Faculdade de Engenharia da Universidade do Porto

FEUP

Performance assessment and prediction of football players: Tailoring an architecture with spatiotemporal positional and physiological features

Sofia de Sousa Almeida

Dissertação realizada no âmbito do Mestrado Integrado em Bioengenharia
Ramo de Engenharia Biomédica

Orientador: Prof. Dr. Jaime Cardoso
Co-orientador: Prof. Dr. Micael Couceiro

julho, 2017

Resumo

O futebol foi organizado pela primeira vez na Inglaterra em 1863 e rapidamente se espalhou para o resto do mundo, apresentando-se como uma das primeiras manifestações da globalização. Hoje em dia, o futebol não é mais apenas um desporto, desempenhando um papel importante nos atuais mecanismos gerais das nossas vidas socioeconómicas. Portanto, como se pode perceber, prever o resultado final de um jogo, decodificar o desempenho das equipas e, adicionalmente, prevenir potenciais ferimentos dos atletas, é visto pela comunidade científica como uma espécie de "Santo Graal". O estado da arte mostra que embora muitos autores tenham apresentado metodologias para entender e, até certo ponto, prever os resultados de um jogo de futebol, nenhum deles explora todas as características relevantes que influenciam o desempenho dos atletas e, por consequência, podem determinar o sucesso, ou falta dele, durante um jogo de futebol. Como tal, nesta tese de mestrado propomos o desenvolvimento de uma arquitetura para a previsão no futebol, capaz de estimar vários parâmetros que influenciam grandemente o funcionamento normal de um jogo de futebol, incluindo a posição do atleta, as ações e o estado de saúde. Para este fim, sinais provenientes de várias fontes, nomeadamente, sinais cinemáticos e fisiológicos, serão obtidos a partir de dois dispositivos *wearable* e integrados na arquitetura final.

Os métodos do Cálculo Fracionário e das Redes Neurais Recorrentes foram utilizados para estimar as coordenadas dos atletas ao longo do tempo, concluindo que o último leva a erros menores. Por outro lado, a classificação das ações dos atletas e a deteção da fadiga muscular foram realizadas usando dados fisiológicos, tendo sido estes recolhidos por um novo dispositivo portátil *TraXports V2*, aqui desenvolvido. Em relação à classificação das ações, foi demonstrado que quando comparado a classificadores simples, como Naïve Bayes, Redes Neurais Artificiais, Máquinas de Vetor de Suporte e o K-Vizinho Mais Próximo, a abordagem proposta, o método de fusão dinâmica de modelos, aumentou claramente o desempenho geral dos resultados da classificação, com uma precisão total de 76% e uma sensibilidade total de 74%. No que diz respeito à deteção da fadiga, foi possível visualizar as diferenças existentes entre os músculos não fatigados e fatigados, corroborando a utilidade e a viabilidade do dispositivo *TraXports V2*.

Os resultados obtidos mostraram o grande potencial da arquitetura para a previsão de futebol, no entanto, algumas modificações ainda precisam ser implementadas para criar um sistema completamente confiável.

Abstract

Association football was first organized in England in 1863 and spread rapidly to the rest of the world, presenting itself as one of the first manifestations of globalization. Nowadays, football is not just a sport anymore, playing an important role in the current overall mechanisms of our socio-economic lives. Therefore, as one may realize, predicting the final outcome of a match, decoding the teams' performance, and additionally preventing any potential athletes' injuries, is seen by the scientific community as a kind of 'Holy Grail'. The state of the art shows that although many authors have been presenting methodologies to understand and, to some extent, predict the football match outcomes, none of them explores all the relevant features that influence athletes' performance and that, by consequence, might determine the success, or lack of it, during a football match.

As such, in this Master thesis we propose the development of a framework for football prediction, capable of estimating various parameters that highly influence the normal functioning of a football match, including athlete's position, actions and muscle fatigue. For this aim, signals from several sources, namely kinematic and physiological signals, were retrieved from two wearable devices and integrated in the final architecture.

Fractional Calculus and Recurrent Neural Networks were used to estimate the athletes' coordinates over time, concluding that the latter leads to minor errors. On the other hand, the classification of athletes' actions and the detection of muscle fatigue was performed by using physiological data, which was collected by a new wearable device TraXports V2, herein developed.

Regarding actions' classification it was demonstrated that when compared to simple classifiers such as Naïve Bayes, Artificial Neural Networks, Support Vector Machines and K-nearest Neighbour, the proposed approach, Dynamic Bayesian Mixture Model, clearly increased the overall performance of the classification results, with a total precision of 76% and a total recall of 74%. As regards of fatigue detection, it was possible to visualize the existent differences between a non-fatigued and a fatigued-muscle, corroborating the utility and viability of the device TraXports V2.

The results obtained showed the great potential of the architecture for football prediction, however, further improvements are still needed to be implemented in order to create a completely reliable system.

Agradecimentos

Ao Professor Doutor Micael Couceiro ao qual quero deixar um grande agradecimento por me ter recebido e por ter apostado em mim e nas minhas capacidades. Gostaria de agradecer pelo seu total apoio, disponibilidade, orientação e dedicação. Sem dúvida uma pessoa fantástica.

Ao Professor Doutor Jaime Cardoso pela sua orientação, pelo seu apoio, pela sua total colaboração na resolução de dúvidas e problemas e pela sua simpatia.

Aos meus amigos da Ingeniarius, André Araújo, Samuel Pereira e David Portugal, que estiveram sempre prontos para ajudar, mesmo estando ocupados, e que também foram essenciais para o sucesso desta tese.

A todos os amigos que me ajudaram na recolha de dados, Elisabete Santos, Diana Pinto, Manuel Silva, Rui Silva, Miguel Fange, Frederico Vaz e Diogo Ramalho.

Ao Ginásio Fange e ao Ginásio do Rui Silva por me terem disponibilizado o espaço durante diversas vezes para fazer a recolha de dados.

Aos meus amigos Rita, Beatriz e André que sempre estiveram lá para me encorajar e nunca desmotivar.

Ao Tiago, ao qual faltam palavras para agradecer o quanto foi essencial neste percurso. Pelo seu apoio incondicional, pela sua paciência, encorajamento e por me fazer sempre ver as coisas pelo lado positivo, um grande obrigada do fundo do coração.

Finalmente, um especial agradecimento aos meus pais que sempre me apoiaram em todas as minhas decisões e sempre me incentivaram para ser a melhor versão de mim mesma. Por serem modelos de empenho e coragem a eles dedico este trabalho.

Acknowledgments

To Professor Doctor Micael Couceiro to whom I want to leave a great thank you for having received me and for believing in me and my abilities. I would like to thank you for your full support, availability, guidance and dedication. Definitely a fantastic person.

To Professor Jaime Cardoso for his guidance, for his support, for his total collaboration in solving doubts and problems and for his sympathy.

To my friends from Ingeniarius, André Araújo, Samuel Pereira and David Portugal, who were always ready to help, even when they were busy, and were also essential to the success of this thesis.

To all the friends who helped me in the data collection, Elisabete Santos, Diana Pinto, Manuel Silva, Rui Silva, Miguel Fange, Frederico Vaz and Diogo Ramalho.

To Fange Gym and Rui Silva's Gym for letting me use the space for several times to collect data.

To my friends Rita, Beatriz and André who have always been there to encourage me and never demotivate.

To Tiago, who there is not enough words to thank him for how essential he has been on this journey. For your unconditional support, for your patience, encouragement and for always making me see things on the positive side, a big thank you from the bottom of my heart.

Finally, a special thanks to my parents who have always supported me in all my decisions and always encouraged me to be the best version of myself. For being models of commitment and courage to them I dedicate this work.

“It’s only after you’ve stepped out your comfort zone that you begin to change, grow and transform”.

