution and uses Maximum Likelihood Estimator (MLE) to estimate
the mean and co-variance parameters. When the model is created, posterior observations are
classified using the Bayes Minimum Error Rate Classification [27].

2.3.2 k-Nearest Neighbors (KNN)

Firstly introduced in the early 1950s, it is considered a "bittersweet" method due to its intensive
computation requirements when dealing with large data sets. KNN is described as a lazy learner,
since it does not actually learn from the data but simply store the training sets to infer predictions.
Specifically:
We consider each classified observation, from the training set, as a multidimensional vector. In
the training phase, each observation vector and class is stored. Considering x_i as an observation i
with n features, we have the multidimensional vector of x represented as $(x_{i1}, x_{i2}, \ldots, x_{in})$. The
predicted class of a new sample l can be obtained using the euclidean distance:

$$d(x_i, x_l) = \sqrt{(x_{i1} - x_{l1})^2 + (x_{i2} - x_{l2})^2 + \cdots + (x_{in} - x_{ln})^2}$$

Resourcing to the k nearest neighbors, each one with a classification result of the observation,
the majority class is then the final prediction.
Despite achieving good results, the KNN method requires some concerns. The value of k should
be an odd number (3, 5 or 7) in order to avoid draws.

2.3.3 Support Vector Machine (SVM)

Introduced in 1992, they originated a new class of algorithms named *kernel machines* [46]. The
general concept of SVMs is to determine the optimal hyperplane (*decision boundary*) to best
separate the classes, which can be linear or non-linear. Despite existing infinite hyperplanes to
be chosen there is one optimal separating hyperplane, being the one with the maximum margins
to ensure a better classification accuracy on unseen data.

When two classes are not linearly separable, the notion of soft margins appears. It defines
the acceptable error to properly find the solution. It is performed by letting some variables
which do not belong to the same class, stay in the same divided space, as represented in Figure 2.6.

To overcome linear models limitations, a different approach was introduced, where the original
data would be projected into a new space with higher dimension, using a projection function (ϕ),
to allow separating the classes linearly, leading to the concept of kernels.
A Kernel is a function K such that $\forall x, y \in \mathcal{X}$:

$$K(x, y) = \phi(x) \cdot \phi(y)$$
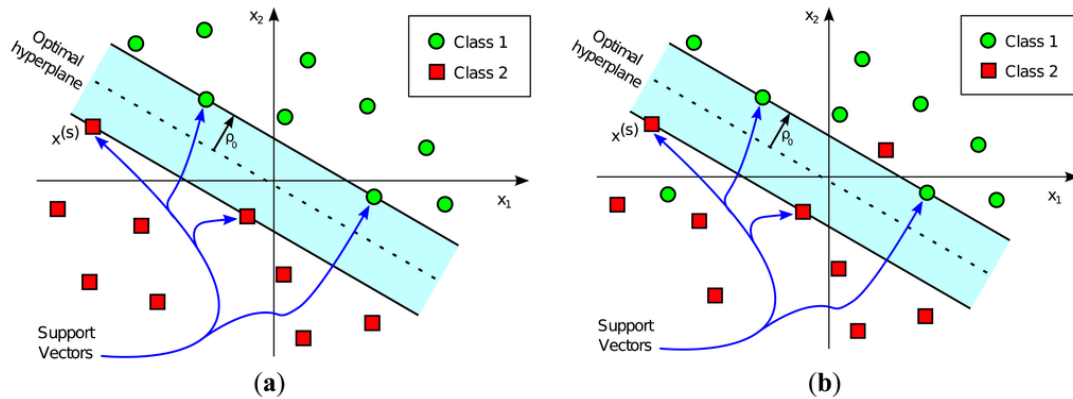
Figure 2.6: SVM classifier representation of linearly and non-linearly separable classes.[40] **a)** The hyperplane separates completely the two classes; **b)** No hyperplane can separate linearly classes 1 and 2.

where $\mathcal{X}$ is the original space and $\mathcal{Y}$ is a new higher dimension space. Some of the most popular kernel functions are:

- Gaussian Kernel

$$K(x_i, x_j) = e^{\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)}$$

- Polynomial Kernel with degree d

$$K(x_i, x_j) = (x_i \cdot x_j)^{\mathsf{d}}$$

- Radial Kernel

$$K(x_i, x_j) = e^{(-\gamma \|x_i - x_j\|^2)}$$

SVM performance and effectiveness lies essentially on the property of kernels, which enables to lower computation requirements by using kernel calculations over dot products on high dimensionality data.

2.3.4 Random Forest (RF)

Random Forest method is based on the concept of *Bootstrap Aggregating (Bagging)*, obtaining several different bootstrap samples from the data used for training purposes, together with random selection of predictors, consisting of a similar concept as Bagging although, instead of varying the samples, the variation lies on the variables .

Consists of sets of uncorrelated decision trees derived from the original data set which is divided in sub-sets where each one generates a different tree, leading to a forest of decision trees.

A *Decision Tree* is then, a tree-based model that propagate branches by dividing the original data set based on the attributes until obtaining a *pure* sub-set, that is, a set with no class variance. The final prediction is based on the vote of each generated tree, and majority class is the predicted outcome.

2.3.5   Naive Bayes (NB)

Inserted in the bayesian classifiers group, Naive Bayes is frequently used due to its competitive performance, when compared with more *sophisticated* methods. It is based on Bayes Theorem and generates probabilistic predictions. It is designated as "naive" since it assumes *class condition independence*, *i.e.* it is assumed that predictors share no dependence relationship among them.

Bayes Theorem   enables to estimate the probability of occurrence of an event by considering conditions possibly related to the event, and it is defined as:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

where

- $P(A|B)$ is the posterior probability of A conditioned on B

- $P(A)$ and $P(B)$ are the prior probabilities of the hypothesis A and B, respectively

- $P(B|A)$ is the likelihood, *i.e.* the probability of the B given A

2.3.6   Evaluation Criteria

The evaluation phase of machine learning regards the process of measuring the performance so the algorithms parameters can be adjusted in order to allow creating a robust and effective solution.

The problem of assessing the quality of the predictions of a classifier is directly related to the data used to evaluate it, *i.e.*, the model can be *overfitted* to the data used in the training phase, and evaluating its performance based on data used to train the model would lead to tendentious results.

The standard procedure to overcome this problem is to use *cross-validation*. The simplest variation, designated as the *holdout method*, consists of using two different data sets, one to train the model and a second to test. In this scenario, the model will be fitted to the first set but evaluated on a *new* group of data, unknown by the model.

Although, using it as an isolated process, can result in misleading evaluations. The division of the data points among the two data sets can introduce variance on the outcome, as the result may be highly dependant on the division performed.

However, using *K-fold cross-validation*, one is able to reduce the influence of how the data is divided. This approach consists of dividing the data into K data sets and perform K iterations of the *holdout method*. The main difference is the usage of several sets to be used as train and test, where in each iteration, one different set is used as test set and the remaining gathered to be used as training set.

Regarding the assessment of the classifiers performance, several different metrics are computed, depending on the application to be given to the classifier. Furthermore, measurements like *accuracy*, *mean absolute error*, among others, typically derived from the confusion matrix, which consists of a matrix where the predicted and the real classes of data are compared, can expose if a classifier is performing good or poorly. When considering binary classification scenarios, the confusion matrix is defined by four conditions, *i.e.* True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN), as illustrated in Figure 2.7.
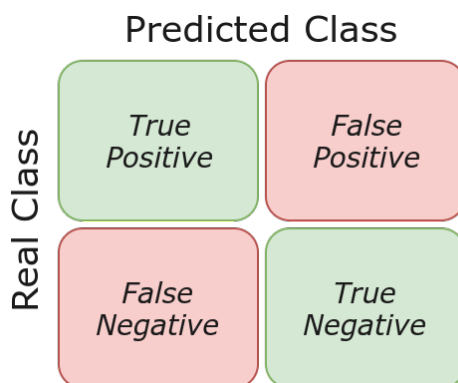
Figure 2.7: Representation of a binary confusion matrix.

A commonly used metric to evaluate the prediction is *Accuracy.*

Accuracy   Consists of a ratio between the correctly predicted values with the total values predicted. Specifically:

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

Moreover, among several metrics based on binary classifiers, the following are specifically relevant when considering authentication scenarios.

Precision   Also known as *positive predicted values*, represents the relationship between the correctly predicted positives with all positive predicted values, *i.e.*:

$$Precision = \frac{TP}{TP+FP}$$

Recall   Like precision, recall uses the correctly predicted positives, however it establishes a relationship with all the real positives. That is:

$$Recall = \frac{TP}{TP+FN}$$

F-Measure   Similarly to *Accuracy*, the *F-Measure*, also known as F_1 *Score*, is a measurement of how accurate the prediction was when comparing to the real values. It takes in consideration both, *Precision* and *Recall* and computes a score to the predictions. It is defined as:

$$F\text{-}Measure = 2 \times \frac{Precision \times Recall}{Precision+Recall}$$

Matthews Correlation Coefficient (MCC)   like the previous metrics, MCC is used to evaluate binary classifications. Resourcing to the confusion matrix components, it computes the correlation coefficient between the predicted and the real value of the classification. Specifically:

$$MCC = \frac{TP \times TN - FP * FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$$

When considering authentication systems, despite evaluating the quality of the predictions, there is also the need to assess the impact that such predictions may have on the system, *i.e.* a threshold has to be defined in order to assure the safety of the system without compromising its usability. For instance, a system may be highly secure, but unusable due to its restrictions in the authentication phase.

To appraise such constraints, The following metrics are computed:

False Positive Rate (FPR)   Also known as False Acceptance Rate (FAR), represents the percentage of misclassified values considered positives, *i.e.* intruders to the system. Is defined as:

$$FPR = \frac{FP}{FP+TN}$$

False Negative Rate (FNR)   Or False Rejection Rate (FRR), consists of the percentage of values misclassified as negatives, *i.e.* real users considered intruders to the system. Specifically:

$$FNR = \frac{FN}{FN+TP}$$

Equal Error Rate (EER)   Like *F-Measure*, EER establishes a relationship between two metrics. It is defined as the rate at which FPR and FNR are equal, *i.e.*, the aforementioned threshold that relates usability with security, as illustrated in Figure 2.8.
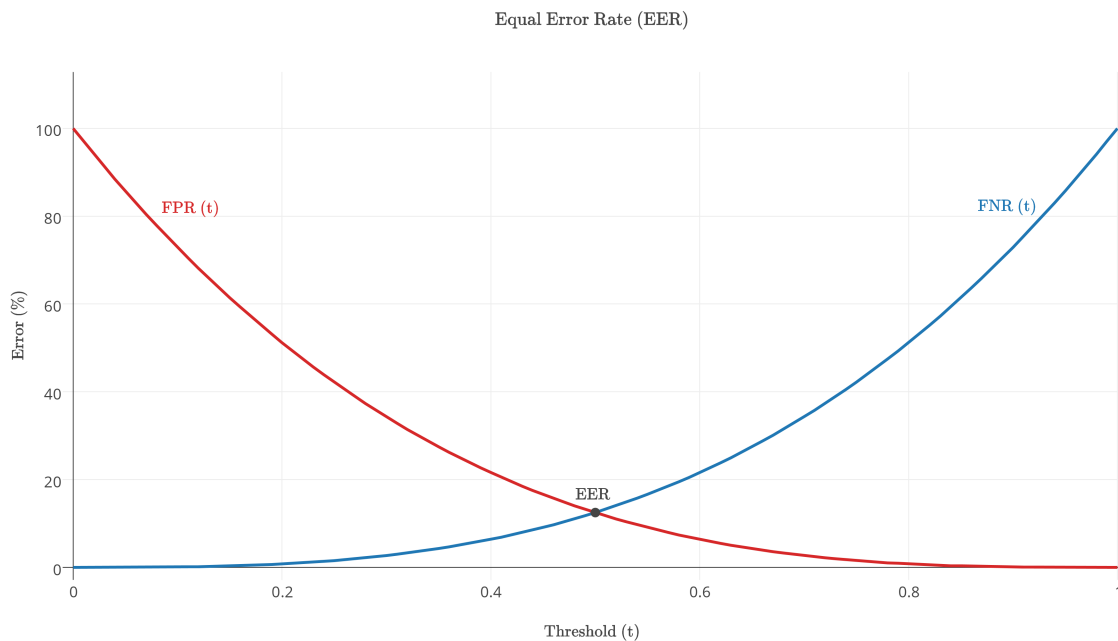
Figure 2.8: Equal Error Rate illustration.

Despite the existence of both sampling and metric evaluation criteria to determine the quality of a classifier, there is also the possibility of assessing the performance of binary classifiers graphically, namely the Receiver Operating Characteristic (ROC), also knows as ROC curve.

ROC curve   Consists of a visual representation of the relationship between True Positive Rate (TPR), also knows as *Recall*, described above, or *sensitivity*, and the FPR, or FAR. The graphical representation illustrates the correspondence between TPR and FPR within a specific decision threshold, *i.e.* the impact that such decision may carry to the classification.

Figure 2.9 illustrates a ROC curve example. The area under the ROC curve, represented with the blue line, designates the Area Under Curve (AUC) which defines how good the relationship between TPR and FPR is, *i.e.*, with a greater AUC value the classifier is more likely to output a correct classification. The more the ROC curve approximates to the top left corner of the plot, the better the classification may be. On the other hand, if the AUC falls below 50%, marked by the lighter line, it means the system is no better than a random classification.

Receiver Operating Characteristic (ROC) Curve
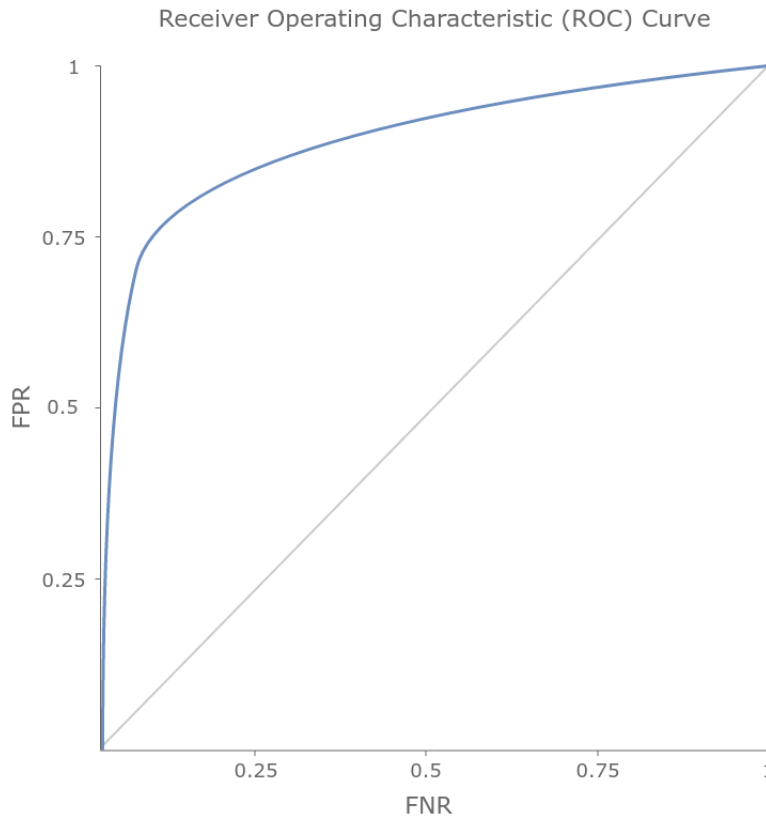
Figure 2.9: ROC curve example.

The high diversity of evaluation metrics is mostly due to some inherent flaws when considering different types of data. Despite the *Accuracy* being good to determine how correct the classifications are, it is highly affected by class distribution of the data, *i.e.*, if a data set is not balanced the weight that class differences may have on the accuracy will increase as the differences between classes aggravate.

As an example, if we consider the case of a two classes, A and B, having 10 and 200 samples each, respectively, if class A has no TP values, but class B has no TN, that is, every instance of A is misclassified and every instance of B is correctly classified, the accuracy would be of, approximately , 95%. In the case of an authentication system, it would be impracticable since no user would be allowed to access the system.

This fact leads to the choice of other metrics, that may better reflect classification performance when dealing with unbalanced data sets, such as the *F-Measure* and *MCC*.

Chapter 3

State of the Art

Every authentication system based on biometric features is essentially a problem of pattern recognition, which can be summarized in three major steps, *i.e. pre-processing, feature extraction classification* [1]. As referred in Section 2.2, several approaches were performed to reduce the interference that external artifacts may produce on the waveform signals. When considering the classification stage, different procedures were chosen, either using machine learning algorithms, some of which exposed in Section 2.3, or using template matching by distance measurements with threshold definition. Regarding the feature extraction phase, several approaches have been attempted and proven to work. Moreover, they can easily be categorized into *fiducial based approaches* and *fiducial independent approaches*, which will be further described in this chapter.

3.1   Fiducial Based Approaches

A fiducial point consists of a mark used as reference for measurements. Hence, fiducial dependent features consists of elements extracted from the fiducials marked in the waveforms, which can be time or amplitude dependent, also referred as analytic features [26]. As an analogy, fiducials can be seen as the minutiae in fingerprints.

The Electrocardiogram (ECG) wave is mostly characterized by the P and T waves, and the QRS complex, thus being prioritized when selecting fiducials, while the Photoplethysmogram (PPG) wave consists of three major characteristics, the systolic and diastolic peaks, and the dicrotic notch, fact that leads to the usage of the first and second derivatives of the wave for feature extraction.

Biel *et al.* experimented in [4] the possibility of human identification using the heartbeat as biometric feature. Using specific hardware to collect the ECG data using 12 leads, they extracted 50 samples from a group of 20 individuals. For each lead, 30 features were obtained, creating a set of 360 features per subject. By calculating the correlation matrix they were able to notice a strong correlation between different leads, for a specific feature, enabling a feature reduction by reducing the number of leads considered. Calculating, again, the correlation matrix

and discarding the ones with high correlation value, they reduced the amount of features to 12. Using Soft Independent Modeling of Class Analogy (SIMCA) as the classification method, they performed 8 different tests with different amount of features, ranging from 7 to 360, naming the experiments after the alphabet, from A to H, and successfully classified 49 out of 50 samples in almost every test. Their best result was test G, where they fully classified the sample, which contained 10 features. They concluded to be possible to identify individuals from a predetermined group using the heartbeat, extracting the ECG data using a single lead.

Shen *et al.* using a one-lead ECG extraction, proved in [42] that an identification system can be created using heartbeat as a biometric feature. Using the QRS complex and normalizing the QT interval to extract features, they were able to successfully identify individuals within their study group with 95% and 80% rates, using two different classification methods, Template Matching and Decision Based Neural-Network (DBNN), respectively. In [43] they again proved the viability of building an identification system using the heartbeat as biometric trait. In this approach, they used a single lead to extract ECG readings and added an additional step, after pre-processing the wave, called *pre-screening* phase, where they would filter the samples to classify using template matching. They obtained 17 features, mostly extracted form the QRS complex, since "this waveform is most easily recognized, easy to detect, essential for life and stable with different heart rates". They shown promising results, obtaining identification rates of 100% in a predetermined group with 10, 20 and 50 individuals, 96% and 95.3% when the number of subjects were 100 and 168, respectively. When comparing the standard three-step procedure with the proposed model, they highlight the fact that with their model, there is need for training process.

Gu *et al.* introduced in [22] the possibility of using PPG for human identification. Their data set consisted of a one minute sample per subject at a rate of 1 KHz. The raw data was first pre-processed using a smoothing technique and then submitted to the feature extraction and classification phases. A feature vector was generated using part of the sample, while the remaining was used as test set. As features, they used:

- Number of peaks in each pulse

- Upward slope between the bottom of each waveform and the first peak

- Downward slope between the last peak of each waveform and the bottom

- The time interval between the bottom point and the first peak point

In order to determine the discriminant potential of the chosen features and the weights to assign to each feature, they used the following equation

$$F_j = \frac{Interclass\ Variability}{Intraclass\ Variablility}$$

which allows to correlate inter and intra classes. Euclidean distance between samples and templates was used to obtain the final decision, classifying the sample as the subject with the minimum distance to the existing templates. The authors obtained an accuracy rate of 94% and claimed the faulty verification was due to the bad quality of the signal.

Israel *et al.* proved in [26] to be possible identifying individuals resourcing to fiducial points, invariant to the state of anxiety of individuals, as biometric features. Using P, T and R which consists of the local maxima and the heartbeat peak, respectively, as fiducials, while also obtaining P', T' and L' being the end points of the P, T and the beginning of the P wave, respectively, they were able to extract 15 features and selecting 12, using a feature selection approach based on a stepwise canonical correlation analysis. Using Linear Discriminant Analysis (LDA) to classify the heartbeats and a standard majority voting to associate heartbeats with individuals, they correctly classified 82% of the heartbeats and 100% of the individuals when using the data collected from the neck, and 79% of the heartbeats and 100% of individuals from the ECG obtain from the chest.

Wübbeler *et al.* tried to prove in [53] the possibility of using the heartbeat to verify and identify humans with a three lead ECG extraction, using the *Einthoven triangle scheme* [35]. Using a public data set Physikalisch-Technische Bundesanstalt (PTB) which consists of more than 27000 ECG records from 74 individuals, they used a bi-dimensional vector (H_x, H_y) as feature vector, where H_x and H_y correspond respectively to the main lead, and the remaining two combined. The data contained in the feature vector was obtained by establishing an interval window of 100 ms, for all the leads, centered in the occurrence of QRS complex.
Using a standard Nearest Neighbors (NN) and threshold schemes they obtained identification rates varying between 98,1% and 99% when using different thresholds, while the Equal Error Rate (EER) varied between 0.2% and 2.5% when considering different False Matching Rate (FMR) and False Non-Matching Rate (FNMR) values, 10% and 3%, respectively.

Yao *et al.* used an approach based on derivatives of the PPG signal for both, identification and subject discrimination, as referred in [54]. The data was collected from 3 subjects using a pulse oximeter sensor with a sample rate of 300 Hz and duration of 70 seconds, further divided into 3 groups, resulting in 9 data sets. The full process of generating the data sets consisted of 9 steps, in which they filtered the raw PPG data using a Chebyshev low-pass filter to remove noise, randomly choosing a complete pulse, fitting the pulse with a 10th order polynomial to obtain the 1st and 2nd derivatives, in order to extract, respectively, the number of maximum and minimum points and inflection points, and repeating the hole process several times to obtain the features to store. The authors concluded that by using high-order derivatives it is possible to obtain more discriminative attributes however they become more sensitive to noise. On the other hand, using lower order derivatives, features become more robust yet less sensitive implying the need for weights assignment when considering such characteristics.