Roy T. Bennet

Contents

List of Figures

List of Tables

List of Abbreviations and Symbols

List of abbreviations

ANN	Artificial Neural Network
ARCANE	Augmented Perception Analysis Framework for Football
DAG	Directed Acyclic Graph
DBMM	Dynamic Bayesian Mixture Models
DT	Decision Tree
ECG	Electrocardiography
EEG	Electroencephalography
EMG	Electromyography
FC	Fractional Calculus
FCB	Futbol Club Barcelona
FFT	Fast Fourier Transform
FMS	Functional Movement Screen
FN	False Negative
FP	False Positive
FVQ	Fuzzy Vector Quantification
GSR	Galvanic Skin Response
GPS	Global Positioning System
HMM	Hidden Markov Model
HR	Heart Rate
IEMG	Integrated EMG
IMU	Inertial Measurement Unit
KNN	K-nearest neighbour
LDA	Linear Discriminant Analysis
LMF	Localized Muscle Fatigue
LR	Logistic Regression
MAV	Mean Absolute Value

MED	Mean Euclidean Distance
MEDF	Median Frequency
MF	Mean Frequency
MLP	Multilayer Perceptron
MUAP	Motor Unit Action Potential
NAR	Non-linear Autoregressive Network
NARX	Non-linear Autoregressive Network with External Input
NB	Naïve Bayes
PPG	Photoplethysmogram
PSO	Particle Swarm Optimization
RMS	Root Mean Square
RNNs	Recurrent Neural Networks
RR	Respiratory Rate
sEMG	surface Electromyography
SKT	Skin Conductance
SLEX	Smooth Localized Complex Exponential
SOM	Self-Organizing Map
SVM	Support Vector Machine
TN	True Negative
TP	True Positive
TR	Tactical Region
UWB	Ultra-wideband

List of symbols

α	Fractional Coefficient
β	Stability Coefficient

Chapter 1

Introduction

Association football, or football, is a team sports endowed with seemingly unpredictable and unrepeatable actions, wherein athletes aim at a superior collective performance in order to maximize the number of scored goals and prevent the opposing team from doing the same. Despite being much studied throughout the years, football is still considered a great subject of study, since there are still many parts that can be further explored. For example, one subject that has for long intrigued researchers is prediction in football. In fact, due to its difficulty and since the outcome of a football match is dependent on many factors [1], [2], from the point of view of researchers, predicting the results of sports matches is seen as a fascinating problem. However, prediction in football is usually quite complex, since this can be affected by the presence of several sources of variation for which a predictive model must account. For instance, team abilities may vary from year to year due to changes in personnel and overall strategy. In addition, team abilities may vary within a season due to injuries, team psychology, and players' transfers. Moreover, team performance may also vary depending on the site of a game [3].

Therefore, given the broad subjectivity of football in general, three questions can be raised: i) *How can we predict the outcomes of football matches?* ii) *What data should we use and how should we use it?* and iii) *How accurate can a model to predict football game outcomes be?*

Indeed, there are nearly infinite variables that one could put into a model to predict football games. For instance, the proficiency of a player's actions during a soccer match is determined by the response to cognitive, physiological, technical and tactical stressors under a high situational unpredictability. So, the prediction of performance during a match play, based on general and soccer-specific capacities, is limited by its own complex nature during competition [4]. Hence, football performance is, in fact, a construct based on many different performance components and their interaction at both individual and collective level, requiring the individual team members' harmonization into an effective unit to achieve the desired outcome [5].

Many authors have been presenting methodologies to predict the football match outcome, from which a considerably large number of studies has been published over the past decade, though none explores all relevant features that influence players' performance and that might determine success, or lack of it, during a match [6]-[8]. Bearing these ideas in mind, one can state that, as there is a wide range of variables that influence the football game outcome, all of them should be taken in account and studied as a whole and not in parts (e.g., result of the match over time, individual performance of players, injuries, playing home or away, among others). Also, the market has been offering disruptive technologies that can extract a wide range of variables. These approaches, combined with the adequate technology, are important mainly in a coaching perspective [7].

Thus, as one may realize, predicting the final outcome of a match, decoding the teams' performance, and additionally preventing any potential athletes' injuries, is seen by the scientific community as a kind of Holy Grail [9]. Under this premise, coaches, observers and researchers are looking for objective answers to understand how professional teams can behave under such complex and dynamic triad of factors, including: i) win the match; ii) present the best performance in the field; and iii) succeed in different and varied contexts (e.g., high number of successful passes, superior ball possession, etc.).

As such, based on these insights, we propose the development of an architecture for football prediction, capable of not only estimating athlete's position and action but also giving relevant information about its health status. For this aim, this architecture integrates information from several sources, namely positional and physiological data of each player.

This thesis is integrated in a national project, denominated ARCANE – Augmented peRCeption ANalysis framEwork for Football, which has as the main objective the development of an architecture for assisting coaches, in terms of player's and team's performance interpretation, during football matches. Thus, it is our expectation that the proposed approach will play an important role in the development of ARCANE, by contributing with pose, action and health status estimation.

This master thesis was performed within an academic environment, in the Engineering Faculty from Porto University, and a corporative environment, in collaboration with Ingeniarius, Lda.

1.1 - Ingeniarius

Ingeniarius is a company founded in 2014, specialized in research and technological development in the many fields of engineering, including robotics and automation, as well as all components within human society, namely the quality of life, sports and health. Besides providing consulting and outsourcing, Ingeniarius also provides training services related to engineering. Furthermore, it organizes and promotes other events, such as, scientific/technical conferences and entertainment, educational and promotional events of technological nature [10].

1.2 - ARCANE

ARCANE is a project that aims to promote an augmented perception of team performance in sports, which resulted from a joint effort of three different institutions, namely, The Laboratory of Expertise in Sport (SpertLab) from the Faculty of Human Kinetics at University of Lisbon, The Centre for Sports Engineering Research from Sheffield Hallam University and Ingeniarius. To fulfil the goals, the framework developed under the ARCANE project, will provide not only an interpretation of athletes' positional data but also information regarding their physiological signals, allowing, by consequence, to assess athletes on-the-fly and, to some extent, predict the health and performance outcomes. As such, the framework should be able to retrieve both positional and physiological data of athletes in real-time, which requires the development of novel wearable technologies. This acquired information is useful for coaches and for the technical support team to the extent that it iteratively provides a 'probabilistic tendency' of what comprises a game over time. Figure 1.1 provides a general overview of ARCANE.

Thus, we can describe the main objectives of ARCANE as follows:

1. Develop a novel wearable tracking system, capable of estimating the position of athletes in real-time.

2. Design a multi-sensor fusion algorithm in order to receive and to analyse data from multiple entities in real-time, including athlete's postural state (i.e., position and orientation) and athlete's state of mind.

3. Perform biosignal monitoring by integrating physiological sensors (e.g., heart rate monitors) within the wearable device.

4. Mathematical formulation of a framework for online match analysis and prediction based on players' position and physiological data over time.

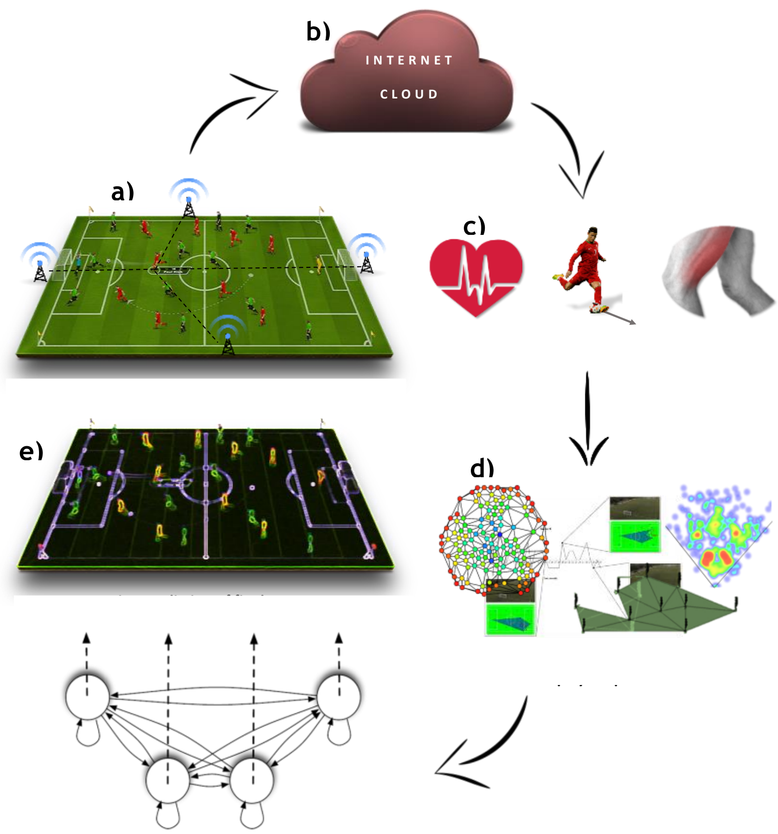


Figure 1.1 - General overview of ARCANE: a) Real-time contextual data acquisition; b) Data sent to internet server to benefit from cloud computing; c) Data cleaning and filtering techniques administered to pre-process and compute biosignals and athlete's pose; d) Pre-processed data feed multiple state of the art performance methods; e) Methods iteratively feed a macroscopic probabilistic model [11].

1.3 - Objectives

The key objectives defined for this Master Thesis project are aligned with the ARCANE project and can be stated as follows:

1. State of the art revision regarding the factors that most influence football match as well as regarding the most used classifiers and features for biosignals.
2. Construction and validation of a wearable device capable of measuring the electromyography (EMG) data from athletes.
3. Data acquisition and preprocessing of relevant athlete's data, by integrating positional and physiological information of players.
4. Recognition of athlete's position within the football field in order to perform Pose Estimation.
5. Recognition of athlete's actions (e.g., running and ball kicking).
6. Comparison of the classification results between several approaches encountered in the literature.
7. Detection of muscle fatigue in athletes.

The ultimate goal is to contribute towards the development of a predictive architecture for football, namely ARCANE project, not only to estimate the performance of teams, but also to prevent injuries that might occur during the match.

1.4 – Main Contributions

Due to the increasing interest in the use of wearable devices in the context of athletes' monitoring, specially, during football games, the development of architectures, such ARCANE has a great potential. Thus, overall, the main contributions herein presented are the following:

1. Construction of a wearable device capable of measuring the EMG data from athletes and further validation within a group of people.
2. Development of an architecture for classification of athletes' actions according to their EMG signals.
3. Introduction to injury prediction by exploring the first steps for detecting muscle fatigue.

1.5 - Structure

In this subsection, the overall structure of the document is presented. This master thesis is divided into seven chapters: 1- Introduction, 2- Preliminaries, 3- State of the Art, 4 – Project Overview, 5- Kinematic Data, 6- Physiological Data and 7– Conclusion. Furthermore, a brief introduction of each chapter is, hereinafter, presented.

Chapter 2

In Chapter 2, a review regarding the most common classification methods is presented, in which these methods are classified in two different groups. Firstly, the Non-Sequential Methods, including the K-nearest Neighbour (KNN), Bayesian Networks, Support Vector Machines (SVMs) and Artificial Neural Networks (ANNs) are addressed, following the Sequential

Methods, in which Recurrent ANNs are included. Finally, a discussion concerning trajectories estimation is performed, in which Fractional Calculus (FC) is introduced, being explained in detail how it can be applied in the context of a football match.

Chapter 3

In this chapter, the state of the art concerning computational performance assessment and, more particularly, prediction in football is discussed. Thus, we start by discussing the main theme of this thesis, which is Prediction in Football. As such, in order to understand how we can predict the outcomes of a football match, firstly, a review regarding the factors that influence football prediction is presented, in which these are divided in match statistics and situational variables or on-the-fly variables. Moreover, a review concerning machine learning research for football prediction is, then, presented. Furthermore, we discuss the importance of some signals, including kinematic and physiological signals, for football prediction, defining the most common features and classifiers for each type of signal.

Finally, we present several technologies used for football prediction that are currently available on the market.

Chapter 4

In this chapter, the overall methodology for this project is addressed. Moreover, a brief description regarding the data used for kinematic data collection is presented. Additionally, the construction of the wearable device used for physiological data collection is explained in detail. Finally, the software used for analysing both kinematic and physiological data is also presented.

Chapter 5

In Chapter 5, the results regarding trajectories' estimation with kinematic data are presented and discussed. For that aim, firstly, a detailed explanation concerning the process of data collection and the choice of features for feature extraction is presented. Afterwards, the FC is used for coordinates' estimation as well as for the calculation predictability and stability coefficients, being also studied the influence of different frequencies in these features. Then, after having decided about the frequency that produces the most suitable features, the data corresponding to this frequency is used for the implementation of several recurrent neural networks (RNNs). Finally, a comprehensive study in which the performance of several RNNs and the performance of FC are compared.

Chapter 6

In here, all of the work performed with physiological data is presented, being this chapter divided in three main parts. Firstly, a detailed explanation about the process of data collection, the pre-processing methods and the choice of features is made, being this phase common to the other two parts. As such, in the second part of this chapter, the results regarding actions' classification are presented. In here, an extensive study is made, in which a comparison between the performance of individual classifiers under different constraints. Afterwards, the Dynamic Bayesian Mixture Models (DBMM) approach is implemented, being its performance discussed and compared with the previous results. Finally, in the last part of this chapter, the EMG data collected with the wearable device *TraXports V2* is used for detecting muscle fatigue in athletes.

Chapter 7

Finally, in Chapter 7, the conclusions of this thesis are presented and a brief discussion regarding future works is made.

Chapter 2

Preliminaries

This chapter presents several theoretical concepts in order to better understand the work herein demonstrated. More specifically, we will present two types of methods for classification, the Non-Sequential Methods, which include the KNN, Bayesian Networks, SVMs and ANNs. Additionally, a method of Ensemble Classification, namely the DBMM, will be explained in detail. Finally, the FC approach for the context of a football match is introduced.

2.1 - Classification Methods for Pattern Recognition

Humans are naturally capable of recognizing patterns without worrying about the conditions of the environment, such as illumination variations, facial rotation, facial expressions and facial biometrical changes [12], [13]. However, when it comes to artificial recognition, the plot thickens as commanding a machine to do this as well as humans is a very complex task. Hence, a new artificial intelligence field emerged enabling the possibility of making artificial recognition something conceivable. This field, known as Pattern Recognition, studies the relation between machines and the environment by analysing how they are able to learn to distinguish various patterns of interest from its background and make reasonable decisions about its categorization. Pattern recognition can be applied to various areas, such as speech and face recognition, classification of handwritten characters and medical diagnosis [14].

In general, pattern recognition requires four stages (Figure 2.1): *i*) data acquisition; *ii*) pre-processing (filtering and normalization); *iii*) features extraction; and *iv*) classification - chosen according to the type of label output [15] - and it can be achieved by computing machine learning algorithms. These can be divided in two different types - supervised or unsupervised learning. Supervised classification designates each input pattern as a member of a predefined class that is previously learnt from supervised training data. In here, training examples are, first, used to train a classifier by determining the descriptor for each class, i.e., the set of common features for the example provided. Put it differently, data from the surrounding environment is acquired by sensors, digitizing machines or scanners and then preprocessed by either removing noise from the data or by extracting patterns of interest from the background. Afterwards, the most relevant features from the processed data are extracted, forming a collective entity that is then classified in the last step, according to pre-trained classifiers. When this step is completed, the rule of classification is formulated being this used after to predict in each class where an object that was not trained yet should be put. So essentially, supervised learning is a method that learns with the past to predict the future [16]-[18].