Wei *et al.* proposed a new algorithm in [50] based on wavelet transformation to eliminate the baseline drift while using conventional methods to remove noise, such as median and Finite

Impulse Response (FIR) filtering, and extracting features resourcing to an improved differential algorithm on PPG signals. The signal processing consisted of three main steps: outliers removal, FIR low-pass filtering and baseline drift elimination, using the proposed algorithm. The authors used a Hamming window of 21 orders to remove high-frequency noise, with sampling frequency of 500 Hz and cut-off frequency set to 30 Hz. The main procedure of the algorithm was to apply a series of low-pass and high-pass filters to the signal, applying a wavelet reconstruction to recompose the wave. To extract the features, they used a differential threshold method consisting of a cubic spline interpolation, an enhanced differentiation approach and outliers removal. The results obtained using the traditional differentiation approach resulted in a 4.1% error rate mitigated by using the improved differential method.

Spachos *et al.* explored the feasibility of using PPG signals as biometric traits of individuals [47]. They obtained the PPG signal from the fingertips of 29 healthy subjects. In the pre-processing phase they applied a sequence of four steps, consisting of peak detection, to locate maximum and minimum peaks, segmentation, normalization and time domain scaling, setting the value to 200 samples to compensate for the variations of the PPG cycle due to the heart rate variability. To collect features from the signal, they used LDA as it is an efficient supervised learning method for feature extraction and dimensionality reduction. The data set was divided in two, using the first half of the recordings as training set and the remaining as test set. Classification was performed using NN and *majority voting* while also using a threshold to accept or reject the input for the identified subject. They tested the proposed method with two different data sets, **OpenSignal PPG Dataset** consisting of 14 healthy subjects and **BioSec PPG Dataset** composed by 15 healthy subjects obtaining 0.5% False Acceptance Rate (FAR) and False Rejection Rate (FRR) with a properly set threshold for the prior, and an EER of 25% for the latter, leading to the conclusion that it is possible to use PPG signals in biometric authentication systems despite being highly affected by the recording sensor and the environment.

Bonissi *et al.* studied in [5] the possibility of continuous authentication techniques using PPG signals, presenting a method based on a correlation and template matching approach. Following the standard procedure, they started by pre-processing the signal, extracting features and proceeding to the classification/verification phase. To pre-process the signal they applied a third order high-pass *Butterworth* filter with a cut-off frequency of 0.1 Hz to normalize the baseline. The feature extraction phase consists of generating templates **T** with a variable and distinct number of heartbeats among them using the following procedure:

- Signal segmentation to generate a matrix **H** containing the relevant points

- Applying Pan-Tompkins algorithm for peak detection

- Compute the average number of heartbeats **S'** per entry in **H**

- Calculate the maximum cross-correlation **C** between every heartbeat stored in the **H** and average heartbeat **S'**

- Store a number **N** of heartbeats corresponding to the maximum number of **C** values in the template **T**

- Discard the signal if the correlation value is lower than the empirically estimated threshold.

The aforementioned procedure is performed iteratively over all **S** generated signals and executed twice to detect and remove noise. Regarding the classification phase, they calculate a similarity score between two templates, generate a matrix **M** by calculating the maximum cross-correlation between every signal belonging to both templates and compute matching scores considering different approaches over the similarity values of **M**, specifically `mean(M)`, `median(M)`, `75%(M)`, `90%(M)`, `95%(M)` and the `max(M)`. The collected data consisted of a 2 minutes long PPG signal for each of the 44 subjects, collected at a rate of 75 Hz. They generated three different data sets varying the sample duration, **DB20** with 242 samples, **DB30** with 158 samples and **DB40** with of 116 samples, with durations of 20, 30 and 40 seconds respectively. Among all the tests they were able to obtain an EER of 3.34% in a regular authentication environment. When testing for continuous authentication scenarios, the best EER obtained was 13.47%. The authors concluded that in a continuous authentication scenarios, the variability of the biometric traits with time may influence the performance of the system while the duration of the collected sample to be used influences the EER obtained in the system.

Carreiras *et al.* designed a model in [8] for a biometric system using the heartbeat. ECG readings were obtained using 12 leads, however they used a single lead, obtained from the subject's fingers, to perform authentication. Using a template matching approach, they firstly processed the ECG data, extracted the QRS complex, removed outliers, obtained a feature set and finally stored the generated template. Using an outlier detection approach, referred as *DMEAN* [30], they calculate the distances between the samples and the templates mean value. In case the results overcome an adaptive threshold, they are considered outliers. Classification is performed using k-Nearest Neighbors (KNN) method, with $k = 3$ as it is known to be lightweight and effective. Results were obtained by applying the model to two different data sets, containing 618 and 63 subjects, which they named **P618** and **Baseline**, respectively. By calculating FAR, FRR using different ratios between False Positive (FP), False Negative (FN), True Positive (TP) and True Negative (TN), they were able to determine the EER. Based on the EER of 9.01% and 13.26%, and Error of Identification (EID) of 15.64% and 36,40% for **P618** and **Baseline** data sets, respectively, they concluded to be possible to use the heartbeat as an authentication factor, since the system performance is not affected by the amount of subjects, when comparing to an identification system.

Akhter *et al.* studied in [2] the possibility of using Heart Rate Variability (HRV) property of the heart for authentication purposes based on PPG signals. The database used consisted of 2430 R-R Interval (RRI) sequences, 10 samples for each of the 81 subjects, using 5 to generate the templates and the remaining 5 for testing. Seven time-domain features were extracted: Root Mean Square of Successive Differences (RMSSD), Mean Heart Rate (MeanHR), Mean, Median,

Standard Deviation of Heart Rate (SDHR) and the Maximum and Minimum Interval duration in a particular RRI. Using the Euclidean distance to classify the samples they obtained an EER of 17% when considering a threshold of 0.07, obtaining a recognition rate of 86.7%.

3.2   Fiducial Independent Approaches

Despite being possible to correctly identify individuals using fiducials, such approach is still highly dependent on both, signal synchronization and heart rate variability, since the waveform signal is not periodic, but highly repetitive [9]. In ECG signals, despite the P wave and QRS complex remaining almost invariant during heart rate variations, the T wave is still affected implying some constraints when it is used in fiducial approaches [15]. Since bio-signals are representations of physiological events, the PPG waveform is also vulnerable to the aforementioned constraints when collecting fiducial information.
In order to bypass such constraints, a different procedure was introduced, where the features would not be extracted directly from time and frequency domains but obtained holistically from the morphology of the entire signal or isolated heartbeats.

Plataniotis *et al.* proved in [38] to be able to successfully identify individuals using a fiducial independent approach. They firstly produced a hybrid framework that combined appearance-based and analytic attributes and, by consecutively applying wave analysis methods, they obtained a set of features that allowed individuals distinction. Posteriorly, to discard the need for appearance-based features, they used Discrete Cosine Transformation (DCT), for dimensionality reduction, after obtaining the Auto Correlation (AC) coefficients, and were able to obtain high rates, some of these reaching 100%. Despite their data set containing few individuals, fourteen, each record had 100 seconds duration, representing a high amount of heartbeats. In order to evaluate the algorithm's performance, they used both, Normalized Euclidean Distance and Normalized Gaussian log likelihood.

Wang *et al.* compared both approaches in [49]. Fiducial features were extracted from P, Q, R, S, T, P' and T', the last two being the endpoints of the P and T waves, respectively. Also, the samples were normalized using P'T' distance to lower the dependence of the obtained features from the heart rate variability. Analytical features were extracted using the AC/DCT method introduced in [38]. The pre-processed signal was segmented in several windows using a *windowing* technique, with the single constraint that each window only needs to be larger than the average length of a heartbeat wave, so it can contain multiple beats. The several windows were then submitted though the AC/DCT process, where for each window, it is estimated the auto-correlation values, applied the DCT and finally, after discarding the less significant values, the remainder is sent for classification. They successfully identify a large amount of subjects using both approaches, in different data sets. Using analytic features, they identified 84.61% and 100% individuals in PTB and MIT-BIH Arrhythmia Database (MIT-BIH) respectively. Regarding

appearance-based features, they obtained 100% and 94.88% of successfully identified individuals when using the PTB and MIT-BIH data sets, respectively.

As a final note, the authors claimed that despite being two different approaches, they complement itself in ECG data analysis.

Fatemian *et al.* demonstrated in [16] the feasibility of human identification using the heartbeat. Their approach consisted of using two different biometrics related to the heartbeat, obtaining both using ECG and Phonocardiogram (PCG). By applying Discrete Wavelet Transform (DWT) to reduce noise and depict the heartbeat on ECG readings, they obtained a set of features, for each biometric sample. Using a threshold approach, and setting its values to 0.85 and 0.98, they obtained recognition rates of 95.37% and 90.6%, respectively.

Chapter 4

Methodology

As previously mentioned, the process of extracting information from the heart is highly dependent on the heart rate variability, health conditions, and the waveform itself. Synchronizing heart beat waveform signals is a time consuming process which may interfere with the usability of the proposed system. Furthermore, due to sensor properties and limitations, which will be further exposed in Section 4.6, using fiducial features would not be feasible. Thus we have used the *AC/DCT* (Auto Correlation (AC) / Discrete Cosine Transformation (DCT)) method, allowing a higher level of abstraction from the fiducial points of the wave, as exposed in Chapter 3.

In order to test the feasibility of the system three major data set groups were created, two of them considering distinct public databases, *PhysioNet* [21] and *CapnoBase* [28], and a third one generated from the sensor used in the system.

Since the proposed system fits into an authentication scheme, two different classes were used to distinguish individuals, where the users id would be 0 (zero) or 1 corresponding to the designations of "Owner" and "Intruder", respectively. To ease the referencing process, *PhysioNet* and *CapnoBase* public databases and respective data sets will be referenced as **ECGID** and **TBME**, correspondingly. The data set collected from the Photoplethysmogram (PPG) sensor will be referred as **Angel Sensor**.

4.1 Development Environments

Before implementing the system, the chosen methodology and associated data was assessed to determine the feasibility of the proposed implementation. Due to the different origins of the public databases, different environments were required for acquiring the data from both the servers, and the files containing the waveform signals. Moreover, different environments were used to test and assess the quality of both, the data sets and the machine learning models.

PhysioToolkit   A software collection provided by *PhysioNet* that allow users to view, analyze, simulate and extract signals directly from their platform. *WFDB Software Package* was used to retrieve the data from *PhysioBank* [21].

Octave   A "A high-level interactive language for numerical computations" [13]. Was used to extract the Electrocardiogram (ECG) and PPG records from *CapnoBase* database, downloaded from the website [28].

Angel Sensor API   Consists of an Application Programming Interface (API) integrating the *Bluetooth Generic Attributes (GATT) Service* with *Angel Sensor* wristband to provide access to the sensor services and data [31].

Android API   Used in both, the concept evaluation and system implementation stages. The prior consisted of a simple application to retrieve bio-signals from individuals to generate the data sets, and the latter, the expansion of the application into the proposed authentication system.

R   A "free software environment for statistical computing and graphics" [39], used to pre-process the wave signals, extract features and generate the data sets, and to test individual models.

WEKA   A machine learning software based on *Java* programming language [23], used to obtain the metrics and evaluate thoroughly the machine learning models chosen.

4.2   Databases Description

This section comprises all the information regarding the data used, specifically ECG and PPG signals contained in the *Physionet* and *CapnoBase* public databases, and the signals collected using the *Angel Sensor* wristband.

4.2.1   PhysioNet ECGID database

PhysioNet's *ECG-ID* database consists of 310 ECG lead I records of 90 distinct healthy individuals. The records were obtained from 44 male and 46 female volunteers aged from 13 to 75 years old. The amount of records per subject varies from a minimum of 2 up to a maximum of 22. Each individual record consisted of 2 groups of data, ECG I and ECG I filtered, where the prior consists of a raw waveform signal, and the latter being the result of filtering the raw signal to attenuate noise artifacts.

4.2.2 CapnoBase TBME database

CapnoBase TBME database consists of ECG and PPG raw signals of 42 individuals with 8 minutes duration per record, with a sampling rate of 300 Hz. The samples were collected from individuals with age ranging from 1 to 76 years old.

4.2.3 Angel Sensor records

Angel Sensor data set was created using the PPG signal obtain from *Angel Sensor* wristband at a rate of 100 Hz. Consists of 342 samples with duration of 25 seconds, collected from 15 volunteers with age ranging from 20 to 50 years old. Each sample consisted of two waveform signals obtained from different Light Emitting Diode (LED) that compose the device sensor. The number of records per individual varies between 3 and 30, with the exception of the user chosen to be the *Owner* of the system, from whom were collected 92 samples (to attenuate the differences between classes).

4.3 Pre-processing

Before extracting features from the individuals signals, some pre-processing was performed in order to handle the data correctly, minimizing the impact of misleading information intrinsic to the data itself, such as noise artifacts or unscaled data. Furthermore, the pre-processing to which signals were submitted can be resumed into the following steps:

4.3.1 Signal normalization

In order to correctly compare data, it should be set into a comparable scale to avoid inaccurate conclusions, specifically, the **Angel Sensor** data, where each wave signal would be outputted in a different range of values, as illustrated in Figure 4.1a.
Using the *Z-Score*, also known as *Standard Score*, defined by:

$$Z = \frac{X - \mu}{\sigma}$$

where $\mu = mean$ and $\sigma = standard\ deviation$, one is able to rescale the samples to ease the process of comparison, as visible in Figure 4.1b.
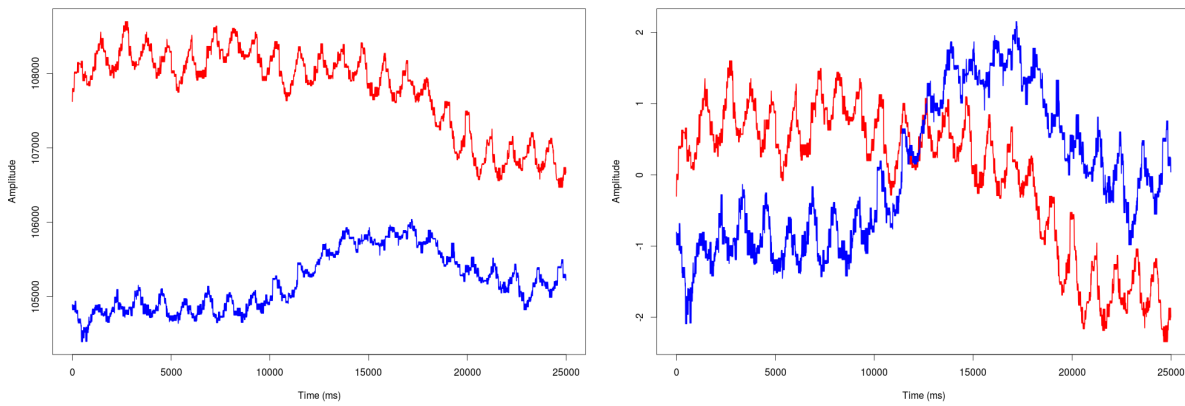
(a) Two different unscaled PPG waves recorded from angel sensor.

(b) Two different PPG waves recorded from angel sensor, scaled using Z-Score.

Figure 4.1: Comparison between two unscaled (a) and scaled (b) PPG wave signals.

By applying the aforementioned method, each signal S was then normalized as follows:

$$Z(x_i) = \frac{x_i - \mu}{\sigma}$$

$\forall x_i \in S, 1 < i < n$ where μ and σ are the *mean* and *standard deviation* of S, and x_i is the ith data point of the signal S of length n.

Among every group of data sets, only the **ECGID** data sets were not submitted to this pre-processing stage, since they were already obtained in a matchable scale.

4.3.2   Signal Filtering and Reconditioning

Regarding the signal filtering stage, the approach was to use a a succession of Finite Impulse Response (FIR) filters in order to adjust and clean the wave samples for feature extraction, using the R functions `fir1` and `signal` from the signal package [44].

Signals obtained from the *Angel Sensor* wristband were submitted to high-pass and low-pass FIR filters, applied subsequently, to remove baseline drift and reduce noise artifacts caused by motion and hyper-sensitivity of the sensor.
Several combinations of both, high-pass and low-pass FIR filter orders were tested based on the maximum cut to be performed on the signal, as exposed in Table 4.1.
After filtering the signals, they were reconditioned to remove the delay caused by the filter application, resulting in a shorter sample. The amount of data points associated with the delay to remove is defined by the following relationship between both high-pass and low-pass filter orders:

$$Maximum\_Cut = Highpass\_Filter\_Order + Lowpass\_Filter\_Order$$

Maximum Cut	High-pass Filter Order	Low-pass Filter Order	Signal duration (milliseconds)
300	294	6	2200
300	292	8	2200
300	290	10	2200
300	288	12	2200
400	394	6	2100
400	392	8	2100
400	390	10	2100
400	388	12	2100
500	494	6	2000
500	492	8	2000
500	490	10	2000
500	488	12	2000

Table 4.1: Relationship between the maximum cut of the signal with the orders of high-pass and how-pass FIR filters.

Signals from **ECGID** were not submitted to this stage since they were previously filtered when the database was created.

4.3.3 Signal Derivatives Computation

Waveform signals may contain many discriminative information regarding the person related to the signal. Although, the condition of the signal may influence the information one can extract from it. Specifically, when testing **Angel Sensor** data, due to the poor quality of the signal, exposed in Section 4.6, the first and second derivatives of the filtered signal of each LED were computed, resourcing to the *R* functions sm.spline [36] and predict [39].

4.3.4 Signal Segmentation

In order to decide the wave length of the signals to be used in the system, each original record from the public databases was segmented into records with varying lengths stored individually to subsequently generate different data sets. Records from **ECGID** and **Angel Sensor** were divided into samples of 5, 10 and 20 seconds, and **TBME** records into 5, 10, 20, 30 and 40 seconds.

4.4   Feature Extraction and Data Set Generation

After the pre-processing stage, the AC/DCT method was applied to every group of signals, defined by the duration of its samples, to extract features. The following content exposes the process of feature vector generation, dimensionality reduction and the final data sets created to proceed to classification.

4.4.1   Feature Vector Generation

In order to extract features from the pre-processed waveform signals, the AC coefficients of each record were computed, using `acf` [39] function, generating an initial feature vector. Figure 4.2 illustrate the AC coefficients of a signal.
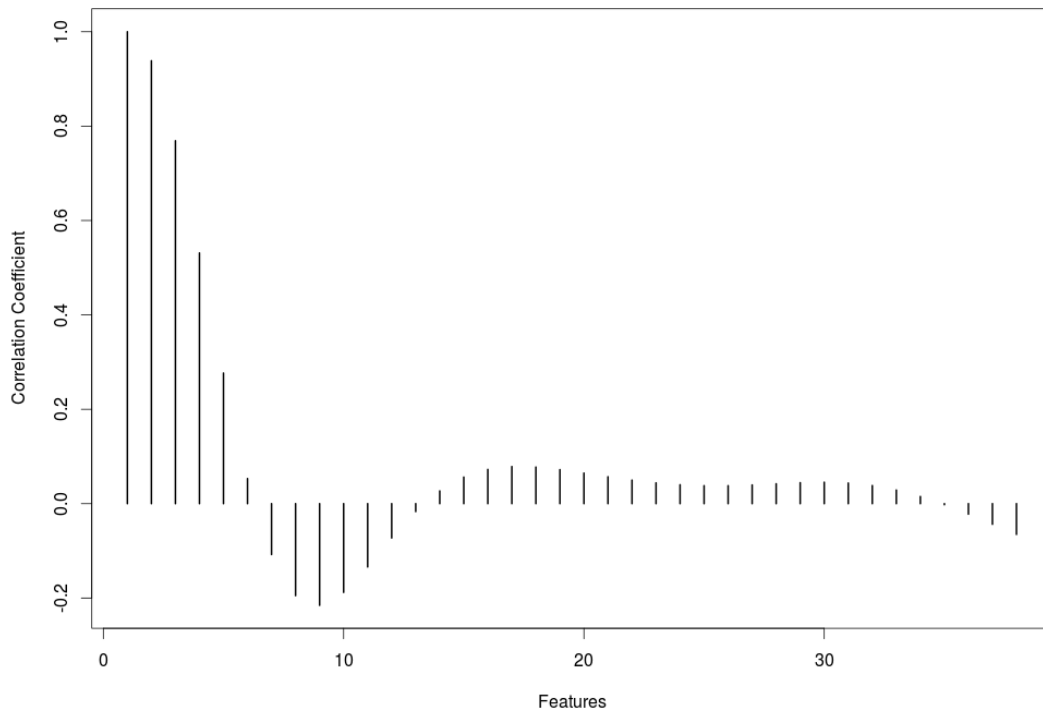
Figure 4.2: Histogram of the AC coefficients.

4.4.2   Dimensionality Reduction

As referred in Chapter 3, The the second phase of the AC/DCT method consists of computing the DCT, using `dct` function from dtt [29] package, of the AC coefficients, which enables to compress information and reduce dimensionality of the final set. The first K values were kept

and considered the minimum amount of features the set should hold, due to the information held
by each value regarding the AC coefficients, as illustrated in the histogram in Figure 4.3.
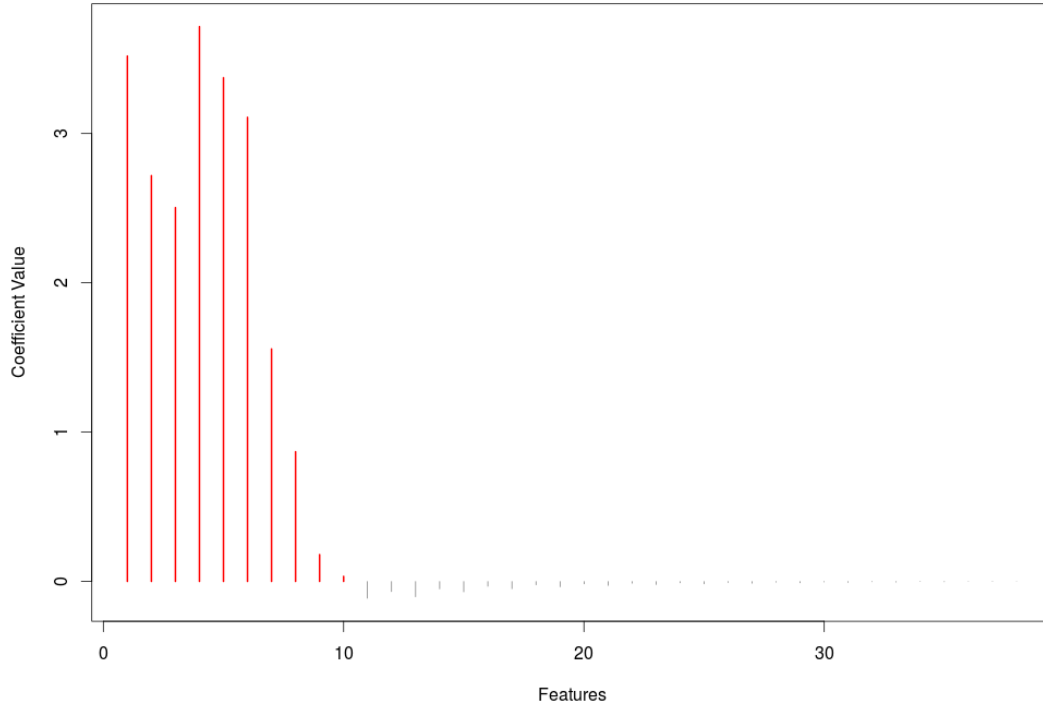
Figure 4.3: Histogram of the AC/DCT coefficients. The highlighted values represent the k
features to be considered.

4.4.3 Data Set Generation

To correctly evaluate the quality of the feature vector, several data sets were created by combining
the generated feature vectors with a user label. For every group of signals, several data sets,
with a varying numbers of features, were generated and labeled, as illustrated in Tables 4.2, 4.3,
4.4 and 4.5.

Dataset	Signal Duration (seconds)	Number of Records	Number of Features
ECG_ID_5	5	1240	10
ECG_ID_{10}	10	620	10
ECG_ID_{20}	20	310	10

Table 4.2: Data sets generated from the PhysioNet **ECGID** database.

Dataset	Signal Duration (seconds)	Number of Records	Number of Features (f)
ECG_TB_5	5	4032	10
ECG_TB_{10}	10	2016	10
ECG_TB_{20}	20	1008	10
ECG_TB_{30}	30	672	10
ECG_TB_{40}	40	504	10

Table 4.3: Data sets generated from the ECG signals of CapnoBase **TBME** database.

Dataset	Signal Duration (seconds)	Number of Records	Number of Features (f)
PPG_TB_5	5	4032	5
PPG_TB_{10}	10	2016	5
PPG_TB_{20}	20	1008	5
PPG_TB_{30}	30	672	5
PPG_TB_{40}	40	504	5

Table 4.4: Data sets generated from the PPG signals of CapnoBase **TBME** database.

Dataset	Signal Duration (seconds)	Number of Records	Number of Features (f) per wave signal
AS_5	5	1300	$3 \leq f \leq 6$
AS_{10}	10	650	$3 \leq f \leq 6$
AS_{20}	20	325	$3 \leq f \leq 6$

Table 4.5: Data sets generated using the collected data from the Angel Sensor wristband. Each data set consists of 2 LED signals and, the first and second derivatives of each LED, making a total of 6 wave signals.

4.5   Machine Learning

In order to determine the best classifier to be used in the authentication system, several models were tested to evaluate its performance and results. Cross-validation *holdout* and *k-folds* techniques were performed, on *R* and *WEKA* environments, respectively.

To choose the better parameters to use in both k-Nearest Neighbors (KNN) and Support Vector Machine (SVM) models, different values for the most significant arguments were tested, *i.e.*:

- KNN - The model was tested using three different values, 1, 3 and 5, for the number of neighbors k, and the distance between samples was computed using the *euclidean distance*.

- SVM - The major alterations were focused on the *Cost* (C) for misclassified instances, being tested with the values 1, 10, 100 and 1000.

4.6 Approaches and Difficulties

The objective of following several approaches was to obtain a deeper knowledge on how the selected approach would be reflected in the outcome when considering different data, specifically, ECG and PPG waveform signals. Despite the validation of the AC/DCT in the literature, it considered ECG data only. Among the several strategies to assess PPG signals biometric potential, the focus was essentially over fiducial features extraction. Since the main objective of this thesis was to assess the feasibility of extracting biometric features from the heartbeat in mobile devices, the usability of the method was considered important. Furthermore, facts like heart rate variability and signal synchronization, issues encountered when considering a fiducial approach, would highly affect the performance of the system.

The following content exposes all the tested approaches and respective reasons for its failure or exclusion from the methodology.

4.6.1 APIs Limitations

After assessing the possibility of using the PPG wave to extract biometric features, some hardware was tested to extract the data, namely *Zeaplus K18 Smartwatch* (referred as **K18**) and *Huawei Smartwatch* (referred as **Huawei**), running Android Operating System (OS) version 4.4 and Android Wear version 1.5, respectively.

The main issues were related to the Android API [24] limitation, *i.e.* each Android OS version is associated with an API level that allows accessing devices features such as accessing to the device location. In the case of **K18**, the API associated is API level 19 which had no specifications to access the heart rate sensor, only available from API level 20 and higher.

As a result, **Huawei** was the second device chosen for the experiment. Since the API for the Android Wear devices already comprised tools to access the heart rate sensor, the access to its data was possible. However, the only data made available by the API was the heart rate in Beats Per Minute (BPM), not the PPG wave itself. Despite the existence of different APIs released for Android devices for fitness purposes, namely the Google Fit APIs [25], it still did not grant access to the PPG waveform signal, making the usage of the device unfeasible.

Finally, a developer oriented device was selected, the *Angel Sensor* wristband, which enabled accessing the data in raw format to proceed with the experiment.

4.6.2 Angel Sensor Limitations

When the data extraction stage started, a different problem was observed, derived from the quality of the hardware. Due to the low sampling rate offered by the sensor, the signal was obtained with low resolution, leading to some constraints regarding the information extracted from the signal.

Moreover, the scarce documentation available further exposed the implementation to hardware

related issues. Specifically, since the communication between the wristband and the device is performed using Bluetooth protocol, the acquired data is extremely dependant on the services provided. As an example, despite the Bluetooth GATT specifications comprising a channel to send R-R Interval (RRI) information, such data was not available on the wristband. Furthermore, if the wristband detects a certain amount of movement, some services are disabled from the hardware, as the case of obtaining the heart rate.

4.6.3   Fiducial Dependent Approach

As aforementioned, the data derived from the angel sensor wristband was delivered at a low frequency leading to a poor definition of the waveform signal. In order to overcome such constraint while experimenting a fiducial dependent model, an approach, exposed in [2], based on time-domain features was tested. It consisted of extracting information from RRI and defining the following components as features:

- **Root Mean Square of Successive Differences (RMSSD)** - Typically used as a measure of Heart Rate Variability (HRV), is defined as:

$$\text{RMSSD} = \sqrt{\overline{R}^2}$$

  where $\overline{R}$ is the mean value of the differences between consecutive RRIs.

- **Mean Heart Rate (MeanHR)** - Is the mean value of the heart rate. The heart rate for each RRI was firstly computed, followed by the calculation of the mean value of the heart rates.

- **MaxInterval** - The maximum interval existing among all RRIs.

- **MinInterval** - The minimum interval existing among all RRIs.

- **MeanRRI** - The mean value of the RRIs.

- **MedianRRI** - The median value of the RRIs.

- **Standard Deviation of Heart Rate (SDHR)** - Was obtained computing the standard deviation of the heart rate values calculated from the RRIs.

The RRIs were computed resourcing to the function `peaks` from pracma package [6]. The heart rate of each RRI was subsequently computed and the time-domain features were extracted. This procedure was applied to each filtered signal of every individual. The feature vectors were then stored into a single data set, with a label to distinguish the users. The same class labels were used, *i.e.* "Owner" and "Intruder" corresponding to id values of 0 and 1, respectively. Furthermore, three data sets were created, corresponding to the length of the signals used, *i.e.* 5, 10 and 20 seconds, as illustrated in Table 4.6.

Dataset	Signal Duration (seconds)	Number of Records	Number of Features
AS_TF_5	5	1300	14
AS_TF_{10}	10	650	14
AS_TF_{20}	20	325	14

Table 4.6: Data sets generated using Time-Domain Features extracted **Angel Sensor** data.

Tests regarding this approach will be exposed in Chapter 6, where evaluation metrics obtained using time-domain feature extraction will be presented, and a comparison between fiducial dependent and fiducial independent approaches will also be performed.

Chapter 5

Architecture Design

After defining the methodology, some validation needs to be performed. Being the objective of this thesis to assess the possibility of using the heartbeat in a biometric system, a practical application of the procedures exposed in Chapter 4 will be defined in this chapter. Moreover, relevant aspects of the experiment, such as the hardware used and the implementation of all the stages of the application will also be referred.

5.1 Hardware

Before explaining the implementation of the system, a reference to the hardware will be provided regarding the two specific devices composing the system, *i.e.* the device that allowed capturing biometric data, and the processing unit to perform the remaining stages of the system.

Photoplethysmogram (PPG) data was acquired using three *Angel Sensor* wristbands, with a dual Light Emitting Diode (LED) PPG sensor, each. As aforementioned, the device communicates over Bluetooth when paired with another device that can access the data and services resourcing to the Application Programming Interface (API) provided. The wristband enables recording the PPG signals, body temperature and the magnitude of the accelerometer data. It is worth mentioning that, despite containing the same hardware, the accelerometer magnitude values would differ between wristbands.

Regarding the device to be paired with the wristband, a LG Nexus 4 was chosen. The relevant aspect of the device is the Bluetooth 4.0 version, due to minimum requirements for the communication with the wristband.

5.2   Software Architecture

The first structural decision was related to the development environment. In order to ensure high
portability potential to the system, it was decided to be implemented using Java programming
language, to enable a posterior port of the implementation to devices running Android Operating
System (OS).
Consecutively, the main actions to be performed by the system were identified and defined, as:

- **Enrollment** - Where a user enrolls himself with the system, *i.e.* when his biometric data
  is collected, processed and stored in the system in order to enable his identity verification
  in posterior usages.

- **Challenge** - Where the user inputs his biometric data, in this specific case, when the
  sensor captures the PPG wave, and initiates the process of verifying the identity.

- **Verification / Response** - When the system sends a feedback regarding the collected
  data. The response will be either a approval or denial of access if the user is verified or
  refuted, respectively.

However, the aforementioned actions are further divided into several smaller procedures
that constitute every biometric system. In a more thorough analysis we can distinguish all the
constituting stages, as visible in Figure 5.1.
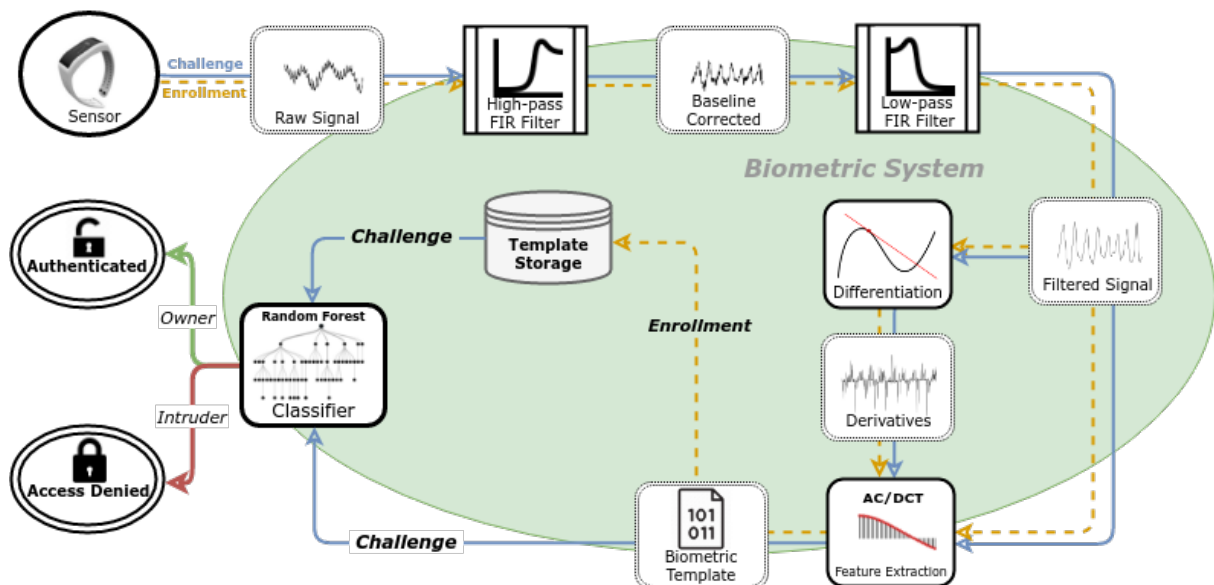
Figure 5.1: Diagram of the implemented system.

Biometric data is firstly obtained from the sensor and send to the processing unit, using
a Bluetooth channel. The received data is then submitted to the pre-processing stage for
interference attenuation and signal reconditioning, specifically the sequence of Finite Impulse

Response (FIR) filters. The filtered signal will then be submitted to a feature extraction algorithm to generate the biometric template that will be labeled and sent to the subsequent phase. When the user is to be enrolled in the system, the generated template will be sent to storage. However, if the action comprises the challenge phase, the template will be sent to the classifier which will determine if the user is to be authenticated or refuted, based on the existing templates.

5.3 System Implementation

In this section, information regarding all the stages of the implemented system is presented, and addressed in detail, from the first stage of the system - the data collection stage - to the last phase - the verification and response.

5.3.1 Data Collection

To collect the data an Android application was developed, resourcing to the Angel Sensor and Android APIs [24, 31]. Consisted of a simple application that recorded the PPG data from the wristband and stored each record.

The data obtained from the sensor would firstly go to the `Window_Buffer` where it would be tested for excessive noise, *i.e.* if the acceleration magnitude exceeded the defined boundaries, the data from both buffers would be excluded and another recording session would take place, otherwise the data contained in the `Window_Buffer` would be sent to the `Record_Buffer` being the prior reset to repeat the process until the end of the recording session. Figure 5.2 illustrate the buffers scheme and the data streams flow.
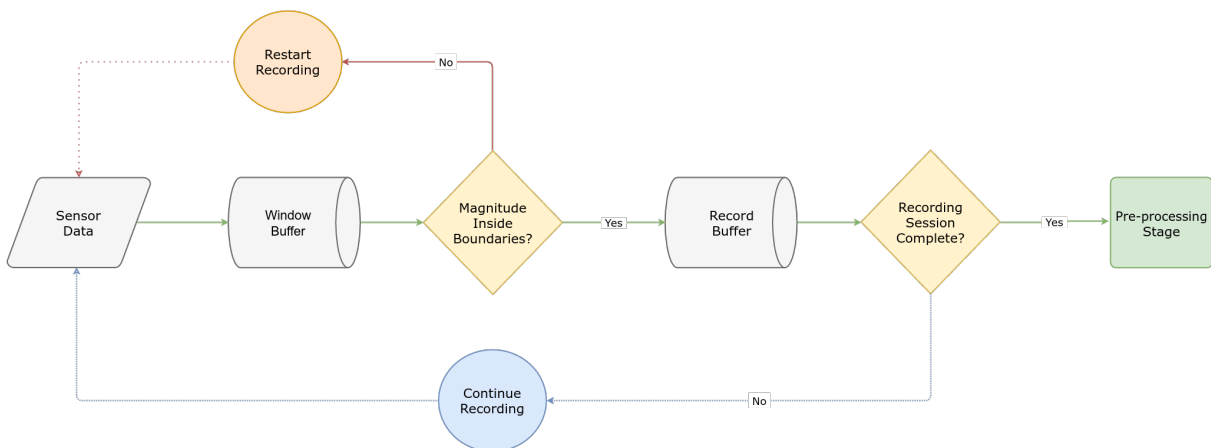
Figure 5.2: Diagram of the dual buffer scheme.

To define a decent range of magnitude values to be accepted, while also dealing with the different magnitude values among wristbands, the acceleration waveform was analyzed as visible in Figure 5.3 and its average values were identified, standing in a range of 15000 and 18000 units,

which were defined as the initial lower and upper boundaries, respectively. The different colored lines represent the magnitude obtain from the sensors used.

Angel Sensor Accelerometer Magnitude
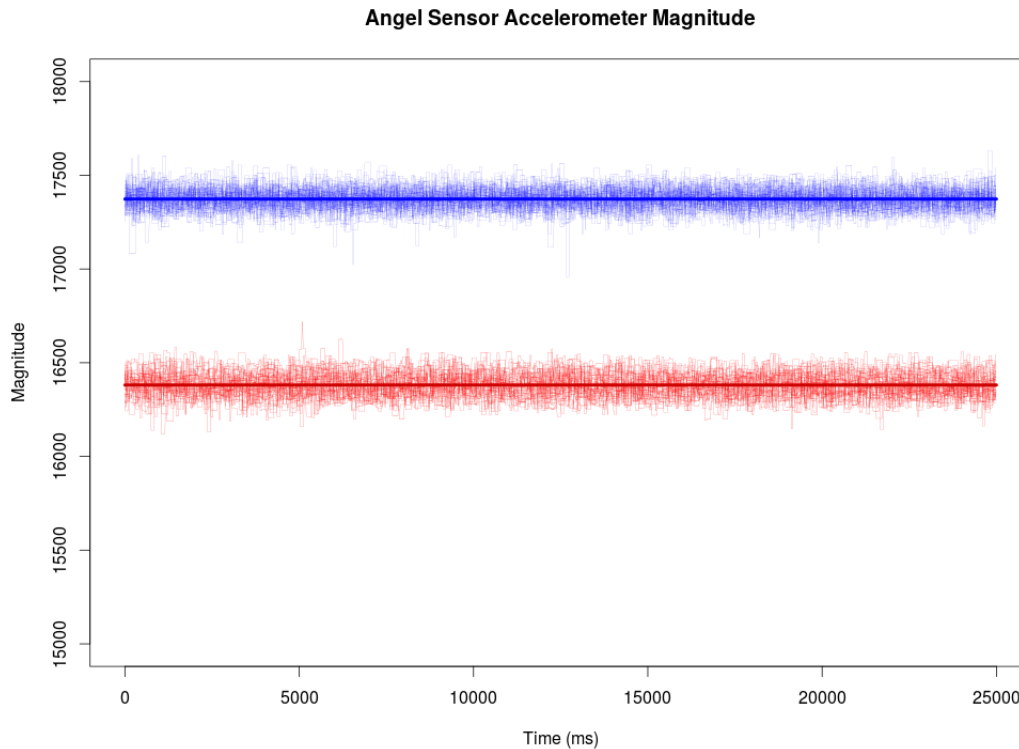
Figure 5.3: Accelerometer magnitude of the recordings obtain from the PPG sensors and the respective average. The thicker lines represent the grand mean of the magnitude oscillations.

During the execution, and while readjusting the margins to guarantee a decent usability and proper collection of data, the process of assessing the magnitude was performed using the magnitude average of the data from the `Window_Buffer`, that is, when the first buffer was filled, the mean value was computed and added or subtracted with a constant value to enable an adaptive evaluation, for its upper and lower bounds, respectively. At each cycle, when the `Window_Buffer` was ready to send the data to the `Record_Buffer`, the magnitude margins would be readjusted regarding the data from the ended cycle.

The recording session would be finished when the `Record_Buffer` was full and its data would be sent to the filtering stage. Algorithm 1 illustrates the recording session and the usage of the buffers scheme.

Algorithm 1 PPG Signals Recording Session

1: **procedure** POCESSPPGSIGNAL($ppg\_signals, acc\_magnitude$)
2:     $InitializeBuffers()$
3:     **while** !($IsFull(record\_buffer)$) **do**
4:         $ppg\_signals \leftarrow GetPPGsFromSensor()$
5:         $acc\_magnitude \leftarrow GetMagnitudeFromSensor()$
6:         $ProcessBuffers(ppg\_signals, acc\_magnitude)$
7:     **end while**
8:     $StartFilteringData(record\_buffer)$ ▷ Finished recording. Filter stage is next
9: **end procedure**

10: **procedure** PROCESSBUFFERS($ppg\_signals, acc\_magnitude$)
11:     **if** !($IsFull(window\_buffer)$) **then**
12:         $window\_buffer.Add(ppg\_signals, acc\_magnitude)$
13:     **else**
14:         $magnitude\_average \leftarrow ComputeMagnitudeAverage(window\_buffer)$
15:         **if** $IsInsideBounds(magnitude\_average)$ **then**
16:             $TransferToRecordBuffer(window\_buffer)$
17:             $UpdateMagnitudeBoundaries(magnitude\_average)$
18:         **else**
19:             $ResetBuffers()$ ▷ Restart recording
20:         **end if**
21:     **end if**
22: **end procedure**

5.3.2   Pre-processing

When the recording stage ends, the PPG signal are then sent to the pre-processing stage. As mentioned in Section 4.3, this stage consists of applying a sequence of FIR filters and recondition the signal to remove the delay imposed.

Since there was no functions implemented to apply a FIR filter directly to the signal, a different approach was followed. The process consisted of precomputing and storing the filter coefficients, obtained using the function `fir1` [44] for R, for high-pass and low-pass filters.

Having the filter coefficients prepared, the subsequent step is a process named *Convolution*.

Convolution   is defined as an operation that receives two functions and merges both into a third function. Specifically, in this scenario, the inputs are the PPG signals and the FIR coefficients for each intended filter, and the output would be the filtered waveform.

As aforementioned, resourcing to the convolution function, it is now possible to perform the FIR filter sequence. The first stage consists of convoluting the raw PPG signal with the high-pass FIR coefficients, of order 294, to attenuate the baseline drift. The resulting signal will be designated `hp_signal` for future references. The second stage is to do a second convolution using the `hp_signal` and the low-pass FIR coefficients of 6th order, to remove high frequency noise, and will be referenced as `lp_signal`.

When the full filtering stage is finished, the signal still needs to be reconditioned before proceeding to further stages of the system, as it will contain misleading data points that correspond to the delay caused by the filter. When using FIR filters with linear phase, as the case of the ones used in this experiment, the delay can be easily determined by the expression

$$Theoretical\ Delay = \frac{(N^{th}order-1)}{(2\times sampling\_rate)}$$

where *sampling_rate* is the rate at which the sensor retrieves data points from the PPG signal, which, in this specific case, was 100 Hz.

However, when testing it while reconditioning the signal, it still contained traces of interference added by the filter delay. To overcome this constraint, and guarantee that only the signal data was being kept, the amount of data points considered as delay was defined to be the sum of both filter orders, *i.e.*,

$$Delay = HP\_Order + LP\_Order$$

where *HP_Order* and *LP_Order* are the orders of the high-pass and low-pass FIR filter coefficients, being 294 and 6, respectively.

After reconditioning, the first and second derivatives of the filtered signals would be computed, D_1 and D_2, respectively, being defined as:

$$D_1(x_i) = \frac{x_{i+1}-x_i}{t_{i+1}-t_i}$$

$$D_2(d_i) = \frac{d_{i+1}-d_i}{t_{i+1}-t_i}$$

where x_i is the ith data point of the signal S, d_i is the ith data point of D_1 and t_i is the time instance of the ith data point of a signal.

The filtered signals and respective derivatives would then be sent to the subsequent stages of the system, where features would be extracted to generate the template.