Figure 2.1 - Pattern Recognition Process.

On the other hand, with unsupervised learning it is only known the input data and not the corresponding output variables. As such, in this self-guided learning algorithm, given a set of data, the goal is to model the underlying structure or distribution in the data in order to learn more about it. Although there are several unsupervised learning tasks, the most commonly addressed is Clustering, in which the task is to establish the existence of clusters in the data. As classes are not previously defined, the system needs to observe the examples and recognize groups by its own. This results in a set of classes' descriptors (one for each class). Basically, given several input patterns, the system determines the similarities between them and clusters them.

There are several algorithms that can be used for pattern recognition. As such, in the following subsections a review is presented, being that we will first discuss various methods suitable for non-sequential data, where observations are assumed to be independent. Afterwards, methods based on sequential and ensemble classification will be also analysed. Finally, a new approach for estimating trajectories will be discussed.

2.1.1 - Non Sequential Methods

2.1.1.1 - K-nearest Neighbours

The KNN algorithm is a simple machine learning method for classifying objects based on a similarity measure (e.g., distance functions), being these assigned to the class which is most common amongst its KNNs [19]. In here, the classification starts with neighbours with the known classification. Although this may be considered as a training set, no explicit training step is required.

The choice of k should be done wisely since this parameter is highly dependent of the data itself. For example, for larger values of k, the effect of noise on classification is reduced, however, the boundaries between classes become less distinct. Moreover, for a 1-NN approach, where k=1, the object is simply assigned to the class of the closest neighbour. On the other hand, by performing a KNN approach, the k closest training points are found, in the majority of times, accordingly with Euclidean distance metrics. To find out the most suitable value of k, a method of cross-validation can be performed [20].

The KNN algorithm is vastly used in statistical estimation and pattern recognition as a non-parametric technique due to its simplicity [14].

2.1.1.2 – Bayesian Models

The Bayesian theory is a fundamental statistical approach for the problem of pattern recognition, which consists in a mathematical model that calculates the probability of an unknown given sample belonging to each of the possible classes, grouping it in the more likely class [21]. This represents the relation between a conditioned probability and its inverse.

As such, given two events A and B,

$$\mathrm{P}(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \tag{2.1}$$

Being P(A) and P(B) the *a priori* probabilities of A and B, respectively, P(A|B) and P(B|A) represent the conditioned probability (or *a posteriori*) of A and B, respectively [22].

Bayesian Networks are probabilistic models based on the Bayesian Theory [23]. These correspond to a graphical model structure known as Directed Acyclic Graph (DAG) which are structures where each graph's node represents a random variable containing the probabilistic information of each event, and the edges between nodes represent probabilistic dependencies among the corresponding random variables, being these estimated by using known statistical and computational methods. Therefore, a DAG can be represented by two different sets, more specifically, the set of nodes (vertices) and the set of directed edges, being that the first represents random variables and is drawn as a circle and labelled by variables' name, whereas the other set represents direct dependence among the variables and is drawn by arrows between nodes [24]. Due to its mathematical rigour and effective representation they are becoming more popular not only in the fields of statistics but also in machine learning and artificial intelligence.

In modelling dependence a Bayesian Network must have inherent assumptions about dependence and independence between variables. In the real world, two variables are virtually never truly and completely independent. However, in order to reduce the algorithm's complexity and cost it is necessary to simplify the Bayesian Networks and assume complete independence. In line with this, Naïve Bayes (NB) Model was developed. The basic principle of this classification can be termed as an "independent feature model", as it considers that the effect of a feature's value over a certain class is independent from the values of the other features. Such an assumption is called class conditional independence. Bayesian classifiers only need a small portion of already classified data, called training data, to estimate all the necessary parameters to proceed to classification. Then, based on this data, the algorithm receives as input a new non-classified sample, returning as output the most likely class for that sample accordingly to the probabilistic calculations [25], [26].

2.1.1.3 – Support Vector Machine

The standard SVM algorithm is a non-probabilistic binary linear classifier that performs pattern recognition through a supervised learning process [27]. In the traditional SVM approach, data classes are separated by building an N-dimensional hyperplane, separating them optimally into two categories (Figure 2.2). Given the set of training examples, where each one is marked as belonging to one of the two categories, the SVM training algorithm builds a model that assigns new examples into one category or the other. Basically, SVM classifies data by finding be most suitable hyperplane to separate one class from the other, being the best hyperplane the one with the largest margin between the two classes [14], [28]. In terms of use, this is one of the most popular machine learning methods since it is easy to use and usually offers good performance results.

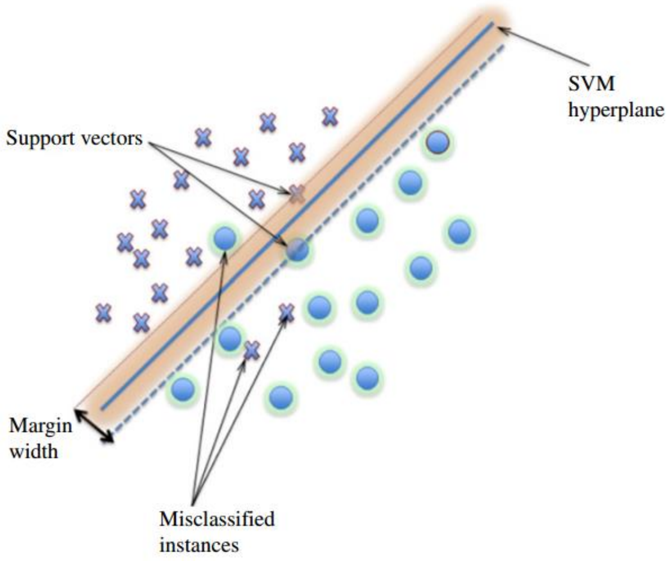


Figure 2.2 - Linear SVM example [29].

Support vectors are the data points that are closer to the separation hyperplane. As the traditional SVM is a linear classifier, when data is non-linearly separable, a mapping from the entrance space to a l-dimensional space can be made, thus allowing classes to be separated by a hyperplane. To make this mapping possible, a kernel is used, being that its choice is crucial to incorporate the a priori knowledge about the application [30]. This can be defined as a function, in which the similarity between observations is computed. By that, kernels offer an alternative to the traditional machine learning algorithms, since by using them there is no need to establish the vector of features. To do this, the data from the input space X is mapped into a high-dimensional feature space $\mathcal{X}$ as follows [31].

$$\Phi : X \rightarrow \mathcal{X} \tag{2.2}$$

As such, the actual function Φ does not need to be known, as the kernel function *k* calculates the inner product in the feature space.

$$\mathrm{k(x,y)} = \Phi(\mathrm{x}) \cdot \Phi(\mathrm{y}) \tag{2.3}$$

According to Schõlkopf [32], it is possible to demonstrate that the kernel function *k(x,y)* can be interpreted as a measure of similarity, since the kernel function defines a distance *d* on the input space by

$$d^2(\mathrm{x,y}) = (\Phi(\mathrm{x}) - \Phi(\mathrm{y}))^2 = k(x,x) - 2k(x,y) + k(y,y) \tag{2.4}$$

There are several kernel functions that can be used for machine learning problems, being some of them presented below.

1. <u>Linear kernel:</u> This is the most simple kernel function and is represented by $k(x,y) = x \cdot y$.

2. Radial Basis Function Kernel: This is in the form of a radial basis function, more specifically, a Gaussian function and is defined by $k_\gamma(x,y) = exp[-\gamma\|x-y\|^2]$, where γ is the parameter that sets the "spread" of the kernel.

3. Fourier Kernel: This is a kernel commonly used for the analysis of time series data, since it uses the Fourier Transform. It can be represented by $k_F(x,y) = \frac{1-q^2}{2(1-2q cos(x-y))+q^2}$, with $0<q<1$.