5.3.3 Feature Extraction and Template Generation

Regarding the feature extraction phase, which, as mentioned in Section 4.4, consisted of applying the AC/DCT method. Both Auto Correlation (AC) and Discrete Cosine Transformation (DCT) functions were implemented, being the latter a 2 dimensional Forward DCT [52], hereinafter referred as `Autocorrelation` and `ForwardDCT`, respectively.

`Autocorrelation` receives two arguments, the previously filtered signal, and the order of the function, that is, the number of coefficients to retrieve. The order of the `Autocorrelation` would depend on the signal itself, and was defined by the following expression:

$$Order = 10 \times \log_{10}(\tfrac{N}{m})$$

where N is the number of data points of the signal, and m the number of records. Since `Autocorrelation` would be applied individually to each record ($m = 1$), the expression can be simplified to be:

$$Order = 10 \times \log_{10} N$$

Such order definition was used to maintain all the processes as similar as possible to the methodology defined. Furthermore, since the pre-processing tests exposed in Section 4.3 were performed using R as a development environment, some default definitions were replicated.

`ForwardDCT` was defined to receive two array arguments, one containing the AC coefficients and a second to be the storage of the DCT spectrum computed.
The resulting DCT array would then be sent to the following stage of the system, either enrolling the user with the system or verifying his identity depending on the action, *i.e.* if the action was the *Enrollment* or the *Challenge*, respectively.

5.3.4 Enrollment

As aforementioned, the enrollment consists of giving information to the system so it can recognize the person trying to authenticate using the biometric template obtained from the data collection stage. Furthermore, in the context of a system that resources to a machine learning algorithm, the enrollment is the process of fitting the model to the provided data.
Regarding the implementation and, since all of the signal treatment is performed in the previous stages, the enrollment was simply the process of labeling the feature template and storing the information into a data set. However, and for the purpose of this thesis, the classifier was submitted to an offline training, *i.e.* when it was imported, it had already been trained to the

data obtained.

As exposed in Chapter 6, the algorithm used was Random Forest, specifically, using the `weka.classifiers.tree.RandomForest` class from Weka API [23].

5.3.5   Verification & Response

Regarding the *Verification and Response* action, the procedure can be defined in three stages:

1. Collect and process biometric data generating a feature template

2. Label the template and submit it to the machine learning model to be classified

3. Output a response based on the classification obtained

In a more thorough analysis, and having a labeled biometric template collected from the user to be authenticated, the system would feed his template to the Random Forest classifier, that would produce a binary result, either 0 or 1, corresponding to the designations of "Owner" and "Intruder", respectively.

If the response value was 0, a confirmation message would be sent, validating the user and consequently granting him access. If, otherwise, the classification returned as 1, the user identity would be refuted and the access he required would be denied.

In this chapter, the system details were exposed. Stages regarding the collection of the PPG signals, the concerns about the quality of the data, the filtering stages, feature extraction and the main actions performed by the system were scrutinized. However, it is necessary a performance evaluation of the system to determine its feasibility in real life scenarios. Such analysis will be performed in Chapter 6.

Chapter 6

Evaluation

In this chapter, a thorough evaluation of the machine learning algorithms chosen is performed, using every data set generated from the two public databases and the Photoplethysmogram (PPG) sensor, and is divided in three sections.

In Section 6.1 metrics regarding different tests are exposed, while exploring some variations of the data and the algorithms used for classifying instances. Specifically, the data sets exposed in Chapter 4 are tested to determine the best classifier to be used in the system.
The first analysis, consists of testing the models with the data sets generated from the public databases, specifically **ECG_ID** from *PhysioNet* and, **ECG_TB** and **PPG_TB** from *CapnoBase.*

In Section 6.2 a second analysis using data sets generated from the PPG sensor is exposed, evaluating both the machine learning algorithm and the number of features to extract in the proposed system. Finally, a comparison between a fiducial dependent approach, using temporal features related to Heart Rate Variability (HRV), and the AC/DCT method is presented.
The presented illustrations and tables may only expose a brief summary of all the evaluation metrics obtained which will be exposed in detail in Appendix A.

Section 6.3 consists of a discussion regarding all the work performed. Relevant aspects of the databases and generated data sets are referred as well as some considerations about hardware issues and major decisions in regard to the chosen methodology and implementation performed.

6.1   Public Data Sets Analysis

Two public databases were used to generate different data sets, as referred in Chapter 4. The main constraints regarding this phase of testing consisted of varying length of the waveform signal used and testing with different machine learning algorithms, specifically k-Nearest Neighbors (KNN), Support Vector Machine (SVM), Random Forest (RF), Linear Discriminant Analysis (LDA) and Naive Bayes (NB).

The tests consisted of performing 5 repetitions of a 10 fold cross-validation strategy, making a total of 50 iterations, per model, resourcing to *WEKA* environment. The metrics obtained are a result of computing the mean value of each metric for all the iterations performed.

6.1.1   PhysioNet Data Sets

The following content illustrate the metrics obtained when testing the five machine learning models with Electrocardiogram (ECG) signals with duration of 5, 10 and 20 seconds, designated as **ECG_ID_5**, **ECG_ID_10** and **ECG_ID_20** data sets, respectively.

Figure 6.1 exposes the F1 Score of the different models when tested with the three aforementioned data sets, while in Figure 6.2, a comparison between the False Positive Rate (FPR), represented with blue, and False Negative Rate (FNR), represented with orange, is exposed. In Figure 6.3 the Matthews Correlation Coefficient (MCC) of all the classifiers used with the *PhysionNet* data sets is illustrated. Tables containing all the metrics computed for the aforementioned data sets are accessible in Appendix A, specifically Tables A.1, A.2 and A.3.
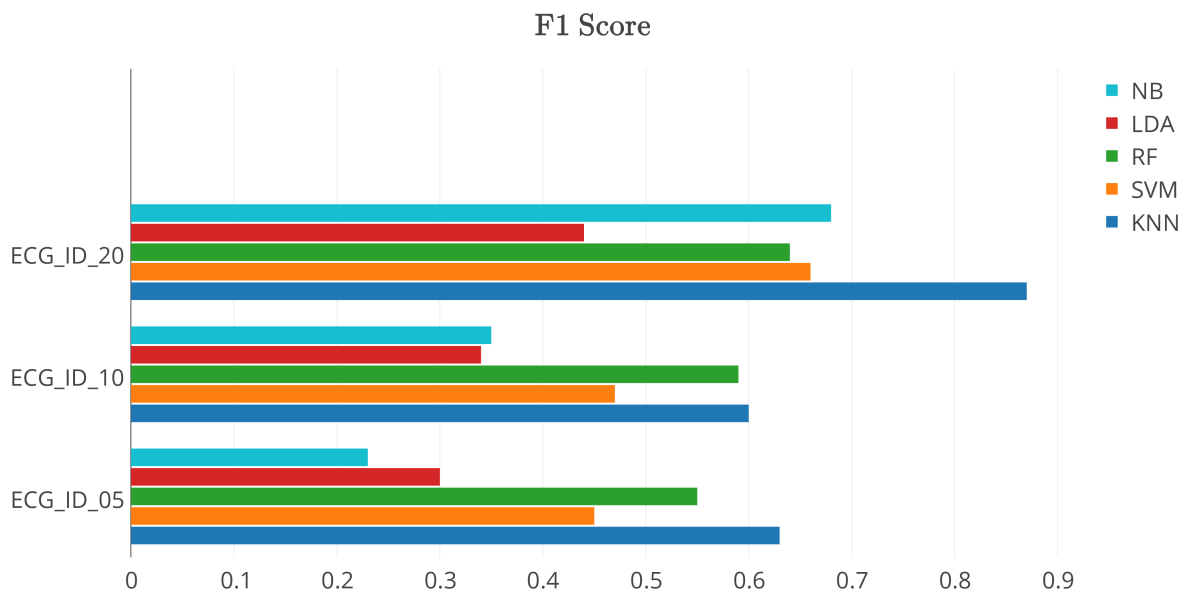
Figure 6.1: F1 Score of **ECG_ID** data sets, varying the duration of the ECG samples.

By observing Figure 6.1, it is noticeable the increase of performance when the duration of the waveform increases, reaching almost 0.90, namely the KNN model, while the second best, specifically the NB, nearly reached 0.70.

(a) FPR and FNR for KNN classifier.

(b) FPR and FNR for SVM classifier.

(c) FPR and FNR for RF classifier.

(d) FPR and FNR for LDA classifier.

(e) FPR and FNR for NB classifier.
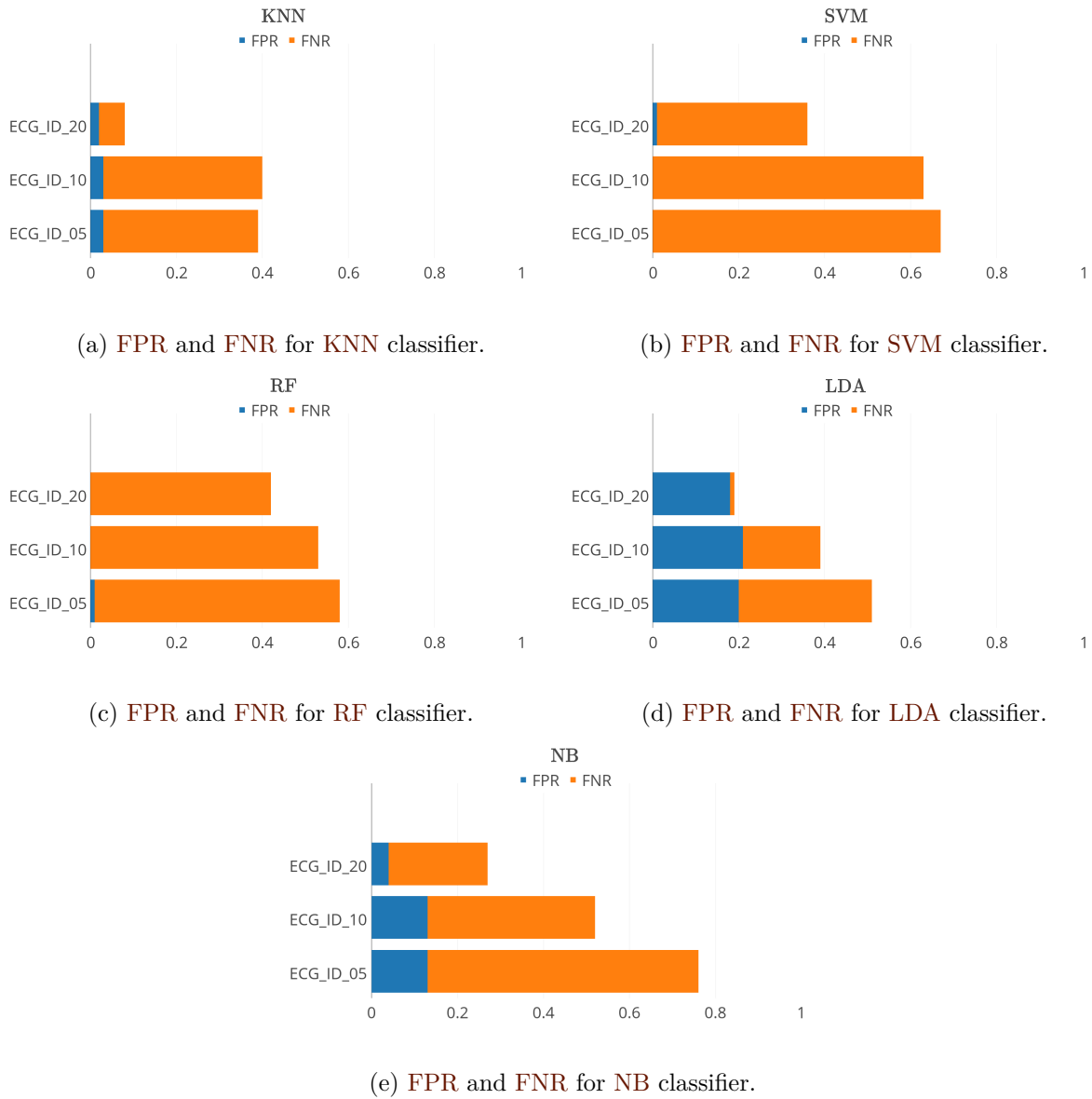
Figure 6.2: False Positive and Negative Rates obtained with five different classifiers using the *PhysioNet* data sets.

Observing Figure 6.2, it is noticeable the predominance of KNN since it has the lowest cumulative rates, in almost every data set, among all the classifiers.
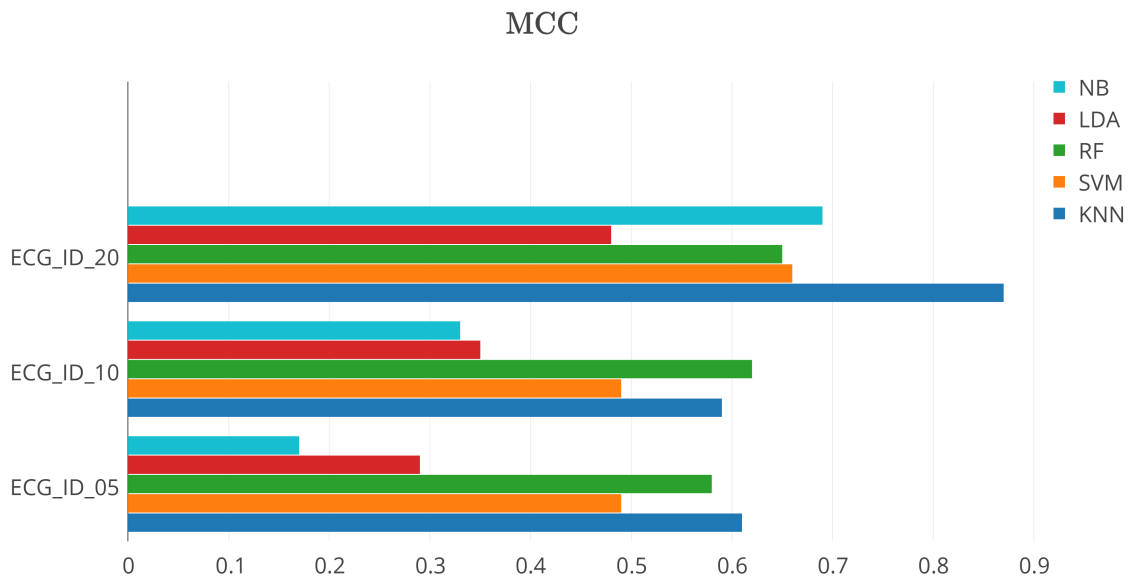
Figure 6.3: Matthews Correlation Coefficient of the classifiers for the **ECG__ID** data sets.

Similarly to Figure 6.1, the performance of the classifiers tend to increase when the duration of the signals that generated the data set increases. As expected after the evaluation of the F1 Score, the KNN is the classifier with highest rate, nearly 0.90 followed by NB with a value close to 0.7.

From a global perspective, by evaluating the metrics illustrated, the first conclusion one may extract is the classifier that best handled the *PhysioNet* data sets and the major differences between the best model and the remainder. KNN was manifestly better than the remaining classifiers, in two of the three data sets tested. A common characteristic among classifiers was the increase of performance in the data sets with highest wavelength duration.

When addressing this study to its application in authentication scenarios, despite the highest rates when considering F1 Score and MCC, a thorough analysis of the false acceptance and rejection of users from the system is needed, *i.e.* a decent trade-off needs to be embraced that enables both usability and security. Thus, by looking into the FPR and FNR values from Figure 6.2 it is possible to start addressing the aforementioned situation.

In terms of False Acceptance Rate (FAR), *i.e.* FPR, the near optimal classifiers are SVM and RF. Although, the the FNR is extremely high for an authentication system, surpassing 30% in its best performance, and reaching 60% when considering shorter ECG signals, which is impracticable. In other words, a user trying to authenticate 10 times in a row, would be rejected, in average, 3 and 6 times, in the best and worst case scenarios, respectively. Therefore, the KNN was elected to be the primal classifier and the duration of the ECG records should be of 20 seconds.

However, since the type of data used in the system is different, *i.e.* PPG signals will be extracted,

this conclusion is merely an indicative of the considerations to have in mind when assessing the remaining data sets.

6.1.2 CapnoBase Data Sets

The following illustrations expose the results obtained when evaluating the *CapnoBase* data sets. As mentioned in Chapter 4, two groups of data sets were generated for the ECG and PPG signals contained in the database. Similarly to the evaluation performed in Section 6.1.1, F1 Score, MCC and FPR and FNR were assessed and all the metrics computed are detailed in Appendix A, specifically, Tables A.4, A.5, A.6, A.7 and A.8 for the ECG signals, and A.9, A.10, A.11, A.12 and A.13 for the PPG signals.

ECG data sets

Figure 6.4 exposes the F1 Score of the classifiers when dealing with ECG with duration of 5, 10, 20, 30 and 40 seconds.
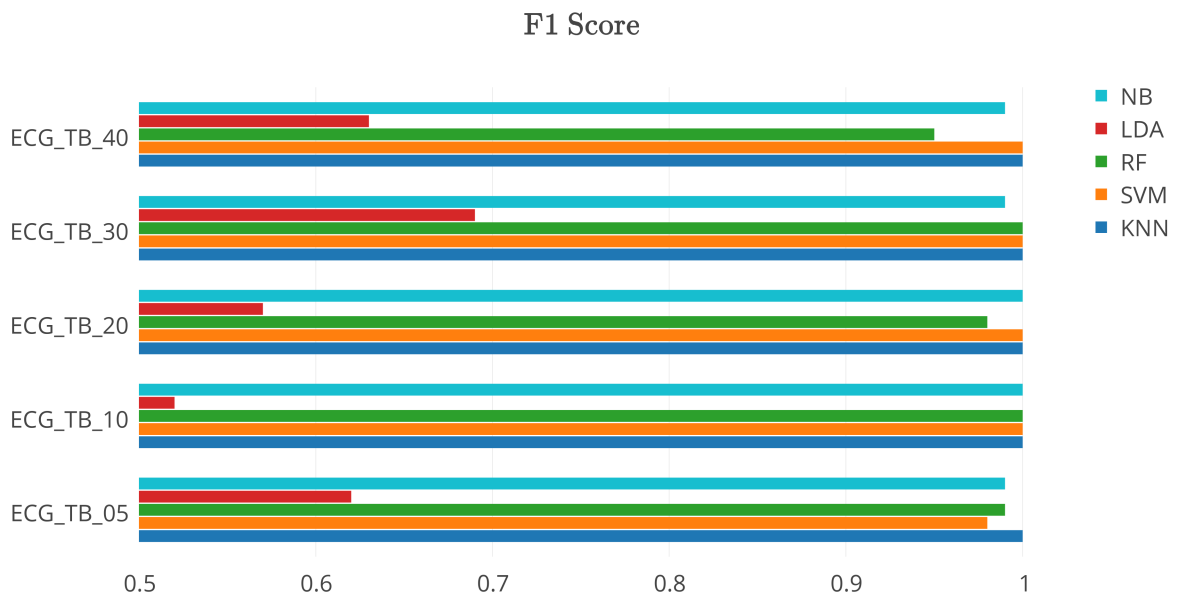
Figure 6.4: F1 Score obtained with five different classifiers using the *CapnoBase* data sets extracted from ECG signals.

By analyzing the F1 Score, it is evident the increase of performance when the duration of the signal exceeds the 5 seconds. With a record of 10 seconds, only the LDA classifier performs poorly, while the remainder obtained a perfect score of 1. As the duration of the wave increases,

it is visible some oscillations on the performance of RF and NB. Similarly to **ECG_ID**, the KNN is the classifier with the best score, regardless of data set.
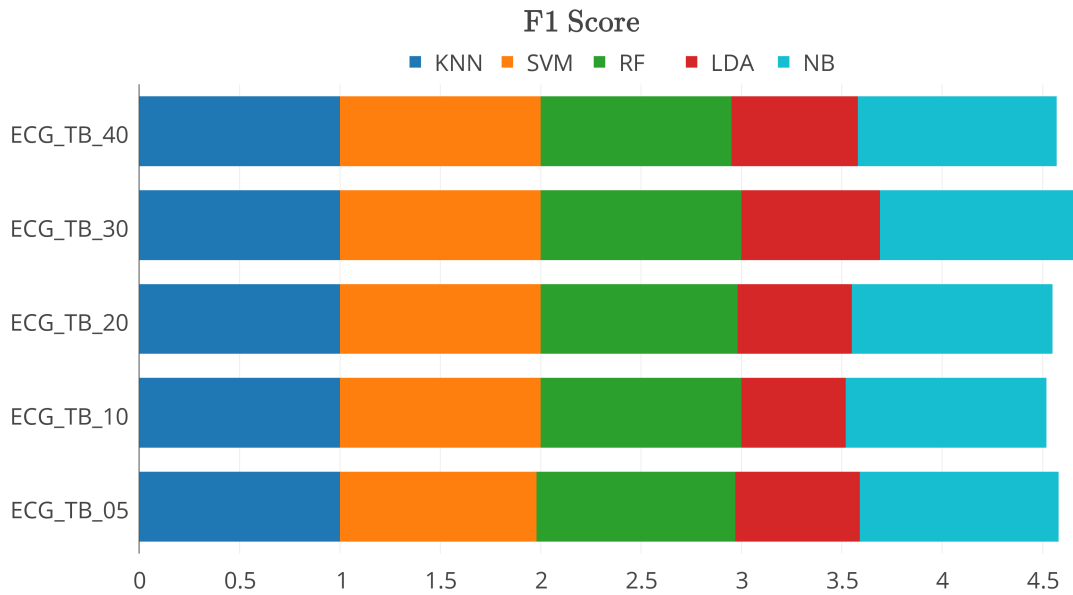
Figure 6.5: F1 Score of the classifiers stacked by data set.

Figure 6.5 exposes scores of the classifiers stacked by data set to evidence the best duration for the ECG records, which is the 30 seconds mark.

Figures 6.6 and 6.7 illustrate the Matthews Correlation Coefficient of the classifiers. The prior exposes the results of each classifier for every data set, while the latter stacks the results by data set.
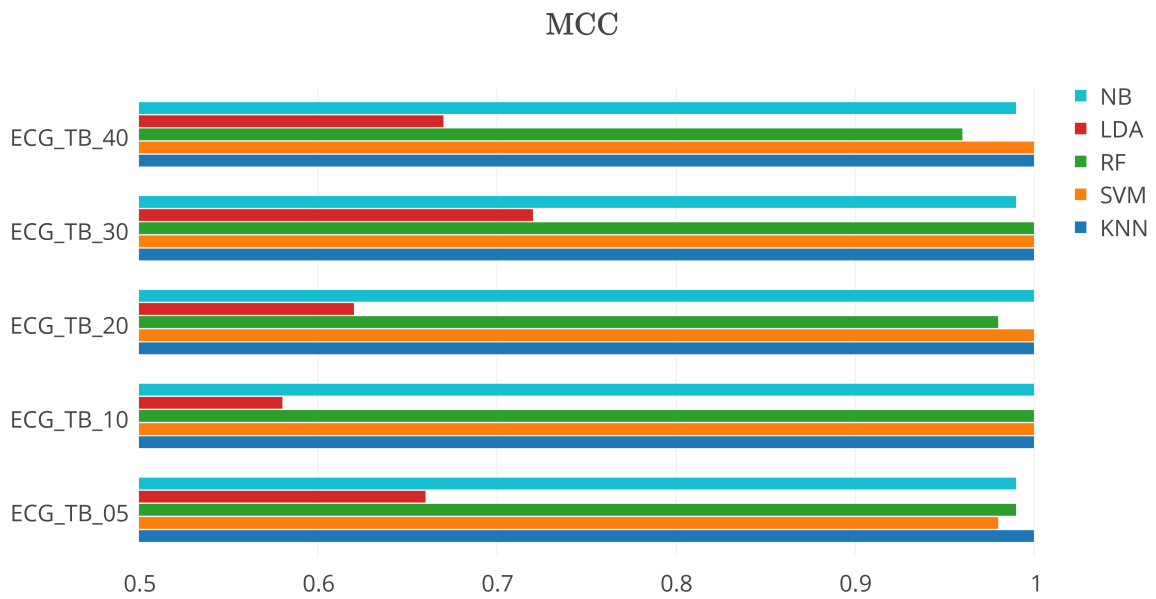
Figure 6.6: MCC of the five different classifiers using the *CapnoBase* data sets extracted from ECG signals.
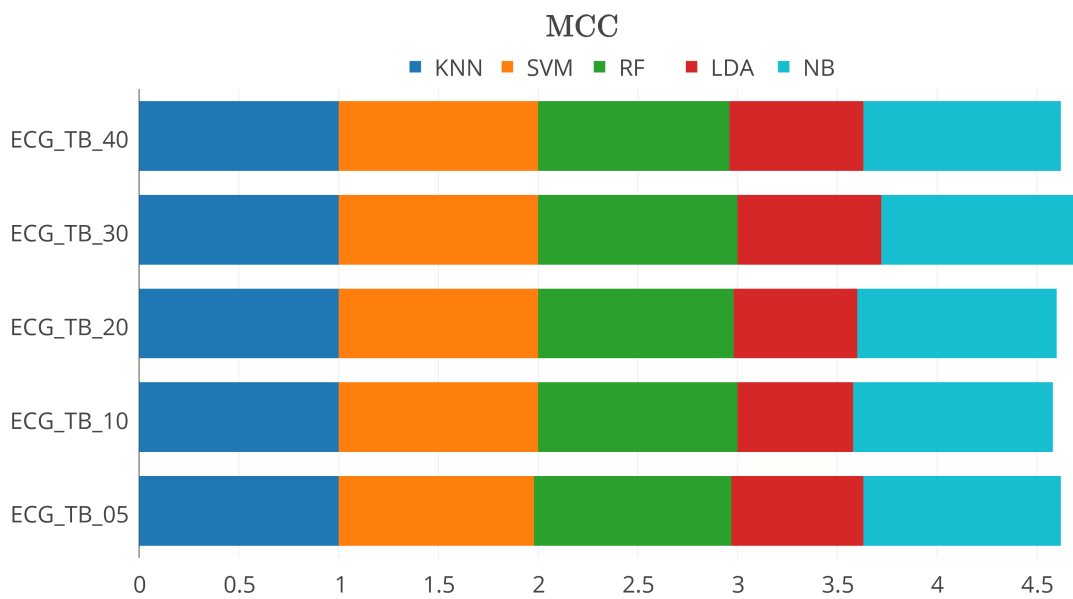
Figure 6.7: MCC of the classifiers stacked by data set.

Similarly to the F1 Score results, the KNN is the best classifier, having the maximum correlation coefficient in every data set tested. Furthermore, the best performance is achieved with 30 seconds ECG records.

In order to assess the false acceptance and rejection of the system, FPR and FNR were computed and exposed in Table 6.1.

Data sets	FPR	FNR
ECG_TB_05	0.00	0.01
ECG_TB_10	0.00	0.00
ECG_TB_20	0.00	0.00
ECG_TB_30	0.00	0.00
ECG_TB_40	0.00	0.00

(a) KNN

Data sets	FPR	FNR
ECG_TB_05	0.00	0.00
ECG_TB_10	0.00	0.00
ECG_TB_20	0.00	0.00
ECG_TB_30	0.00	0.00
ECG_TB_40	0.00	0.00

(b) SVM

Data sets	FPR	FNR
ECG_TB_05	0.00	0.03
ECG_TB_10	0.00	0.00
ECG_TB_20	0.00	0.04
ECG_TB_30	0.00	0.00
ECG_TB_40	0.00	0.07

(c) RF

Data sets	FPR	FNR
ECG_TB_05	0.03	0.00
ECG_TB_10	0.05	0.00
ECG_TB_20	0.04	0.00
ECG_TB_30	0.02	0.00
ECG_TB_40	0.03	0.00

(d) LDA

Data sets	FPR	FNR
ECG_TB_05	0.00	0.01
ECG_TB_10	0.00	0.00
ECG_TB_20	0.00	0.00
ECG_TB_30	0.00	0.01
ECG_TB_40	0.00	0.01

(e) NB

Table 6.1: False Positive and Negative Rates of all classifiers using the *CapnoBase* ECG data sets.

Regarding the FPR, only LDA does not perform optimally, having the lowest rate of 2% with records of 30 seconds. When considering the False Rejection Rate (FRR), *i.e.* FPR, Random Forest has the highest rates with signals with 5, 20 and 40 seconds of duration.

In a general appreciation of the results obtained, two classifiers can be isolated among all, namely KNN and SVM. Despite being slightly worst when considering the F1 Score and MCC, the SVM presented better FPR and FNR, in average. However, it is worth mentioning that the results were already returned, by *WEKA*, as the mean value of all the iterations performed, thus the values may contain some deviation. As an example, when considering the F1 Score of KNN and SVM for ECG records of 40 seconds, they obtained the scores of 1 and 0.93 with standard deviation of 0.02 and 0.05, respectively. Notwithstanding, all the classifiers performed better when using records with 30 seconds, fact that will be focused when evaluating the results of the

classifiers when using the PPG signals.

PPG data sets

The following content will expose the results obtained when using the PPG signals in a biometric system. F1 Score and Matthews Correlation Coefficient will be illustrated in Figures 6.8 and 6.9, and, 6.10 and 6.11, respectively, and the False Positive and Negative Rates will be exposed in Table 6.2.
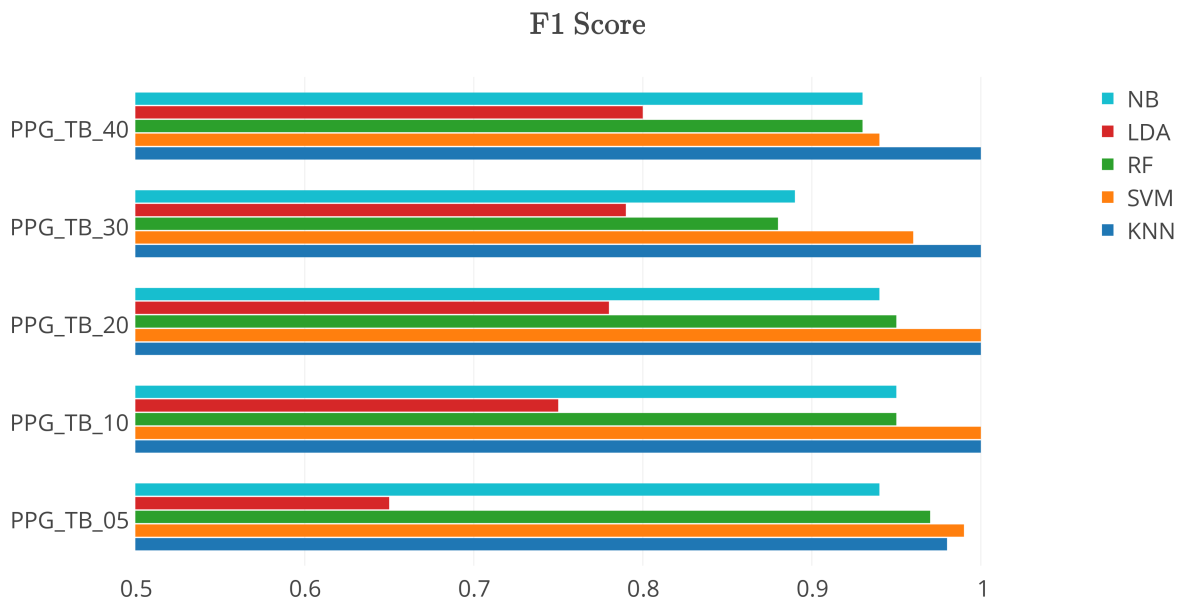
Figure 6.8: F1 Score obtained with five different classifiers using the *CapnoBase* data sets extracted from PPG signals.

In Figure 6.8 it is visible the predominance of KNN as the classifier with the better overall performance among in every data set. Despite SVM performing better with 5 seconds records, and matching KNN with 10 and 20 seconds records, when the duration of the samples increases, the score of SVM decreases.
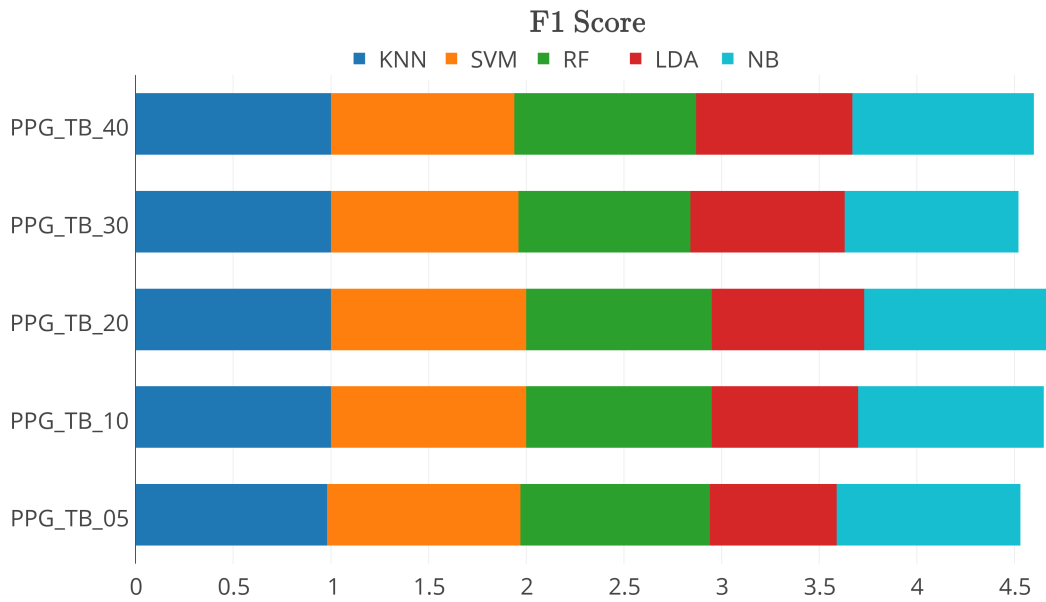
Figure 6.9: F1 Score of the classifiers stacked by data set.

By observing Figure 6.9 it is possible to obtain a better understating regarding the duration of the samples that enables better performances among every classifier with the cumulative score. The 10, 20 and 40 seconds PPG records allows the overall performance to surpass the 4.5 mark where, among the three data sets, the **PPG_TB_20** is the one with the highest value.
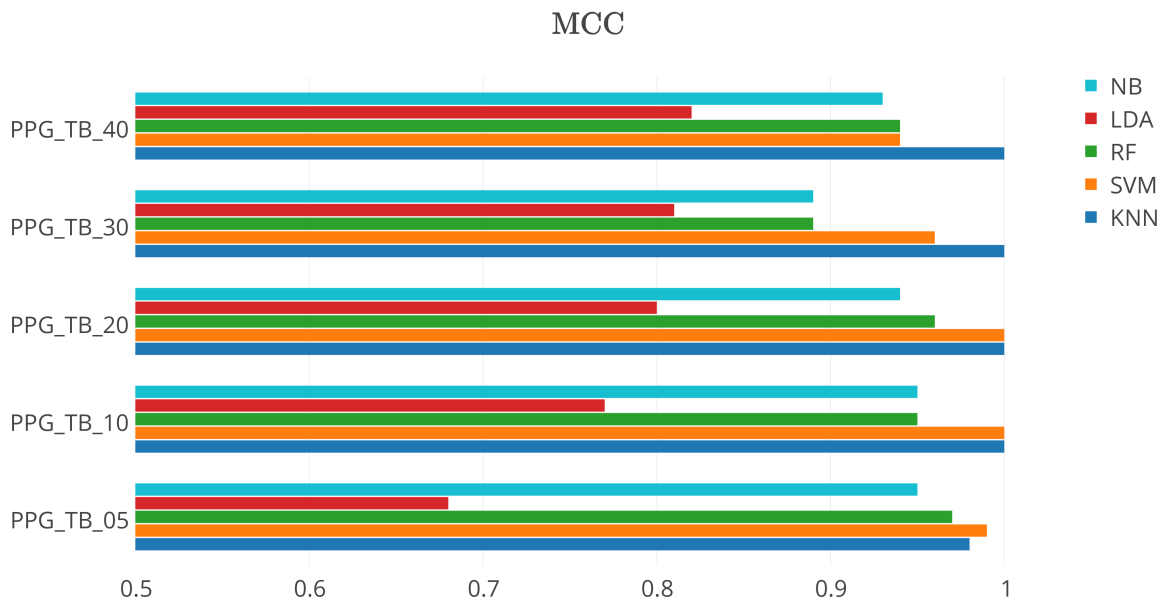
Figure 6.10: MCC of the five different classifiers using the *CapnoBase* data sets extracted from PPG signals.

Regarding the Matthew Correlation Coefficient, KNN and SVM remain being the best classifiers and, similarly to the F1 Score, SVM correlation coefficient decreases when the duration of the signal surpasses 20 seconds.
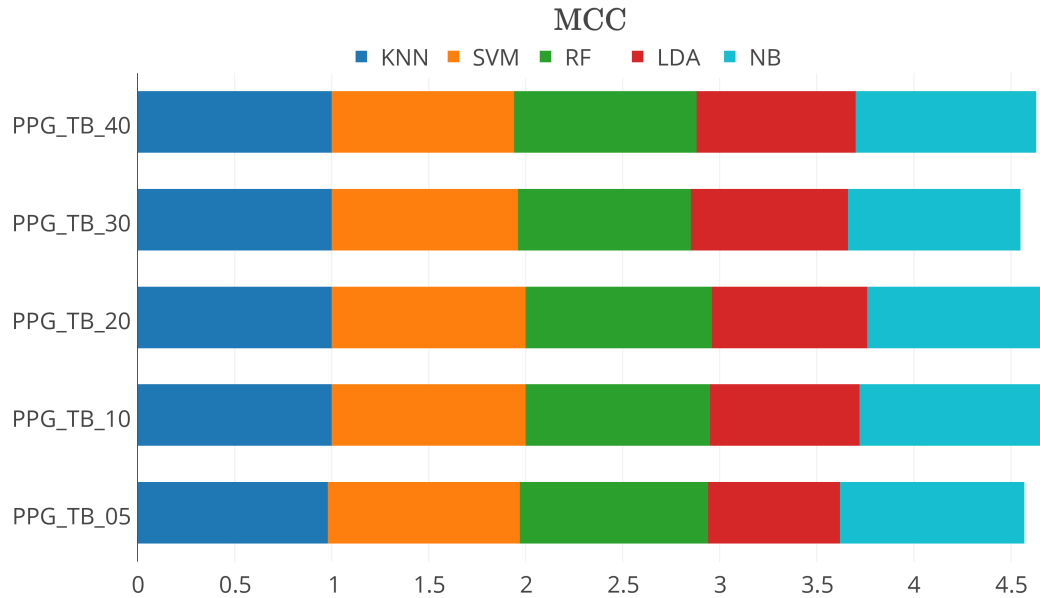
Figure 6.11: MCC of the classifiers stacked by data set.

Similarly to F1 Score, Figure 6.11 exposes the increase of performance by the classifiers when considering PPGs with 20 seconds.

Data sets	FPR	FNR
PPG_TB_05	0.00	0.01
PPG_TB_10	0.00	0.00
PPG_TB_20	0.00	0.00
PPG_TB_30	0.00	0.00
PPG_TB_40	0.00	0.00

(a) KNN

Data sets	FPR	FNR
PPG_TB_05	0.00	0.01
PPG_TB_10	0.00	0.00
PPG_TB_20	0.00	0.00
PPG_TB_30	0.00	0.07
PPG_TB_40	0.00	0.09

(b) SVM

Data sets	FPR	FNR
PPG_TB_05	0.00	0.05
PPG_TB_10	0.00	0.07
PPG_TB_20	0.00	0.08
PPG_TB_30	0.00	0.16
PPG_TB_40	0.00	0.09

(c) RF

Data sets	FPR	FNR
PPG_TB_05	0.03	0.00
PPG_TB_10	0.02	0.00
PPG_TB_20	0.02	0.00
PPG_TB_30	0.01	0.00
PPG_TB_40	0.01	0.00

(d) LDA

Data sets	FPR	FNR
PPG_TB_05	0.00	0.10
PPG_TB_10	0.00	0.09
PPG_TB_20	0.00	0.10
PPG_TB_30	0.00	0.16
PPG_TB_40	0.00	0.11

(e) NB

Table 6.2: False Positive and Negative Rates of all classifiers using the *CapnoBase* PPG data sets.

Regarding the False Positive and Negative Rates, KNN is still the preeminent classifier, outperforming every model.

From the evaluation performed, two major conclusions regarding the best classifier and the duration of the samples to collect were extracted. Since the data to be used in the proposed implementation is highly related with the data tested, the results obtained with PPG signals will have higher impact on the decisions regarding the implementation. Furthermore, since every classifier performed better when the PPG had 20 seconds of duration, the signals extracted from the *Angel Sensor* wristband will have that duration.

6.2 System Analysis

As aforementioned, the results from assessing the *CapnoBase* PPG data sets defined some baselines to test the data obtained from the PPG sensor. Having set the limit to 20 seconds, the

remaining data sets to be generated were then defined to contain records of 5 and 10 seconds. Regarding the classifier and despite KNN exceeding the remaining models, every classifier will be considered to subsequent tests.

Furthermore, the following illustrations expose the evaluations performed to the **Angel Sensor** data sets, specifically the **AS** and **AS_TF** sets, obtained from the AC/DCT and Temporal Features approaches, respectively.

Despite having a clear idea about the classifiers that better handle the data sets generated from the PPG signals due to the evaluation of the *CapnoBase* data sets, a final test comparing all the classifiers is required. Posteriorly, the number of features to extract from each signal considered, namely the PPG signals obtained from both Light Emitting Diode (LED)s and the respective first and second derivatives, also needs to be defined.

6.2.1   Fiducial Independent Approach

Figures 6.12, 6.14 and 6.13 expose the F1 score, Matthews Correlation Coefficient and a comparison of the False Positive and Negative Rates of all the classifiers used. The classifier elected as the one with highest performance, when considering the desired constraints, will be then tested with several data sets varying the number of features used per signal, specifically in Figures 6.15, 6.16 and 6.17, where the F1 Score, MCC and, FPR and FNR, will be exposed.
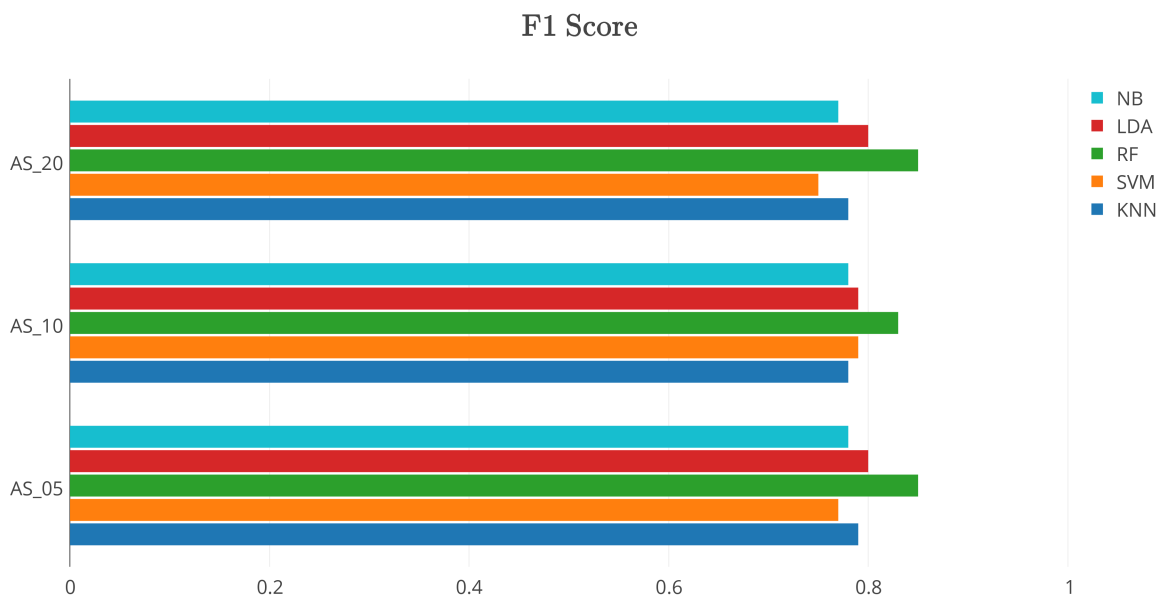
Figure 6.12: F1 Score obtained with five different classifiers using the *Angel Sensor* data sets extracted from PPG signals.

As visible in Figure 6.12, the classifier with the highest F1 Score is Random Forest, regardless

of duration. However, the best results are obtained when considering a signal of either 5 or 20 seconds.

(a) FPR and FNR for KNN classifier.

(b) FPR and FNR for SVM classifier.

(c) FPR and FNR for RF classifier.

(d) FPR and FNR for LDA classifier.