2.1.1.4 – Artificial Neural Networks

ANNs are computational models inspired in the central nervous system of humans – more precisely, in the natural neurons of the human brain - in which simple subunits (neurons) organize to form an enormous parallel structure [33]. These artificial neurons are basically a logical and mathematical structure that consists of inputs (identical to synapses), which are further multiplied by a parameter known as weight (strength of each signal), and then computed by a mathematical function which determines the activation of the neuron [14]. Then, the computation of the artificial neuron's output is performed, enabling the formation of artificial networks by combining these artificial neurons to process information. As such, the ANN can then be considered a network of weighted directed graphs in which the artificial neurons are the nodes and the connections between the several nodes are directed edges with weights [34]. The interconnection between neurons of an ANN is executed in different layers of each system. The first layer is composed by input neurons that send data to the second neurons' layer through synapses, being this repeatedly performed until the data get to the last layer of neurons, *i.e.*, the output layer, as it is illustrated in Figure 2.3 [35].

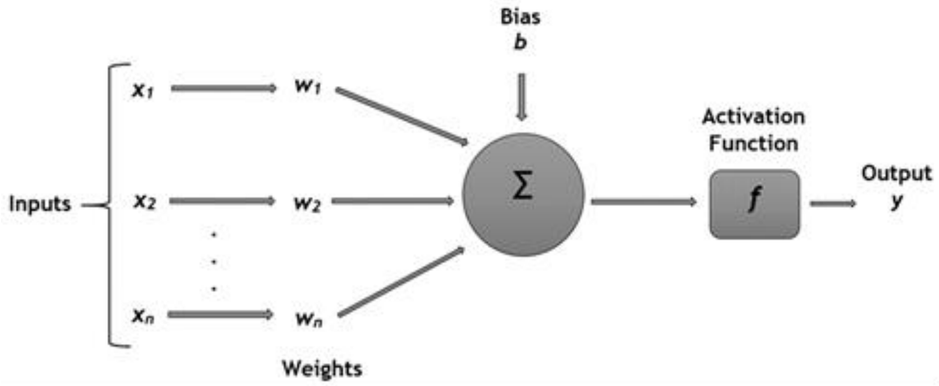


Figure 2.3 - Representation of the Activation Function in an ANN. Adapted from [36].

There are several neural networks-based algorithms being these defined depending on the following:

1. The type of connection model between the different layers of neurons and the allowed direction of information flow;
2. Learning process to update the weights of connections;
3. The activation function that converts the input in an output [37].

Regarding the first point, ANNs can be classified in two main categories: feedforward and RNNs. In the first case, the information flows strictly forward from inputs towards the outputs

(Figure 2.4) and does not keep record of its previous output values, whereas in the second one, the information can flow on both directions [38]. The latter is presented in Section 2.1.2.2.

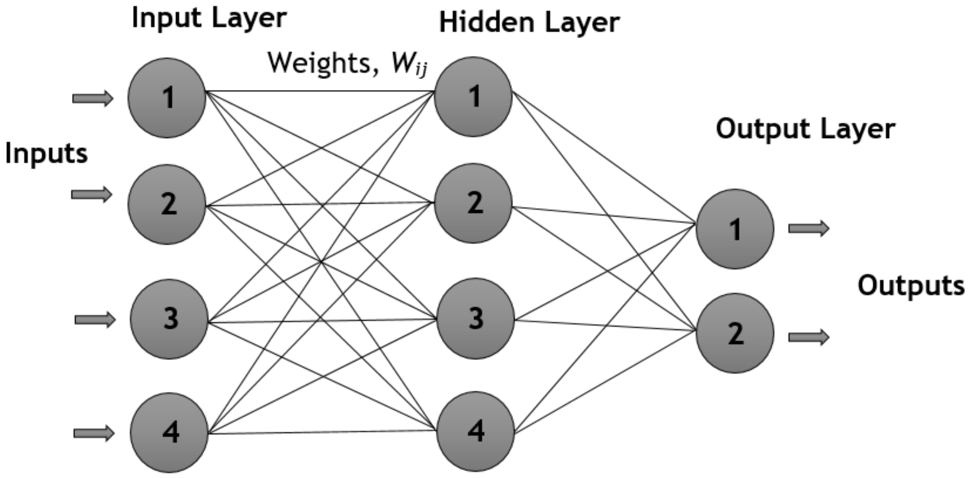


Figure 2.4 - Architecture of a feedforward ANN. Adapted from [36].

The learning process of an ANN can be either supervised or unsupervised [39]. In supervised learning, there is an external agent that iteratively assesses how much the network is close to an acceptable solution, adapting the weights between neurons during training so that a desirable classification is obtained. As the weights are adjusted through the learning process, these are responsible for pattern memorization. On the other hand, in unsupervised learning, as there is no *a priori* knowledge of the networks' output, the algorithm tries to form classes by detecting similar patterns and by clustering them. In here, the network is adjusted accordingly to the statistics of the input data, being this a method that learns through observation.

Typically, up to a certain extent, the performance of neural networks increases with the number of hidden layers and the number of neurons in each layer. Yet, it is important to maintain a balanced trade-off between size of network and the complexity it brings; the algorithm should have a number of neurons large enough to represent the problem domain, but small enough to allow generalizing the training data [17].

Due to its ability of learning complex nonlinear input-output relationships and changing its weight iteratively accordingly to the data, neural networks are able to provide efficient results in the field of classification [40], [41]. Moreover, their low dependence on domain-specific knowledge and availability of efficient learning algorithms led this type of model to gain an increasing popularity in the field of pattern recognition [15].

2.1.2 - Sequential Methods

Despite of the already proved utility of non-sequential methods, such as SVMs, logistic regression (LR) and feedforward networks, these are not capable of modelling time, which, accordingly to the literature, is a key element in countless learning tasks, including image captioning, speech synthesis (that require models capable of producing outputs in the form of sequences), time series prediction, video analysis and musical information retrieval (that require models that can learn from inputs in the form of sequences), and others, such as translating natural language and robot's control (that often demand both capabilities) [42]. As a result, new models that are capable of dealing with sequential data were developed, being two of them presented below.

2.1.2.1 - Recurrent ANNs

Although Markov Models and their hidden counterpart have already proved to be efficient in several areas, these are still quite limited since their states must be drawn from a discrete state space S with small dimensions, which leads to high computational costs, especially for HMM, when the set of possible hidden states is large. Furthermore, each hidden state can be only dependent on the previous state. As a result, a new ANN approach emerged [42].

As it was already presented before, recurrent ANNs, or RNNs, are a class of ANNs formed by a feedback connection structure. Besides the typical flow of information from the inputs to the outputs, it has additional connections that allow to either connect directly to the same layer or even to lower layers, enabling the flow of information from the outputs towards the inputs (Figure 2.5) [38].

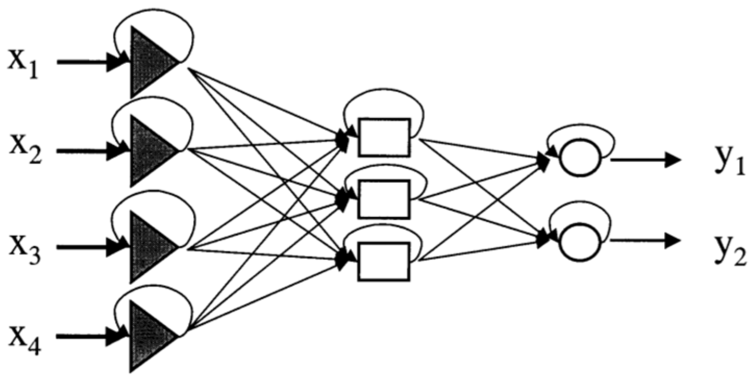


Figure 2.5 - Architecture of a feedback ANN (RNN) [38].