(e) FPR and FNR for NB classifier.
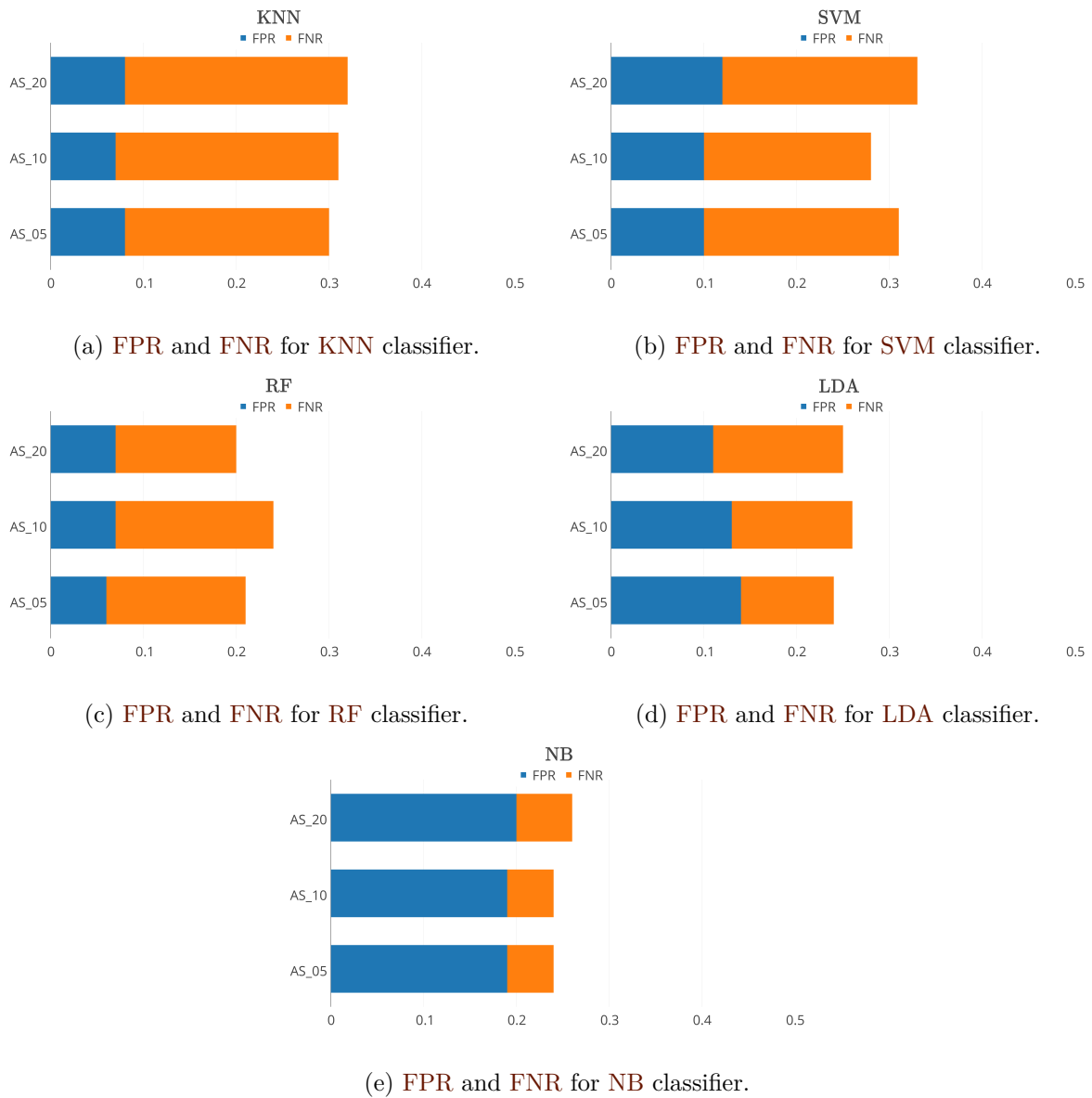
Figure 6.13: False Positive and Negative Rates obtained with five different classifiers using the *Angel Sensor* data sets.

Regarding the comparison of False Positive and Negative Rates, RF and KNN obtained similar values of FPR. Despite having the best FNR among classifiers, NB presents a rate of false acceptance of roughly 20% in all data sets tested. Overall, the lowest cumulative rates belong to RF and LDA, however, the latter has a FPR exceeding the 10% mark in every data set, being approximately two times greater than RF.

MCC
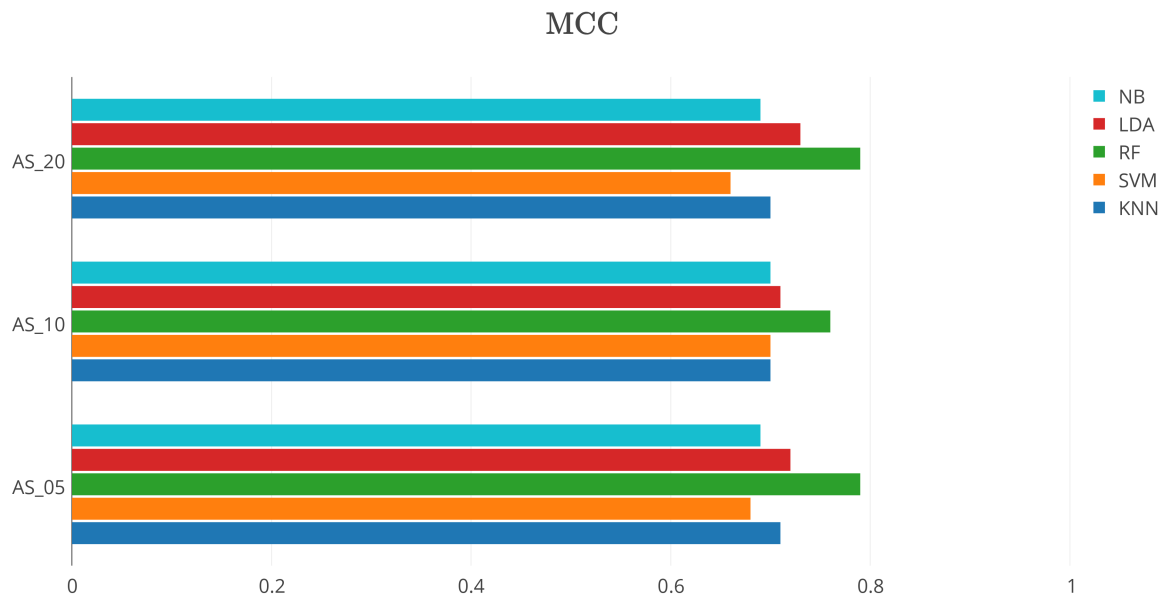
Figure 6.14: MCC of the five different classifiers using the *Angel Sensor* data sets extracted from PPG signals.

By evaluating the figures above, it is noticeable the predominance of Random Forest, since it has the highest rates, either F1 Score and MCC, and the best FPR and FNR ratio, among all the data sets tested, thus it will be chosen as the classifier for the proposed implementation. Having a classifier defined, the subsequent evaluations regards the number of features to be extracted. Figures 6.15, 6.16 and 6.17 illustrate the F1 Score and respective Precision and Recall values, False Positive and Negative Rates, and Matthews Correlation Coefficient, respectively, of Random Forest for the **AS** data sets with PPG signals of 5, 10 and 20 seconds, and having 3, 4, 5 and 6 features per signal extracted.
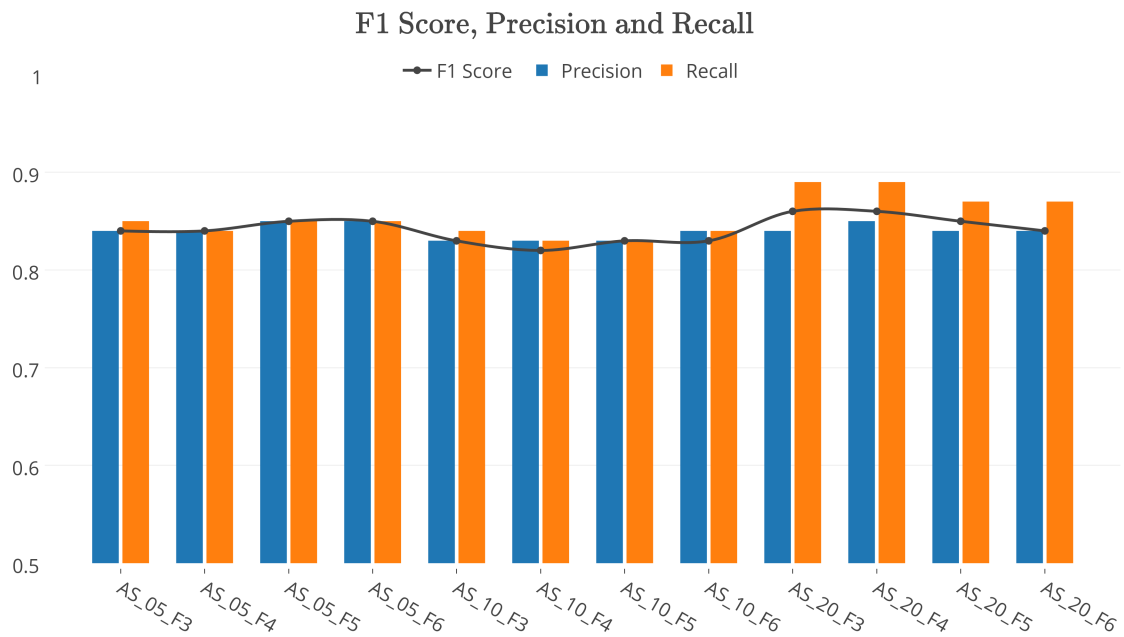
Figure 6.15: F1 Score of Random Forest with a variation over the number of features of the **AS** data sets.

From Figure 6.15 is is possible to verify a slightly increase of performance when considering **AS_20_F3** and **AS_20_F4** data sets. Despite the Precision being roughly the same, the Recall is clearly higher with the 20 seconds PPG records.
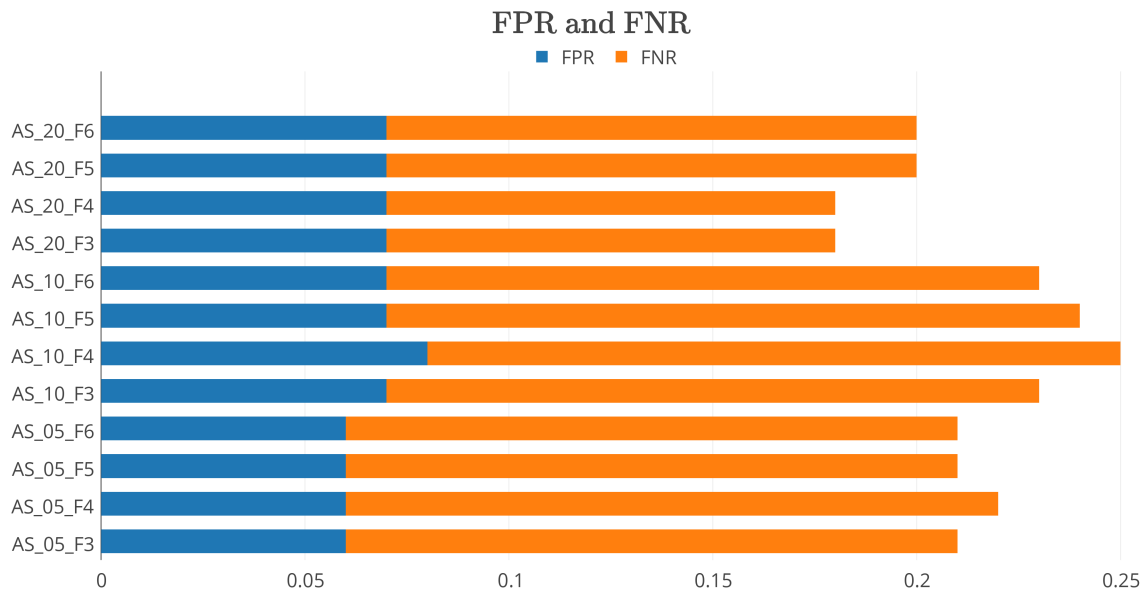
Figure 6.16: False Positive and Negative Rates of Random Forest with a variation over the number of features of the **AS** data sets.

In regard to false acceptance, the lowest values are achieved with 5 seconds PPG signals. However, the best cumulative rates, *i.e.* FPR and FNR combined, are registered with 20 seconds PPG records and having 3 and 4 features extracted.

MCC
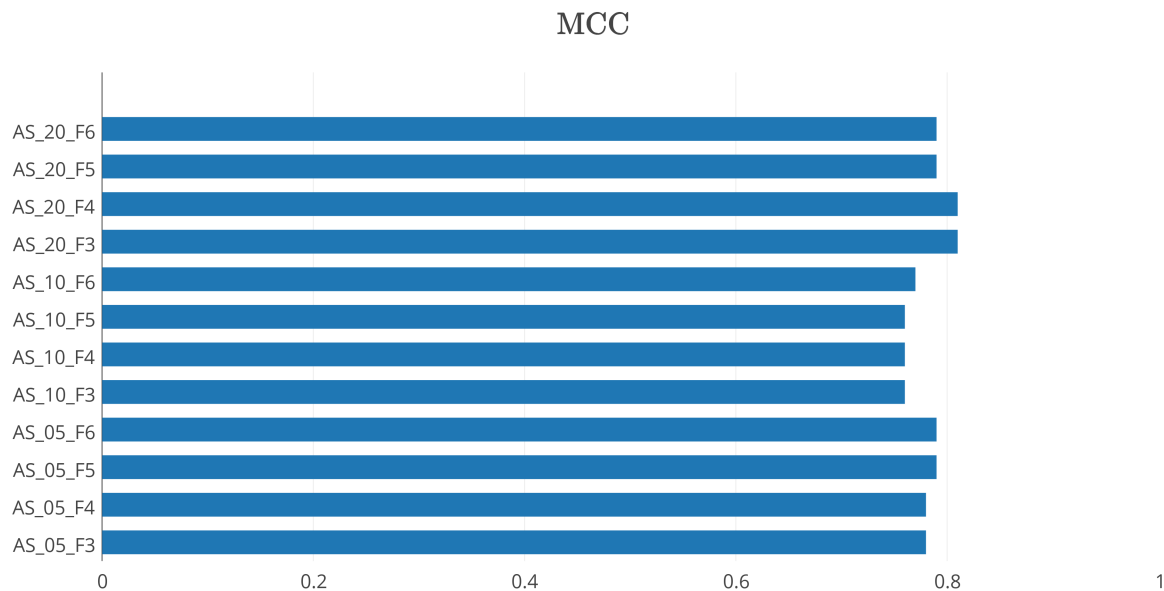
Figure 6.17: Matthews Correlation Coefficient of Random Forest with a variation over the number of features of the **AS** data sets.

Similarly to the results obtained with F1 Score, the best MCC values are obtained when using **AS_20_F3** and **AS_20_F4** data sets, surpassing the 0.8 mark.

Through the evaluation process previously exposed, some conclusions regarding relevant aspect of the proposed system were achieved. It is evident the supremacy of Random Forest when using the PPG waveform signals acquired from the *Angel Sensor*, thus making it the selected classifier to use in the authentication system. Furthermore, the number of features to extract when using the AC/DCT method was also determined. Despite the similarity of results obtained with **AS_20_F3** and **AS_20_F4** data sets, the latter presents a higher Precision value, which indicates a better ratio of positive values correctly predicted and all the values predicted as being positives, *i.e.* True Positive (TP) and False Positive (FP) values, leading to the decision of extracting 4 features for each signal derived from the PPG waves. As a final analysis, the representation of the Receiver Operating Characteristic (ROC) curve is illustrated in Figure 6.18.
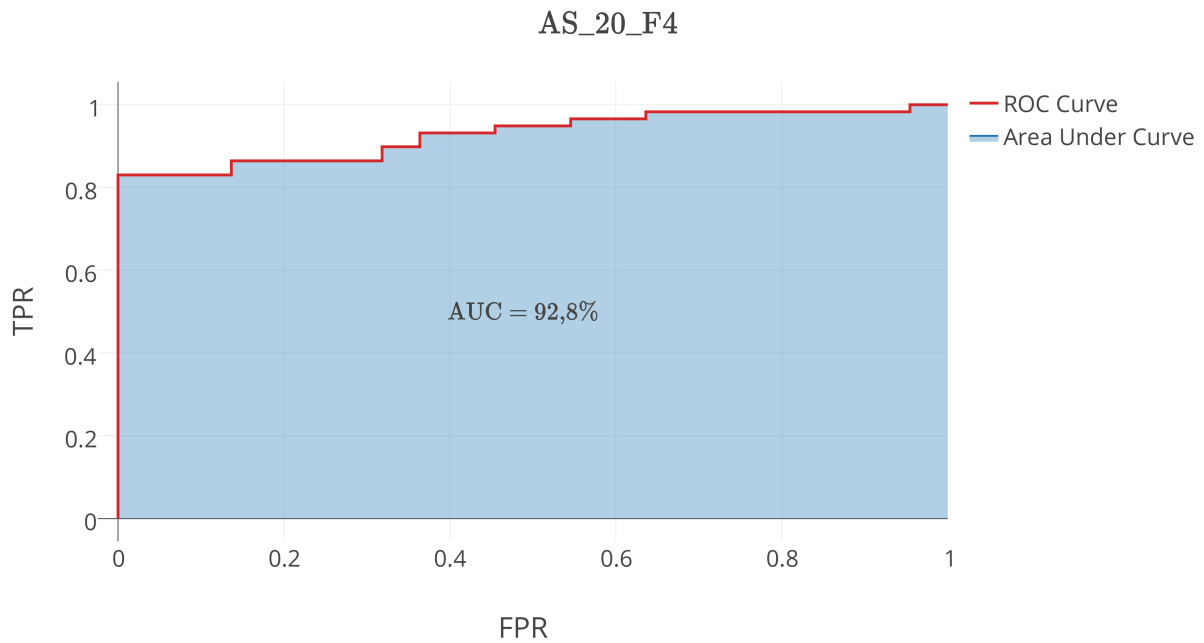
Figure 6.18: ROC curve of *Angel Sensor* data set with samples of 20 seconds, with 4 features extracted from each signal derived from PPG waveforms.

From the ROC curve representation in Figure 6.18, it is possible to identify the Area Under Curve (AUC), being 92.8%, which, despite not being perfect, it is considered to be good.

6.2.2   Fiducial Dependent Approach

Despite the quality of the PPG signals acquired from the *Angel Sensor* wristband lacking resolution, the fiducial approach described in Section 4.6 was also evaluated to determine its feasibility.

Tables A.26, A.27 and A.28 contain the metrics for the different models selected. Similarly to the results obtained when evaluating the **AS** data sets, it is visible that RF performs better with this specific data, thus being used in performed tests, illustrated in Figures 6.19, 6.20 and 6.21. Equivalently to the evaluation of the fiducial independent approach, the data sets used to consisted of PPG signals of 5, 10 and 20 seconds, referred as **AS_TF_05**, **AS_TF_10** and **AS_TF_20**, respectively.
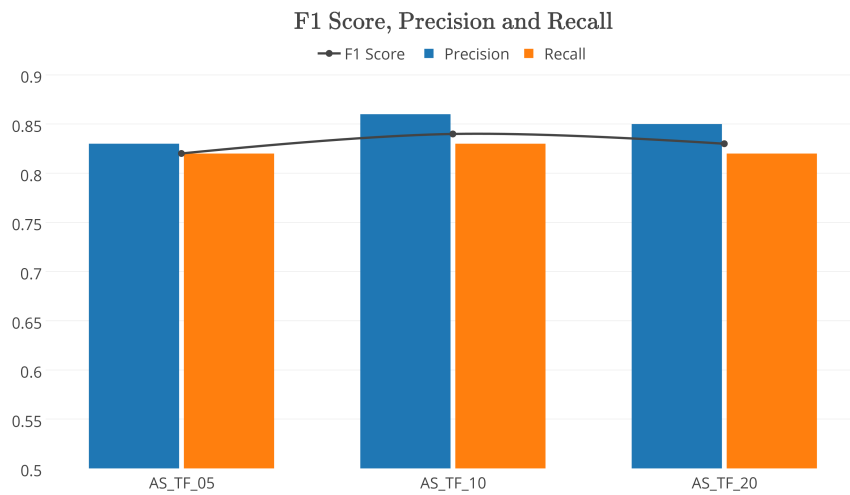
Figure 6.19: F1 Score to compare **AS_TF** data sets.

Figure 6.19 exposes the F1 Score, Precision and Recall of Random Forest for each data set. It is noticeable a higher score for the data sets of 10 and 20 seconds, where the prior presents highest values.
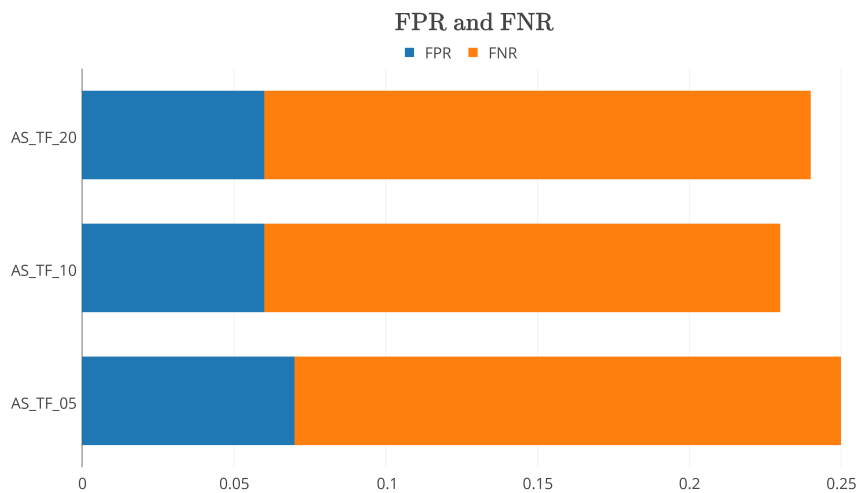
Figure 6.20: False Positive and Negative Rates of **AS_TF** data sets.

Regarding the False Positive and Negative Rates, the best rates are also achieved with PPG signals of 10 and 20 seconds. However, the **AS_TF_10** has a lower FNR than **AS_TF_20**.
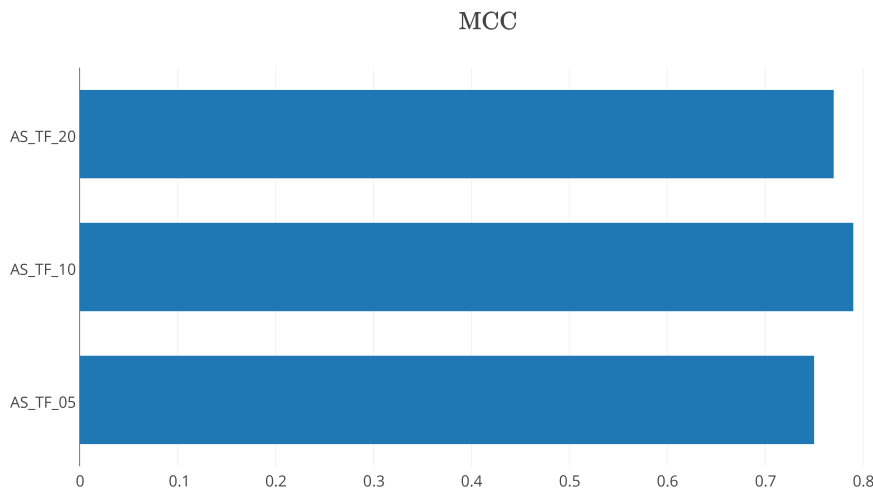
Figure 6.21: Matthews Correlation Coefficient of **AS_TF** data sets.

Similarly to the results obtained for the F1 Score and ,FPR and FNR, the data set with the best correlation coefficient is **AS_TF_10**, while **AS_TF_05** is has the lowest MCC value.

Unlike the results obtained in the fiducial independent approach, where the best outcome would rely on PPG signals of 20 seconds, the data set that allows a better performance comprises PPG signals of 10 seconds. Therefore, a final analysis considering all the data sets tested is required to correctly identify the better approach to implement.

6.2.3 AC/DCT vs. Time-Domain Features

The following illustrations comprise the aforementioned comparison between fiducial independent (AC/DCT) and fiducial dependent (Time-Domain Features) approaches. The evaluation metrics obtained are exposed in Table A.29.
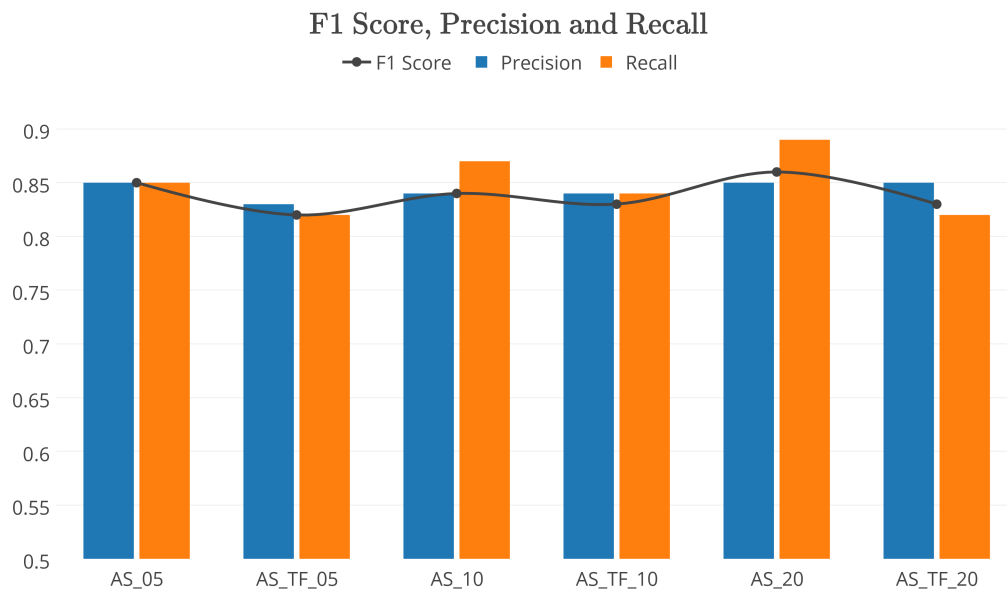
Figure 6.22: Comparison of the F1 Score between fiducial dependent (**AS** data sets) and fiducial independent approaches (**AS_TF** data sets).

Figure 6.22 exposed the F1 Score and respective Precision and Recall values of Random Forest when using testing **AS** and **AS_TF** data sets. It is visible that **AS** data sets outperform the **AS_TF** data sets, and the best score is obtained **AS_20**.
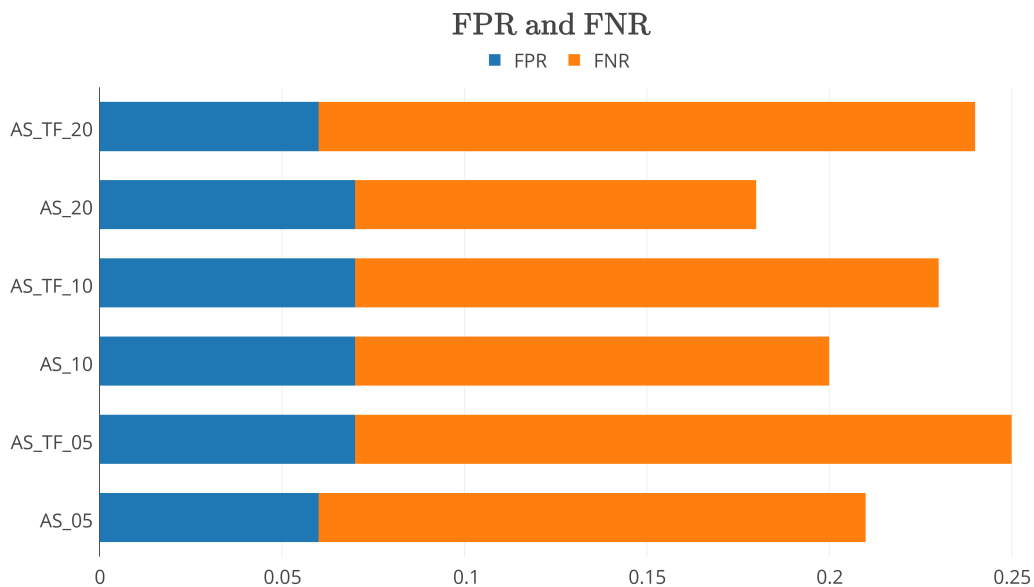
Figure 6.23: Comparison of False Positive and Negative Rates between fiducial dependent (**AS** data sets) and fiducial independent approaches (**AS_TF** data sets).

Regarding the FPR, both **AS__05** and **AS__TF__20** achieve the lowest rate of 5%. However, when considering also the FNR, the data set with the lowest cumulative rate is **AS__20**.

MCC
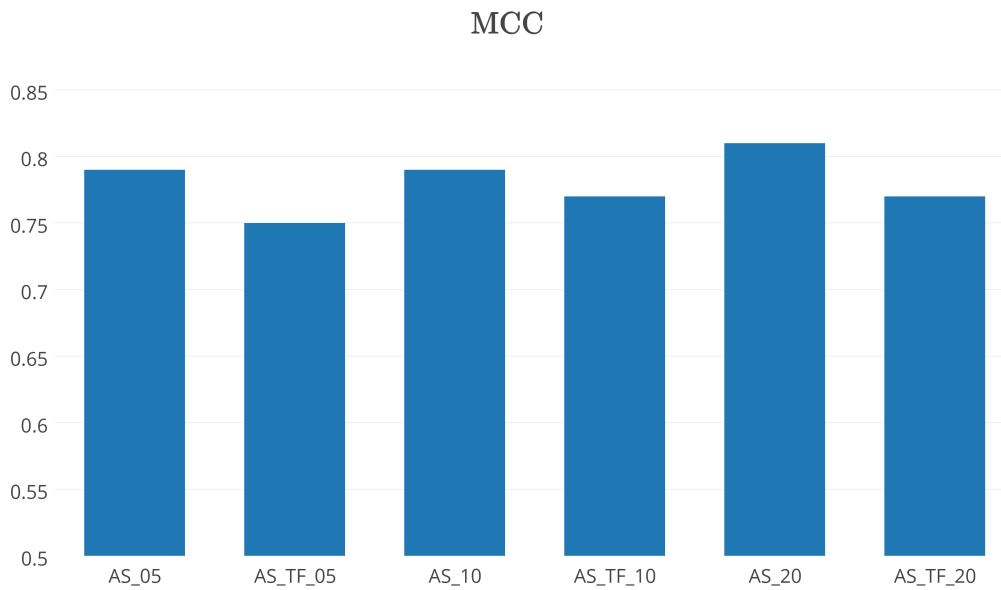
Figure 6.24: Comparison of Matthews Correlation Coefficient between fiducial dependent (**AS** data sets) and fiducial independent approaches (**AS__TF** data sets).

In Figure 6.24 the MCC is compared. Similarly to the results obtained when comparing the F1 Score, the best results were obtained with **AS__20**, surpassing the 0.8 mark.

From the comparative evaluation performed, it is evident that **AS** data sets outperform the **AS__TF** in every metric computed.

As a result, the conclusions obtained regarding the best configuration for the proposed system can be summarized to:

- **Approach** : Fiducial Independent (AC/DCT method)

- **Classifier** : Random Forest

- **Record Duration** : 20 seconds

- **Number of Features per signal** : 4

6.3   Discussion

Databases and data sets

After evaluating every classifier and the data sets, there are some considerations regarding the constraints that led to the results previously exposed. Sampling frequency of the sensors, the quality of the sensors, environmental interference, health conditions, among others, may have a high impact factor on the waveform signals studied.

In regard to the tests performed with data sets generated form the public data bases, *i.e.* **ECG_ID**, **ECG_TB** and **PPG_TB**, the main difference lies on the duration of the signals. In spite of having only 1 record of each waveform per individual, each sample of *CapnoBase* is 8 minutes long, opposing to the records of *PhysioNet* which lasts 20 seconds. However, each individual can have between 2 and 20 samples. Furthermore, the records of *PhysioNet* were collected in within different periods of time, which reflects into some minor variations of the wave signals of an individual, as exposed in [32]. In addition, the age group of the individuals is also relevant to consider. The **TBME** data set contain data from a group with age ranging 1 to 75 years old, and the majority of the signals belong to children which may have affected the results obtained.

Hardware and Sensors

Regarding the data obtained with the *Angel Sensor* wristband, the major constraints are related with the quality of the sensor. As referred in Section 4.6, the sampling rate was quite low to enable a decent definition of the waveform signal, contrasting with the sensors used in the public databases, where the sampling rate was of 500 and 300 Hz, respectively for the *PhysioNet* and *CapnoBase*, making the extraction of fiducial features impracticable. Moreover, signal would be highly affected by movement and, due to firmware implementation, some data would not be possible to extract, as the case of the heart rate.

Methodology

In regard to choices made in methodology, such as the order of the Finite Impulse Response (FIR) filters, they were based on a relationship between the minimum amount of data points to discard contingent on the quality of the signal to keep, *i.e.* the remaining signal should comprise minimal interference while keeping a decent amount of information to be evaluated. In order to guarantee at least 20 seconds of PPG signals, and having recorded 25 seconds worth of data, the maximum delay to cut was set to be 3 seconds (300 samples), guaranteeing a margin of 2 seconds.

Regarding the classifier used the main aspects considered were FPR and FNR, with focus on the prior. For instance, NB presented the lowest rate of FNR, however the FPR was extremely high,

the highest among all the classifiers. As the objective was an authentication system for a mobile platform, a system involving continuous authentication would be possible, as the wristband would be worn as a daily accessory, allowing a higher FNR.

When considering the amount of features to extract, the decision was based on experimentation, as exposed in Figure 4.2. not disregarding the fact that different classifiers might perform slightly better with a different amount of features.

Implementation

In terms of implementation, the choice over the buffers scheme might be debatable. It was, however, a consequence of the low quality of the data obtained from the sensor, otherwise the restrictions regarding the PPG data could be sightly relaxed which would be reflected in terms of computational performance. Moreover, all the decision derived from the limitations imposed by the hardware used, as the recordings were dependent on the sensor rate and quality, and the firmware implementation. However, and considering the hardware limitations, the approach followed produced acceptable results while not being too specific or too restrictive.

General Considerations

Globally, and despite the differences between hardware and the data sets, the outcome of this study was definitely positive. The main differences between all the results is directly affected by the quality of the data acquired, either from public databases or or recordings from the sensor. When the results to be produced are highly dependent on the data to be treated, the hardware component is very important, as it may impose heavy limitations regarding the objectives to be achieved.

Chapter 7

Conclusions

This work aimed to evaluate the feasibility of an authentication system based on waveform signals derived from the heartbeat. Several approaches exposed in the literature referred some unique traits that heartbeat related signals contain, in comparison with the remaining biometric traits. To ensure a higher usability, an application that allowed an abstraction of Heart Rate Variability (HRV) artifacts from Electrocardiogram (ECG) waveform signals, exposed in [38], was tested with Photoplethysmogram (PPG) signals.

Resourcing to two public databases which allowed the generation of three types of data sets, namely **ECG_ID**, **ECG_TB** and **PPG_TB**, and wristband with a PPG sensor that generated two more data sets, **AS** and **AS_TF**, several tests were performed to properly assess the feasibility of the desired approach. Among five different classifiers, Random Forest (RF) was selected to be applied to the system, since it provided the best evaluation metrics when dealing with the data obtained from the sensor. In order to determine how long the authentication session should last, in order to guarantee both, security and usability, signals with different duration were tested. Moreover, a range of features, from 3 to 6, was also tested to minimize the classification error while maintaining a low dimensionality of the data set.

In terms of results, the fiducial independent approach revealed to be slightly better than the approach using time-domain features. Using a waveform signal of 20 seconds, and extracting 4 features from each signal, *i.e.* two PPG signals, two 1st derivatives and two 2nd derivatives, from each signal, the system obtained an Area Under Curve (AUC) value of 92.8% and an accuracy of 91.78%.

When considering the feasibility of the proposed system, it might be a viable application depending on the relevance and the context of which such system would be applied. For a mobile application, and considering the user could be still for some seconds in order to perform a first authentication, it would be a decent alternative to standard authentication methods. Despite

having a considerably high value of false rejections, in a continuous authentication scenario, the user would not need to be demanded to perform the authentication process every time, *i.e.* after being successfully authenticated, the system could periodically confirm its identity without requiring any type of feedback.

However, the usage and security of the system is dependent on the hardware. During this experiment it was noticeable the interference of the sensor. If the sensor to use would be built from scratch, the limitations could be diminished, as the implementation would be distributed through all the cycle, specifically, issues related with the sensor itself could be more easily controlled.

7.1   Future Work

When considering a future perspective for the implementation performed, there are some aspects liable to be improved. The first milestone would be to port the implemented system to an application for a device running Android Operating System (OS).
Subsequently, the following improvements regarding specific stages of the application could be performed.

Biometric Data

Regarding the biometric data collected, it could be complemented with other forms of biometric information as security improvement factor. An example would be using behavioural biometrics, as exposed in [20], where, by using movements defined by the user, capturing "signature" gestures or even gesture patterns, a higher level of security could be achieved.

Pre-processing Stage

As previously mentioned, the data obtained from the *Angel Sensor* wristband was highly affected by the quality of the hardware and the conditions on which the recordings were performed, inducing high level of interference on the extracted signals. However, as exposed in [11], using an *Adaptive Noise Cancellation* filter based on the accelerometer data, interference derived from body movements could possibly be more efficiently filtered.

Enrollment

Regarding the enrollment stage, the implementation performed consisted of an offline training. However, and for the system to be used by any person, and not only the ones enrolled in the experiment a different type of learning would be required.
For a first version, when a user enrolls with the system, each biometric template will be stored into

a single data set which will contain records from several individuals to represent the "Intruders", so the classifier can have a base of comparison to learn from. When the recording session of the enrollment ends, the classifier will then be trained with the data set containing all the data.

In a second version, when trying to optimize the system, a technique referred in [18] as *online learning* could be used. It consists of using the data in a sequential order to improve the predictor. Concretely in this scenario, each time the user was validated during the authentication process, the data could be used to retrain the classifier to improve its future predictions.

Appendix A

Appended Tables

ECG_ID data sets generated from *PhysioNet* database

Model	Accuracy (%)	FPR	FNR	Precision	Recall	F-Measure	MCC
KNN	95.10	0.03	0.36	0.64	0.64	0.63	0.61
SVM	95.21	0.00	0.67	0.84	0.33	0.45	0.49
RF	95.87	0.01	0.57	0.84	0.43	0.55	0.58
LDA	79.32	0.20	0.31	0.19	0.69	0.30	0.29
NB	83.63	0.13	0.63	0.17	0.38	0.23	0.17

Table A.1: Evaluation Metrics of the five classifiers for the *PhysioNet* ECG records with duration of 5 seconds.

Model	Accuracy (%)	FPR	FNR	Precision	Recall	F-Measure	MCC
KNN	94.74	0.03	0.37	0.63	0.63	0.60	0.59
SVM	95.55	0.00	0.63	0.71	0.37	0.47	0.49
RF	96.26	0.00	0.53	0.90	0.47	0.59	0.62
LDA	78.74	0.21	0.18	0.21	0.82	0.34	0.35
NB	85.10	0.13	0.39	0.25	0.61	0.35	0.33

Table A.2: Evaluation Metrics of the five classifiers for the *PhysioNet* ECG records with duration of 10 seconds.

Model	Accuracy (%)	FPR	FNR	Precision	Recall	F-Measure	MCC
KNN	98.06	0.02	0.06	0.84	0.94	0.87	0.87
SVM	96.58	0.01	0.35	0.73	0.65	0.66	0.66
RF	97.10	0.00	0.42	0.76	0.58	0.64	0.65
LDA	82.84	0.18	0.01	0.29	0.99	0.44	0.48
NB	95.23	0.04	0.23	0.69	0.77	0.68	0.69

Table A.3: Evaluation Metrics of the five classifiers for the *PhysioNet* ECG records with duration of 20 seconds.

ECG_TB data sets generated from *CapnoBase* database

Model	Accuracy (%)	FPR	FNR	Precision	Recall	F-Measure	MCC
KNN	99.98	0.00	0.01	1.00	0.99	1.00	1.00
SVM	99.89	0.00	0.00	0.96	1.00	0.98	0.98
RF	99.94	0.00	0.03	1.00	0.97	0.99	0.99
LDA	97.02	0.03	0.00	0.45	1.00	0.62	0.66
NB	99.97	0.00	0.01	1.00	0.99	0.99	0.99

Table A.4: Evaluation Metrics of the five classifiers for the *CapnoBase* ECG records with duration of 5 seconds.

Model	Accuracy (%)	FPR	FNR	Precision	Recall	F-Measure	MCC
KNN	100	0.00	0.00	1.00	1.00	1.00	1.00
SVM	100	0.00	0.00	1.00	1.00	1.00	1.00
RF	99.99	0.00	0.00	1.00	1.00	1.00	1.00
LDA	95.53	0.05	0.00	0.35	1.00	0.52	0.58
NB	100	0.00	0.00	1.00	1.00	1.00	1.00

Table A.5: Evaluation Metrics of the five classifiers for the *CapnoBase* ECG records with duration of 10 seconds.

Model	Accuracy (%)	FPR	FNR	Precision	Recall	F-Measure	MCC
KNN	100	0.00	0.00	1.00	1.00	1.00	1.00
SVM	100	0.00	0.00	1.00	1.00	1.00	1.00
RF	99.90	0.00	0.04	1.00	0.96	0.98	0.98
LDA	96.32	0.04	0.00	0.40	1.00	0.57	0.62
NB	100	0.00	0.00	1.00	1.00	1.00	1.00

Table A.6: Evaluation Metrics of the five classifiers for the *CapnoBase* ECG records with duration of 20 seconds.

Model	Accuracy (%)	FPR	FNR	Precision	Recall	F-Measure	MCC
KNN	100	0.00	0.00	1.00	1.00	1.00	1.00
SVM	100	0.00	0.00	1.00	1.00	1.00	1.00
RF	100	0.00	0.00	1.00	1.00	1.00	1.00
LDA	97.68	0.02	0.00	0.54	1.00	0.69	0.72
NB	99.97	0.00	0.01	1.00	0.99	0.99	0.99

Table A.7: Evaluation Metrics of the five classifiers for the *CapnoBase* ECG records with duration of 30 seconds.

Model	Accuracy (%)	FPR	FNR	Precision	Recall	F-Measure	MCC
KNN	100	0.00	0.00	1.00	1.00	1.00	1.00
SVM	100	0.00	0.00	1.00	1.00	1.00	1.00
RF	99.83	0.00	0.07	1.00	0.93	0.95	0.96
LDA	97.06	0.03	0.00	0.47	1.00	0.63	0.67
NB	99.96	0.00	0.01	1.00	0.99	0.99	0.99

Table A.8: Evaluation Metrics of the five classifiers for the *CapnoBase* ECG records with duration of 40 seconds.

PPG_TB data sets generated from *CapnoBase* database

Model	Accuracy (%)	FPR	FNR	Precision	Recall	F-Measure	MCC
KNN	99.92	0.00	0.01	0.98	0.99	0.98	0.98
SVM	99.96	0.00	0.01	1.00	0.99	0.99	0.99
RF	99.87	0.00	0.05	1.00	0.95	0.97	0.97
LDA	97.33	0.03	0.00	0.48	1.00	0.65	0.68
NB	99.75	0.00	0.1	1.00	0.9	0.94	0.95

Table A.9: Evaluation Metrics of the five classifiers for the *CapnoBase* PPG records with duration of 5 seconds.

Model	Accuracy (%)	FPR	FNR	Precision	Recall	F-Measure	MCC
KNN	100	0.00	0.00	12.00	12.00	12.00	12.00
SVM	100	0.00	0.00	12.00	12.00	12.00	12.00
RF	99.77	0.00	0.07	0.98	0.93	0.95	0.95
LDA	98.4	0.02	0.00	0.60	12.00	0.75	0.77
NB	99.8	0.00	0.09	12.00	0.91	0.95	0.95

Table A.10: Evaluation Metrics of the five classifiers for the *CapnoBase* PPG records with duration of 10 seconds.

Model	Accuracy (%)	FPR	FNR	Precision	Recall	F-Measure	MCC
KNN	100	0.00	0.00	1.00	1.00	1.00	1.00
SVM	100	0.00	0.00	1.00	1.00	1.00	1.00
RF	99.81	0.00	0.08	1.00	0.92	0.95	0.96
LDA	98.53	0.02	0.00	0.65	1.00	0.78	0.80
NB	99.75	0.00	0.10	1.00	0.90	0.94	0.94

Table A.11: Evaluation Metrics of the five classifiers for the *CapnoBase* PPG records with duration of 20 seconds.

Model	Accuracy (%)	FPR	FNR	Precision	Recall	F-Measure	MCC
KNN	100	0.00	0.00	1.00	1.00	1.00	1.00
SVM	99.84	0.00	0.07	1.00	0.93	0.96	0.96
RF	99.62	0.00	0.16	0.96	0.84	0.88	0.89
LDA	98.57	0.01	0.00	0.68	1.00	0.79	0.81
NB	99.62	0.00	0.16	0.96	0.84	0.89	0.89

Table A.12: Evaluation Metrics of the five classifiers for the *CapnoBase* PPG records with duration of 30 seconds.

Model	Accuracy (%)	FPR	FNR	Precision	Recall	F-Measure	MCC
KNN	100	0.00	0.00	1.00	1.00	1.00	1.00
SVM	99.78	0.00	0.09	1.00	0.91	0.94	0.94
RF	99.74	0.00	0.09	0.99	0.91	0.93	0.94
LDA	98.70	0.01	0.00	0.68	1.00	0.80	0.82
NB	99.74	0.00	0.11	1.00	0.89	0.93	0.93

Table A.13: Evaluation Metrics of the five classifiers for the *CapnoBase* PPG records with duration of 40 seconds.