These non-forward connections are termed as recurrent connections, having a time-delay of usually 1 time step (discrete time), which allows the model to be aware of its previous inputs [43]. Moreover, with this feedback architecture, besides the normal weight, each neuron possesses one additional weight as an input, allowing an extra degree of freedom when trying to minimize the training error. By having such structure it allows to keep the memory of the previous state so that the next state will depend not only on the input signals, but also on the previous states of the network, allowing to perform temporal processing and learn sequences [38].

Hence, unlike feedforward neural networks, RNNs are capable of using their internal memory in order to process arbitrary sequences of inputs. Furthermore, as RNNs are dynamic, their state is always changing until they reach an equilibrium point, being that they are unaffected until the input changes and a new equilibrium needs to be achieved [42]. Moreover, unlike the HMMs, RNNs are capable of capturing long-range time dependencies, i.e., although in this model each state also depends only on the current input and the state of the network at the previous time step, RNNs' hidden state can contain, at any time step, information from a nearly arbitrarily long context window [42].

Therefore, RNNs are considered effective models for sequential data, having already demonstrated to be computationally more powerful and biologically more plausible when compared to other adaptive approaches, such as feedforward networks, SVMs and HMMs [44]-

[46]. Moreover, as these allow to model relationships between a set of variables, they are considered to be very appropriate for any functional mapping problem. RNNs can be used for a range of applications, including adaptive robotics and control, handwriting recognition, keyword spotting, protein analysis, stock market prediction and many other sequence problems [47]–[49].

2.1.3 - Ensemble Classification

Due to the increasing interest in the machine learning field a new concept started to emerge, the Ensemble Classification. This can be defined as the use of learning algorithms that construct a set of classifiers, which can be based on different approaches, and then classify new data points by taking a weighted vote of their predictions, resulting in more accurate and precise results. The first ensemble method developed was the Bayesian averaging, which led to the development of further algorithms [50].

2.1.3.1 – Dynamic Bayesian Mixture Models

One example is the DBMM introduced in [51] and in [52], which is designed to combine multiple classifiers in a dynamic way, by combining the outputs of the different conditional probabilities of the different classifiers. To do this, a weight is assigned to each of the classifiers, according to previous knowledge, by using an uncertainty measure as a confidence level, which can be updated locally during online classification. When the local weight is updated, it assigns priority to the classifier that presents more confidence along the temporal classification, since this can vary along the different frame classifications. Figure 2.6 depicts an overview of the DBMM architecture applied to the human activities recognition by using body motion from RGB-D images, in which the base classifiers are integrated as weighted posterior distributions.

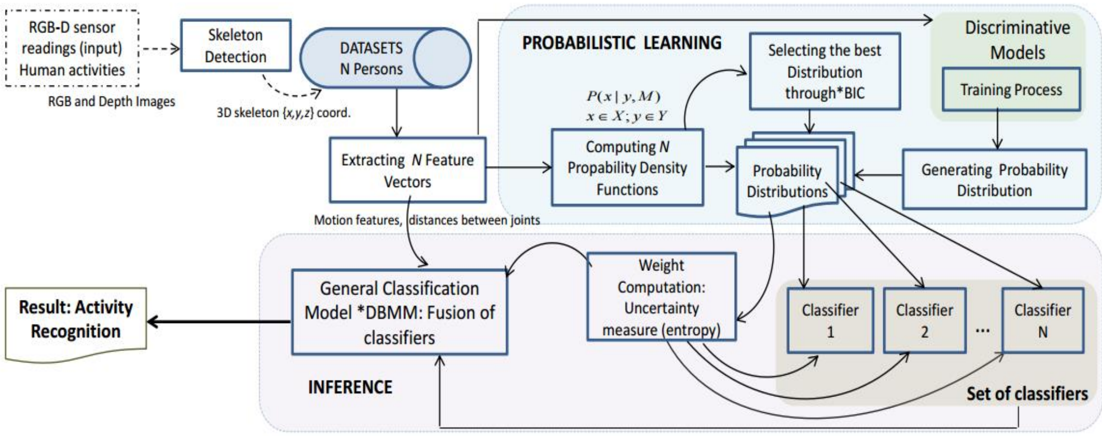


Figure 2.6 - DBMM architecture for activities' recognition [51].

The DBMM assumes a Markov property of first order that includes the temporal information as a simple dynamic probabilistic cycle and, similarly to the Hidden Markov Model, this architecture accepts the matrix of state transition probability.

Moreover, the set of dynamic probabilistic classifiers considers the Bayesian probability as a means of classification, in which each model contributes to the inference result in the proportion of its probability. By other word, this fusion model is directly presented as a weighted sum of the distributions, where it is possible to obtain the combination of different models in one.

Thus, the DBMM combines a set of models A= $\{A^1_m, A^2_m, \ldots, A^T_m\}$ where A^t_m is a model with *m* attributes, such as, variables or features, in the time t={1,2,…,T}. The global model of DBMM can be expressed as follows:

$$P(C^t|A) = \beta \; x \; M_{trans} \; x \sum_{i=1}^{N} w_i^t x \; P_i(A|C^t) \tag{2.5}$$

Where:

- M_{trans} is the model for the state transition probability between the variables class, or states, over time. This model can represent *a priori* in DBMM, as a dynamic probabilistic cycle, where the present *a posteriori* becomes the new *a priori,* or may represent the state transition matrix. Thus, M_{trans} is calculated as a dynamic *a priori* information of the model.
- $P_i(A|C^t)$ is the *a posteriori* result for the base classifier *i* at the instant of time u, which becomes the probability *i* of the mixing model, with *i* = {1, ..., N} and N being the total number of classifiers considered in the model.
- The weight w_i^t is estimated by using a measure of confidence that is based on entropy.
- The normalization factor is defined by

$$\beta = \frac{1}{\sum_j (P(C_j^t | C_j^{t-1}) x \sum_{i=1}^{N} w_i^t x P_i(A|C_j^t))} \tag{2.6}$$

The global function of probability distribution can be expressed by the following equation:

$$P(C^t|A) = \beta \; x \; P(C^t|C^{t-1}) \; x \sum_{i=1}^{N} w_i^t \; x \; P_i(A|C^t) \tag{2.7}$$

Where $P(C^t|C^{t-1})$ represents the distribution of the transition probability between the variables of each class over time.

Note that class C in time t is conditioned to the class in time t-1. This step describes the non-stationary behaviour that is recursively applied, in which the posterior of the anterior time of each class becomes the present prior in order to reinforce the classification in the timestep t, by using the information regarding the timestep t-1.

2.1.3.1.1 – Weights' Assignment for the fusion model using entropy

In DBMM, $H_i(L)$ represents the level of confidence used to assign weights w_i^t to the base classifiers and, by consequence, to update the probabilistic model. Thus, the weights are calculated considering the entropy for each base classifier $H_i(L)$, through the analysis of the previous results. The entropy of the posterior probabilities can be expressed by the following equation:

$$H_i(\mathcal{L}) = -\sum_{j}^{s} \mathcal{L}_i \; x \log(\mathcal{L}_i) \tag{2.8}$$

Where $\mathcal{L}$ represents the set of conditional probabilities $P_i(C|A)$ which are given by their base classifier *i* and *j* is the index for the set of posteriors of a specific base classifier. Thus, knowing H_i, the weight w_i^t for each base classifier can be obtained by, first, calculating the global value of the weight:

$$\forall w_i, w_i^t = \left[1 - \left(\frac{H_i}{\sum_{i=1}^{N} H_i}\right)\right] \quad (2.9)$$

Where $H_i = H_i(L)$ is the present value of entropy. Then, the weight w_i^t is normalized as follows:

$$w_i^t = \frac{w_i^t}{\sum_i^N w_i^t} \quad (2.10)$$

This step assures that $\Sigma_i w_i = 1$.
During the process of classification, the base classifiers can alter their performance frame by frame. By that way, a local update of the weights during the classification allows to produce a better confidence, assigning priority to the base classifiers with more confidence over the previous classification. Assuming a system memory as a Markov property during the online classification, it is possible to obtain temporal information from the set of posteriors for each base classifier. This information is then used along with the weights in the instance w_i^{t-1} in order to update the weights of each base classifier during the classification of each frame, as follows:

$$w_i^t = \frac{w_i^{t-1}\ x\ P(w_{i_{new}}|H_i(\mathcal{L}))}{\sum_{i=1}^{n} w_i^{t-1}\ x\ P(w_{i_{new}}|H_i(\mathcal{L}))} \quad (2.11)$$

Where w_i^t is the estimated weight that is updated by each base classifier in each instance and w_i^{t-1} is given by the previous calculated error in t-1.