AS data sets created using the *Angel Sensor* wristband

5 seconds records

Model	Accuracy (%)	FPR	FNR	Precision	Recall	F-Measure	MCC	AUC
KNN	88.49	0.08	0.22	0.81	0.78	0.79	0.72	0.85
SVM	86.98	0.1	0.21	0.76	0.79	0.78	0.69	0.94
RF	90.98	0.06	0.15	0.84	0.85	0.84	0.78	0.97
LDA	85.91	0.16	0.08	0.69	0.92	0.79	0.70	0.95
NB	84.48	0.20	0.05	0.66	0.95	0.78	0.69	0.93

Table A.14: Evaluation Metrics of the classifiers for the *Angel Sensor* PPG records with duration of 5 seconds, using three features per signal, *i.e.* PPG signals from LEDs green and blue, and the first and second derivatives of each PPG signal.

Model	Accuracy (%)	FPR	FNR	Precision	Recall	F-Measure	MCC	AUC
KNN	87.03	0.08	0.26	0.79	0.74	0.76	0.68	0.82
SVM	86.86	0.10	0.21	0.76	0.79	0.77	0.68	0.94
RF	91.03	0.06	0.16	0.84	0.84	0.84	0.78	0.97
LDA	86.37	0.15	0.09	0.70	0.91	0.79	0.71	0.95
NB	84.75	0.19	0.06	0.66	0.94	0.78	0.69	0.93

Table A.15: Evaluation Metrics of the classifiers for the *Angel Sensor* PPG records with duration of 5 seconds, using four features per signal, *i.e.* PPG signals from LEDs green and blue, and the first and second derivatives of each PPG signal.

Model	Accuracy (%)	FPR	FNR	Precision	Recall	F-Measure	MCC	AUC
KNN	88.09	0.08	0.22	0.80	0.78	0.79	0.71	0.84
SVM	86.94	0.10	0.21	0.76	0.79	0.77	0.68	0.94
RF	91.38	0.06	0.15	0.85	0.85	0.85	0.79	0.97
LDA	87.11	0.14	0.10	0.72	0.90	0.80	0.72	0.95
NB	84.58	0.19	0.05	0.66	0.95	0.78	0.69	0.93

Table A.16: Evaluation Metrics of the classifiers for the *Angel Sensor* PPG records with duration of 5 seconds, using five features per signal, *i.e.* PPG signals from LEDs green and blue, and the first and second derivatives of each PPG signal.

Model	Accuracy (%)	FPR	FNR	Precision	Recall	F-Measure	MCC	AUC
KNN	87.35	0.09	0.23	0.78	0.77	0.78	0.69	0.83
SVM	87.06	0.10	0.21	0.77	0.79	0.77	0.69	0.94
RF	91.37	0.06	0.15	0.85	0.85	0.85	0.79	0.97
LDA	86.83	0.14	0.12	0.72	0.88	0.79	0.71	0.95
NB	84.60	0.19	0.06	0.66	0.94	0.78	0.69	0.94

Table A.17: Evaluation Metrics of the classifiers for the *Angel Sensor* PPG records with duration of 5 seconds, using six features per signal, *i.e.* PPG signals from LEDs green and blue, and the first and second derivatives of each PPG signal.

10 seconds records

Model	Accuracy (%)	FPR	FNR	Precision	Recall	F-Measure	MCC	AUC
KNN	87.29	0.09	0.23	0.79	0.77	0.77	0.69	0.82
SVM	86.31	0.11	0.21	0.75	0.79	0.77	0.67	0.94
RF	90.15	0.07	0.16	0.83	0.84	0.83	0.76	0.96
LDA	87.60	0.13	0.10	0.73	0.90	0.81	0.73	0.95
NB	85.38	0.18	0.07	0.68	0.93	0.79	0.7	0.94

Table A.18: Evaluation Metrics of the classifiers for the *Angel Sensor* PPG records with duration of 10 seconds, using three features per signal, *i.e.* PPG signals from LEDs green and blue, and the first and second derivatives of each PPG signal.

Model	Accuracy (%)	FPR	FNR	Precision	Recall	F-Measure	MCC	AUC
KNN	87.14	0.08	0.24	0.79	0.76	0.77	0.68	0.81
SVM	86.37	0.11	0.20	0.75	0.80	0.77	0.68	0.94
RF	89.85	0.08	0.17	0.83	0.83	0.82	0.76	0.96
LDA	87.02	0.14	0.11	0.73	0.89	0.80	0.72	0.96
NB	85.51	0.18	0.05	0.68	0.95	0.79	0.71	0.94

Table A.19: Evaluation Metrics of the classifiers for the *Angel Sensor* PPG records with duration of 10 seconds, using four features per signal, *i.e.* PPG signals from LEDs green and blue, and the first and second derivatives of each PPG signal.

Model	Accuracy (%)	FPR	FNR	Precision	Recall	F-Measure	MCC	AUC
KNN	88.00	0.07	0.24	0.81	0.76	0.78	0.70	0.82
SVM	87.32	0.10	0.18	0.76	0.82	0.79	0.70	0.94
RF	90.03	0.07	0.17	0.83	0.83	0.83	0.76	0.96
LDA	86.86	0.13	0.13	0.73	0.87	0.79	0.71	0.96
NB	85.02	0.19	0.05	0.67	0.95	0.78	0.70	0.94

Table A.20: Evaluation Metrics of the classifiers for the *Angel Sensor* PPG records with duration of 10 seconds, using five features per signal, *i.e.* PPG signals from LEDs green and blue, and the first and second derivatives of each PPG signal.

Model	Accuracy (%)	FPR	FNR	Precision	Recall	F-Measure	MCC	AUC
KNN	88.58	0.08	0.20	0.80	0.80	0.80	0.72	0.85
SVM	87.11	0.10	0.20	0.77	0.80	0.78	0.69	0.95
RF	90.46	0.07	0.16	0.84	0.84	0.83	0.77	0.96
LDA	87.11	0.13	0.13	0.74	0.87	0.79	0.71	0.96
NB	84.25	0.20	0.06	0.66	0.94	0.77	0.69	0.94

Table A.21: Evaluation Metrics of the classifiers for the *Angel Sensor* PPG records with duration of 10 seconds, using six features per signal, *i.e.* PPG signals from LEDs green and blue, and the first and second derivatives of each PPG signal.

20 seconds records

Model	Accuracy (%)	FPR	FNR	Precision	Recall	F-Measure	MCC	AUC
KNN	87.33	0.07	0.26	0.80	0.74	0.76	0.68	0.81
SVM	85.88	0.12	0.19	0.73	0.81	0.76	0.67	0.94
RF	91.79	0.07	0.11	0.84	0.89	0.86	0.81	0.97
LDA	87.18	0.13	0.12	0.74	0.88	0.80	0.72	0.96
NB	86.38	0.17	0.05	0.70	0.95	0.80	0.73	0.93

Table A.22: Evaluation Metrics of the classifiers for the *Angel Sensor* PPG records with duration of 20 seconds, using three features per signal, *i.e.* PPG signals from LEDs green and blue, and the first and second derivatives of each PPG signal.

Model	Accuracy (%)	FPR	FNR	Precision	Recall	F-Measure	MCC	AUC
KNN	85.39	0.1	0.25	0.76	0.75	0.74	0.65	0.82
SVM	86.86	0.11	0.19	0.76	0.81	0.77	0.69	0.94
RF	91.78	0.07	0.11	0.85	0.89	0.86	0.81	0.97
LDA	87.97	0.13	0.11	0.75	0.89	0.81	0.74	0.95
NB	85.04	0.18	0.07	0.68	0.93	0.78	0.70	0.93

Table A.23: Evaluation Metrics of the classifiers for the *Angel Sensor* PPG records with duration of 20 seconds, using four features per signal, *i.e.* PPG signals from LEDs green and blue, and the first and second derivatives of each PPG signal.

Model	Accuracy (%)	FPR	FNR	Precision	Recall	F-Measure	MCC	AUC
KNN	87.60	0.08	0.24	0.81	0.76	0.78	0.70	0.83
SVM	85.32	0.12	0.21	0.73	0.79	0.75	0.66	0.94
RF	91.11	0.07	0.13	0.84	0.87	0.85	0.79	0.97
LDA	88.14	0.11	0.14	0.77	0.86	0.80	0.73	0.96
NB	83.98	0.20	0.06	0.67	0.94	0.77	0.69	0.93

Table A.24: Evaluation Metrics of the classifiers for the *Angel Sensor* PPG records with duration of 20 seconds, using five features per signal, *i.e.* PPG signals from LEDs green and blue, and the first and second derivatives of each PPG signal.

Model	Accuracy (%)	FPR	FNR	Precision	Recall	F-Measure	MCC	AUC
KNN	87.66	0.10	0.17	0.77	0.83	0.79	0.71	0.84
SVM	86.80	0.10	0.21	0.77	0.79	0.77	0.69	0.94
RF	91.05	0.07	0.13	0.84	0.87	0.84	0.79	0.97
LDA	87.53	0.12	0.13	0.75	0.87	0.80	0.72	0.96
NB	83.25	0.20	0.07	0.66	0.93	0.76	0.67	0.94

Table A.25: Evaluation Metrics of the classifiers for the *Angel Sensor* PPG records with duration of 20 seconds, using six features per signal, *i.e.* PPG signals from LEDs green and blue, and the first and second derivatives of each PPG signal.

AS_TF data sets created using the *Angel Sensor* wristband

Model	Accuracy (%)	FPR	FNR	Precision	Recall	F-Measure	MCC
KNN	86.11	0.08	0.29	0.78	0.71	0.74	0.65
SVM	86.29	0.09	0.25	0.77	0.75	0.76	0.66
RF	89.88	0.07	0.18	0.83	0.82	0.82	0.75
LDA	87.37	0.12	0.13	0.74	0.87	0.80	0.71
NB	84.11	0.18	0.10	0.66	0.90	0.76	0.67

Table A.26: Evaluation Metrics of the five classifiers for the *Angel Sensor* PPG records with of 5 seconds, using temporal features.

Model	Accuracy (%)	FPR	FNR	Precision	Recall	F-Measure	MCC
KNN	88.62	0.07	0.23	0.82	0.77	0.79	0.72
SVM	86.89	0.10	0.21	0.76	0.79	0.77	0.69
RF	91.26	0.06	0.17	0.86	0.83	0.84	0.79
LDA	87.48	0.12	0.13	0.74	0.87	0.80	0.72
NB	85.23	0.17	0.08	0.68	0.92	0.78	0.69

Table A.27: Evaluation Metrics of the five classifiers for the *Angel Sensor* PPG records with of 10 seconds, using temporal features.

Model	Accuracy (%)	FPR	FNR	Precision	Recall	F-Measure	MCC
KNN	88.39	0.07	0.23	0.82	0.77	0.79	0.72
SVM	86.74	0.10	0.21	0.76	0.79	0.77	0.68
RF	90.49	0.06	0.18	0.85	0.82	0.83	0.77
LDA	87.16	0.12	0.15	0.75	0.85	0.79	0.71
NB	86.87	0.15	0.07	0.71	0.93	0.80	0.73

Table A.28: Evaluation Metrics of the five classifiers for the *Angel Sensor* PPG records with of 20 seconds, using temporal features.

Metrics comparison between AS and AS_TF data sets

Data sets	Accuracy (%)	FPR	FNR	Precision	Recall	F-Measure	MCC
AS_05	91.38	0.06	0.15	0.85	0.85	0.85	0.79
AS_TF_05	89.88	0.07	0.18	0.83	0.82	0.82	0.75
AS_10	91.05	0.07	0.13	0.84	0.87	0.84	0.79
AS_TF_10	90.46	0.07	0.16	0.84	0.84	0.83	0.77
AS_20	91.78	0.07	0.11	0.85	0.89	0.86	0.81
AS_TF_20	90.49	0.06	0.18	0.85	0.82	0.83	0.77

Table A.29: Evaluation Metrics comparison between fiducial dependent and independent approaches of the Random Forest classifier for *Angel Sensor* data sets. Three data sets of each approach are compared varying the duration of the signal, *i.e.* 5, 10 and 20 seconds.

Bibliography

[1] Foteini Agrafioti. *ECG in biometric recognition: Time dependency and application challenges.* PhD thesis, University of Toronto, 2011.

[2] Nazneen Akhter, Hanumant Gite, Gulam Rabbiani, and Karbhari Kale. *Heart Rate Variability for Biometric Authentication Using Time-Domain Features*, volume 377. Springer, 2015. ISBN: 978-3-642-40575-4. doi:10.1007/978-3-642-40576-1.

[3] AngioScan-Electronics. Schematic representation of an optical sensor mounted on the terminal phalanx of the finger. (Online; accessed march 15, 2016).

[4] Lena Biel, Ola Pettersson, Lennart Philipson, and Peter Wide. Ecg analysis: a new approach in human identification. *Instrumentation and Measurement, IEEE Transactions on*, 50(3): 808–812, 2001.

[5] Angelo Bonissi, Ruggero Donida Labati, Luca Perico, Roberto Sassi, Fabio Scotti, and Luca Sparagino. A preliminary study on continuous authentication methods for photoplethysmographic biometrics. *2013 IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications, BioMS 2013 - Proceedings*, (March 2016): 28–33, 2013. doi:10.1109/BIOMS.2013.6656145.

[6] Hans Werner Borchers. *pracma: Practical Numerical Math Functions*, 2016. R package version 1.9.3.

[7] Roberto Brunelli and Tomaso Poggio. Face recognition: Features versus templates. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (10):1042–1052, 1993.

[8] Carlos Carreiras, André Lourenço, Ana Fred, and Rui Ferreira. Ecg signals for biometric applications-are we there yet? In *Informatics in Control, Automation and Robotics (ICINCO), 2014 11th International Conference on*, volume 2, pages 765–772. IEEE, 2014.

[9] Chuang-Chien Chiu, Chou-Min Chuang, and Chih-Yu Hsu. A novel personal identity verification approach using a discrete wavelet transform of the ecg signal. In *Multimedia and Ubiquitous Engineering, 2008. MUE 2008. International Conference on*, pages 201–206. IEEE, 2008.

[10] V. S. Chouhan and S. S. Mehta. Total removal of baseline drift from ECG signal. *Proceedings - International Conference on Computing: Theory and Applications, ICCTA 2007*, pages 512–515, 2007. doi:10.1109/ICCTA.2007.126.

[11] Gary W Comtois. *Implementation of Accelerometer-Based Adaptive Noise Cancellation in a Wireless Wearable Pulse Oximeter Platform for Remote Physiological Monitoring and Triage.* PhD thesis, Worcester Polytechnic Institute, 2007.

[12] D Dumn. Using a multi-layer perceptron neural for human voice identification. In *Proc. Fourth Int. Conf. Signal Process. Applicat. Technol*, 1993.

[13] John Wesley Eaton, David Bateman, and Søren Hauberg. *Gnu octave.* Network thoery London, 1997.

[14] Mohamed Elgendi. On the analysis of fingertip photoplethysmogram signals. *Current cardiology reviews*, 8(1):14–25, 2012. ISSN: 1875-6557. doi:10.2174/157340312801215782.

[15] S. Zahra Fatemian and Dimitrios Hatzinakos. A new ECG feature extractor for biometric recognition. *DSP 2009: 16th International Conference on Digital Signal Processing, Proceedings*, 2009. doi:10.1109/ICDSP.2009.5201143.

[16] S Zahra Fatemian, Foteini Agrafioti, and Dimitrios Hatzinakos. Heartid: Cardiac biometric recognition. In *Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on*, pages 1–5. IEEE, 2010.

[17] Proyecto Fin De Máster, Autor : Víctor, and Barbero Romero. ECG baseline wander removal and noise suppression analysis in an embedded platform. 2008.

[18] Óscar Fontenla-Romero, Bertha Guijarro-Berdiñas, David Martinez-Rego, Beatriz Pérez-Sánchez, and Diego Peteiro-Barral. Online machine learning. *Efficiency and Scalability Methods for Computational Intellect*, pages 27–54, 2013.

[19] Robert W Frischholz and Ulrich Dieckmann. Biold: a multimodal biometric identification system. *Computer*, 33(2):64–68, 2000.

[20] Hugo Gamboa. A behavioral biometric system based on human-computer interaction. *Proceedings of SPIE*, (i):381–392, 2004. ISSN: 0277786X. doi:10.1117/12.542625.

[21] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. Ch. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 2000 (June 13). Circulation Electronic Pages: http://circ.ahajournals.org/cgi/content/full/101/23/e215 PMID:1085218; doi: 10.1161/01.CIR.101.23.e215.

[22] YY Gu, Y Zhang, and YT Zhang. A novel biometric approach in human verification by photoplethysmographic signals. *Information Technology Applications in Biomedicine, 2003. 4th International IEEE EMBS Special Topic Conference on*, 2003.

[23] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H Witten. The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18, 2009.

[24] Google Inc. Android api, 2016. (Online; accessed April 10, 2016).

[25] Google Inc. Googlefit api, 2016. (Online; accessed April 15, 2016).

[26] Steven A Israel, John M Irvine, Andrew Cheng, Mark D Wiederhold, and Brenda K Wiederhold. Ecg to identify individuals. *Pattern recognition*, 38(1):133–142, 2005.

[27] Alan Julian Izenman. *Modern multivariate statistical techniques*, volume 1. Springer, 2008.

[28] W. Karlen, M. Turner, E. Cooke, G. Dumont, and J. M. Ansermino. Capnobase: Signal database and tools to collect, share and annotate respiratory signals. In *Annual Meeting of the Society for Technology in Anesthesia (STA)*, West Palm Beach, 2010.

[29] Lukasz Komsta. *dtt: Discrete Trigonometric Transforms*, 2013. R package version 0.1-2.

[30] André Lourenço, Carlos Carreiras, Hugo Silva, and Ana Fred. Ecg biometrics: A template selection approach. In *Medical Measurements and Applications (MeMeA), 2014 IEEE International Symposium on*, pages 1–6. IEEE, 2014.

[31] Seraphim Sense Ltd. Angel sensor api, 2015. (Online; accessed May 1, 2016).

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

[33] Tsutomu Matsumoto, Hiroyuki Matsumoto, Koji Yamada, and Satoshi Hoshino. Impact of artificial gummy fingers on fingerprint systems. In *Electronic Imaging 2002*, pages 275–289. International Society for Optics and Photonics, 2002.

[34] Michael Negin, M Salganicoff, and Grace G Zhang. An iris biometric system for public and personal use. *Computer*, 33(2):70–75, 2000.

[35] Clifford V Nelson and David B Geselowitz. *The theoretical basis of electrocardiology*, volume 1. Oxford University Press, 1976.

[36] S original by Jim Ramsey. R port by Brian Ripley <ripley@stats.ox.ac.uk>. *pspline: Penalized Smoothing Splines*, 2015. R package version 1.0-17.

[37] P Jonathon Phillips, Alvin Martin, Charles L Wilson, and Mark Przybocki. An introduction evaluating biometric systems. *Computer*, 33(2):56–63, 2000.

[38] Konstantinos N Plataniotis, Dimitrios Hatzinakos, and Jimmy KM Lee. Ecg biometric recognition without fiducial detection. In *Biometric Consortium Conference, 2006 Biometrics Symposium: Special Session on Research at the*, pages 1–6. IEEE, 2006.

[39] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2015.

[40] R. Ruiz-Gonzalez and *et al.*. Representation of a support vector machine (svm)), 2014. (Online; accessed march 10, 2016).

[41] Ashok Samal and Prasana A Iyengar. Automatic recognition and analysis of human faces and facial expressions: A survey. *Pattern recognition*, 25(1):65–77, 1992.

[42] Tsu-Wang Shen, WJ Tompkins, and YH Hu. One-lead ecg for identity verification. In *Engineering in Medicine and Biology, 2002. 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society EMBS/BMES Conference, 2002. Proceedings of the Second Joint*, volume 1, pages 62–63. IEEE, 2002.

[43] Tsu-Wang Shen, Willis J Tompkins, and Yu Hen Hu. Implementation of a one-lead ecg human identification system on a normal population. *Journal of Engineering and Computer Innovations*, 2(1):12–21, 2011.

[44] signal developers. *signal: Signal processing*, 2014.

[45] Yogendra Narain Singh, Sanjay Kumar Singh, and Phalguni Gupta. Fusion of electrocardiogram with unobtrusive biometrics: An efficient individual authentication system. *Pattern Recognition Letters*, 33(14):1932–1941, 2012.

[46] Alex Smola and Bernhard Schölkopf. Kernel machines website. http://www.kernel-machines.org/, 2007. (Online; accessed July 14, 2016).

[47] Petros Spachos, Jiexin Gao, and Dimitrios Hatzinakos. Feasibility study of photoplethysmographic signals for biometric identification. *17th DSP 2011 International Conference on Digital Signal Processing, Proceedings*, (March 2016), 2011. ISSN: Pending. doi:10.1109/ICDSP.2011.6004938.

[48] Ya-Ting Tsao, Tsu-Wang Shen, Tung-Fu Ko, and Tsung-Hsing Lin. The morphology of the electrocardiogram for eevaluating ecg biometrics. In *e-Health Networking, Application and Services, 2007 9th International Conference on*, pages 233–235. IEEE, 2007.

[49] Yongjin Wang, Foteini Agrafioti, Dimitrios Hatzinakos, and Konstantinos N Plataniotis. Analysis of human electrocardiogram for biometric recognition. *EURASIP journal on Advances in Signal Processing*, 2008:19, 2008.

[50] Chen Wei, Lei Sheng, Guo Lihua, Chen Yuquan, and Pan Min. Study on conditioning and feature extraction algorithm of photoplethysmography signal for physiological parameters detection. *Proceedings - 4th International Congress on Image and Signal Processing, CISP 2011*, 4:2194–2197, 2011. doi:10.1109/CISP.2011.6100581.

[51] Psychology Wiki. Electrocardiography. (Online; accessed march 10, 2016).

[52] Ja-Ling Wu. Discrete cosine transformation (dct), 2016. (Online; accessed June 1, 2016).

[53] Gerd Wübbeler, Manuel Stavridis, Dieter Kreiseler, Ralf-Dieter Bousseljot, and Clemens Elster. Verification of humans using the electrocardiogram. *Pattern Recognition Letters*, 28 (10):1172–1175, 2007.

[54] Jianchu Yao, Xiaodong Sun, and Yongbo Wan. A pilot study on using derivatives of photoplethysmographic signals as a biometric identifier. *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings*, 2007:4576–4579, jan 2007. ISSN: 05891019. doi:10.1109/IEMBS.2007.4353358.