2.2 - Fractional Calculus

FC is a field of mathematics that goes beyond the traditional definitions of integral calculus and derivative operators, by considering the possibility of generalizing the operation of differentiation to non-integer orders, such as real number powers, real number fractional powers and complex number powers [53]. As such, by operating as a natural extension of the integer derivatives, fractional derivatives are capable of acting as an excellent tool for the description of memory and hereditary properties of processes [54]. Furthermore, another property that is usually highlighted is that while an integer-order derivative just implies a finite series, the fractional-order derivative requires an infinite number of terms [55].

Although there are several approaches described in the literature [53], for the purpose of this thesis we will only focus on the *Grünwald-Letnikov* formulation, which approaches the problem from the derivative side. As such, taking into consideration the fundamental definition of a derivative.

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \quad (2.12)$$

By applying this formula again, we can easily obtain the second derivative.

$$f''(x) = \lim_{h \to 0} \frac{f'(x+h) - f'(x)}{h} = \tag{2.13}$$

$$\lim_{h_1 \to 0} \frac{\lim_{h_2 \to 0} \frac{f(x+h_1+h_2) - f(x+h_1)}{h_2} - \lim_{h_2 \to 0} \frac{f(x+h_2) - f(x)}{h_2}}{h_1} \tag{2.14}$$

When $h=h_1=h_2$, the expression simplifies to

$$f''(x) = \lim_{h \to 0} \frac{f(x+2h) - 2f(x+h) + f(x)}{h^2} \tag{2.15}$$

Thus, for the n^{th} derivative, we can summarize this formulation, in which the operator D^n is used to represent the n-repetitions of the derivative.

$$D^n[x(t)] = \lim_{h \to 0} \left[\frac{1}{h^n} \sum_{k=0}^{n} (-1)^k \binom{n}{k} x(t-kh) \right] \tag{2.16}$$

The gamma function is intrinsically tied to FC by definition, being considered a simple generalization of the factorial for all real numbers. Let the gamma function be Γ, this can be defined as follows [53]:

$$\Gamma(k) = \int_0^\infty e^{-u}\, u^{k-1} \mathrm{du} \text{ for all z} \in \mathbb{R} \tag{2.17}$$

For integer numbers it can be expressed as follows:

$$\Gamma(k+1) = \mathrm{k}\Gamma(\mathrm{k}), \qquad \text{when k } \in \mathbb{N}_+, \qquad \Gamma(k) = (k-1)! \tag{2.18}$$

Equation 2.16 can be generalized for non-integer values for n with $\alpha \in R$ if the binomial coefficient is understood as using the Gamma Function in place of the standard factorial. Thus, the expression $D^\alpha[x(t)]$ given by

$$D^\alpha[x(t)] = \lim_{h \to 0} \left[\frac{1}{h^\alpha} \sum_{k=0}^{+\infty} \frac{(-1)^k \Gamma(\alpha+1)}{\Gamma(k+1)\Gamma(\alpha-k+1)} x(t-kh) \right] \tag{2.19}$$

is said to be the Grünwald-Letnikov fractional derivative of order $\boldsymbol{\alpha}$, $\alpha \in \mathbb{C}$, of the signal $x(t)$. This formulation can be readjusted for discrete time calculation as it is presented below:

$$D^\alpha\big[x[t]\big] = \frac{1}{T^\alpha} \sum_{k=0}^{r} \frac{(-1)^k \Gamma[\alpha+1]}{\Gamma[k+1]\Gamma[\alpha-k+1]} x[t-kT] \tag{2.20}$$

Where T is the sampling period and r is the truncation order, is the approximate discrete time Grünwald-Letnikov fractional difference of order α, $\alpha \in \mathbb{C}$, of the discrete signal $x[t]$.

It is possible to extend an integer discrete difference, *i.e.*, classical discrete difference, to a fractional-order one, using equation 2.21 [56].

$$\Delta^{\varpi} x[t] = \begin{cases} x[t] & , \varpi = 0 \\ x[t] - x[t-1] & , \varpi = 1 \\ \Delta^{\varpi-1} x[t] - \Delta^{\varpi-1} x[t-1], & \varpi > 1 \end{cases} , \tag{2.21}$$

Where $\varpi \in \mathbb{N}_0$ is the order of the integer discrete difference.

Hence, one can extend the integer-order $\Delta^{\varpi}x[t]$ assuming that the fractional discrete difference satisfies the following inequalities:

$$\varpi - 1 < \alpha < \varpi \tag{2.22}$$

Thus, FC is a mathematical tool suitable for describing many phenomena, such as irreversibility and chaos, because of its inherent memory property. With this in mind, one can easily understand the potential of FC in the dynamic phenomena of player's trajectory. As such, in the following subsection, FC approach for the study of football players' trajectories will be discussed.

2.2.1 - FC approach for football player's trajectories estimation

During a football match, to track football players, whether we use a manual or an automatic system, a matrix containing the planar position of each player n of team δ over time is generated, called positioning matrix $X_\delta[t]$.

$$X_\delta[t] = \begin{bmatrix} x_1[t] \\ \vdots \\ x_{N_\delta}[t] \end{bmatrix}, \quad x_n[t] \in \mathbb{R}^2, \tag{2.23}$$

Wherein N_δ represents the current number of players in team δ at sample/time t. $X_\delta[t]$, wherein row n represents the planar position of player n of team δ at time t. It is also noteworthy that each element from $x_n[t]$ is independent from each other as they correspond to the (x, y) coordinates of the n^{th} player planar position.

Taking into consideration that a football team has 11 players, we have $N_\delta = 11$. Finally, using the equations presented previously and considering players' dynamics one can define an approximation of player n to next position, *i.e.*, $x_n^s[t+1]$, as:

$$x_n^s[t+1] = x_n{}^0 + x_n[t] - x_n[t-1] - \frac{1}{T^\alpha} \sum_{k=0}^{r} \frac{(-1)^k \Gamma[\alpha+1]}{\Gamma[k+1]\Gamma[\alpha-k+1]} x[t+1-kT] \tag{2.24}$$

Wherein $x_n[t] = 0, \forall\, t < 0$ in such a way that $x_n[0] = x_n{}^0$ corresponds to the initial tactical position of player n in the field, $x_n{}^0 \in \mathbb{R}^2$.

Since players may not be able to drastically change their velocity between two consecutive samples, it is important to choose an adequate sampling period, more specifically, a small one (e.g., $T \leq 1$ second). However, by doing this the memory requirements increase, since the last r positions of each player, i.e., $O[rN_\delta]$ are always being memorized. Nonetheless, the truncation order r does not need to be too large and will always be inferior to the current iteration/time t, i.e., $r \leq t$. Moreover, by analysing Equation 2.24 one can conclude that the influence of past events (i.e., previous positions) of a given players depends on the fractional coefficient α. Thus, analysing carefully the fractional coefficient may be very helpful to understand the level of predictability of each player.

Furthermore, during a football match, each athlete plays a different role (e.g., defender, goalkeeper, midfielder), having, by consequence, a more restricted intervention region, which provides some organization to the team's collective dynamics. As such, despite the different movements that a player might do to support the defensive and offensive phases, he will always return to his main tactical region (TR) due to his positional role.

Note that regardless on its size, the geometric centre of the TR of player n, herein denoted as tactical position ${x_n}^0$, can be defined as a specific planar position a player converges during the game, since it is directly dependent on the player's tactical mission [55].

As it is possible to observe in equation 2.24, the player's trajectory is dependent on the fractional coefficient α, meaning that its estimation can only be achieved by adjusting the coefficient along time. Thus, the best fitting α for player n at time t, i.e., α_n [t] based on its last known positions so far, can be formulated by the following minimization problem:

$$\min_{\alpha_n[t]} d_n^{min}(\alpha_n[t+1]) = \Big| -x_n[t+1] + x_n[t] - x_n[t-1] - \\ - \frac{1}{T^\alpha} \sum_{k=0}^{r} \frac{(-1)^k \Gamma[\alpha_n[t+1]+1]}{\Gamma[k+1]\Gamma[\alpha_n[t+1]-k+1]} x[t+1-kT] \Big| \\ ,s.t \;\; \alpha_n[t+1] \in [0,1] \tag{2.25}$$

During a football match, in order to be able to estimate players' trajectories it is important to not only have information about each player's predictability but also about its stability within the field. As such, the classification of players as stable or unstable can be formulated by the following equation.

$$\beta_n^\tau[t] = \frac{v_n^\tau[t]}{v_n^\tau[t-1]} = \frac{x_n^\tau[t] - x_n^\tau[t-1]}{x_n^\tau[t-1] - x_n^\tau[t-2]}, \in [0,1]. \tag{2.26}$$

2.3 - Summary

In this chapter the theoretical concepts necessary to understand the work herein presented were discussed. There are several classification methods for pattern recognition, which can be divided in two different types: Non-Sequential Classification Methods that comprise simple machine learning algorithms, including KNN, Bayesian Models, DTs, SVM and ANNs, and that do not take in consideration the factor time and Sequential Classification Methods, such as Markov Models and RNNs, which were developed for modelling time and, thus, dealing with sequential data. It is noteworthy that the methods previously discussed can be further enhanced by using ensemble classification approaches, i.e., performing the fusion of multiple classification methods. Finally an introduction to a possible approach for trajectories estimation was performed, in which FC was presented.

Chapter 3

State of the art

This chapter presents a literature review regarding Prediction in Football, in which the following concepts are addressed: Factors that influence Football Prediction, Machine Learning for Football Prediction and Relevant Signals for Football Prediction. Concerning the latter, an extensive discussion regarding the most common features and classifiers used for biosignals is presented. Finally, some examples of current technologies for football monitoring are presented.

3.1 - Prediction in Football

The problem of modelling football data has become increasingly popular in the last few years. Several models have been proposed with the purpose of estimating the characteristics that bring a team to lose or win a match or to predict the final score. Undeniably, predicting the outcome of football matches is very motivating as a research problem, not only because of its intrinsic challenge, but also because the result of a football match relies on many factors, or contextual information [7]. Concerning this, in this section it is first presented a review regarding the main factors that should be considered when trying to forecast football results. Then, a literature review concerning the implementation of machine learning algorithms for football prediction is presented. Finally, a brief discussion regarding the choice of the most relevant signals for football prediction is made.

3.1.1 - Factors that influence football prediction

According to the literature one reasonable way to forecast future match results is to use the information already known from past results combined with relevant information such as the scoring frequency [57]. However, when developing models for football prediction, the history of matches should not be the only factor to be taken in account [58].

As a result, in order to understand which factors should be considered during the development of an architecture for football prediction, the following subsection will present a discussion regarding several factors that, according to the literature, influence teams' performance.

3.1.1.1 - Match Statistics and Situational Variables

According to Courneya and Carron [59], home advantage can be defined as “the consistent finding that home teams win over 50% of the games played under a balanced home and away schedule”. In fact, during the last years, several authors have demonstrated evidences that home advantage is real and that indeed influences the outcome of a football match [60], being this subject of analysis of statistical companies from all over the world. One example is presented in Figure 3.1. In here, match results comparisons were made, revealing that not only home teams are more likely to win matches but also that the concept of home advantage is present in most of all sports. These types of observations allowed analysts to understand the real role of home advantage in football prediction, being already established that playing at home stadium indeed provides a major advantage in football.

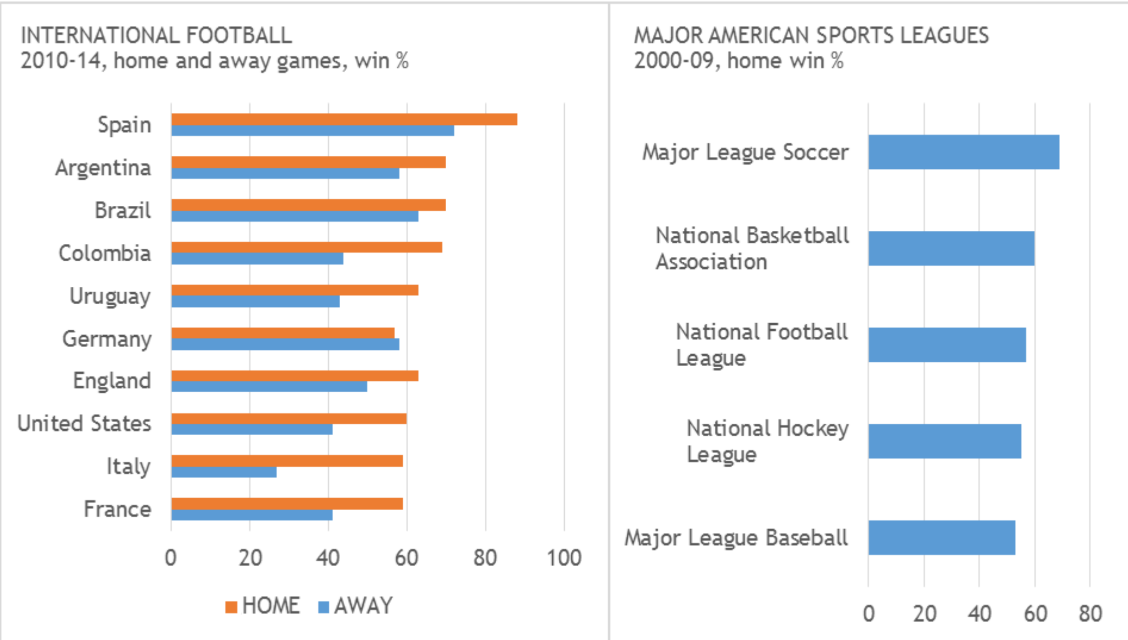


Figure 3.1 - Influence of Home and Away Games in match results. Adapted from [61].

Up to now, one may understand that the concept of home advantage is undeniable. However, the reasons for this event are still not clearly known. Although there are several explanations, the literature highlights the influence of the crowd as the main plausible one [62]. In this context, Schwartz and Barsky [63] investigated the influence of the crowd factor, hypothesizing that this exists because of the greater social support to home teams, not only in terms of adepts but also in terms of local networks that usually give more audience to the home team. Another reason regarding the crowd effect was presented in the study of Greer [64] in which he concluded that the home crowd may also have an influence in referees’ actions as the crowd has the power of intimidating them. Concerning this, Neville et al. [65] randomly allocated forty qualified referees to either a noise group or a silent group and ask them to review 47 challenges and incidents recorded during an English Premier League game between Liverpool (home) and Leicester (away) from the 1998/1999 season in order assess if they were legal (no foul) or illegal (foul), being that in the case of a foul they should also assess if it was a home foul or an away foul. Basically, the referees had to choose between four categorical response variables (1) home foul (2) away foul (3) no foul and (4) uncertain, in the case of uncertainty. The experimental results showed that the referees from the silent group demonstrated a higher certainty in their responses, awarding a greater number of fouls against home players. On the other hand, the group of referees that was subjected to crowd noise awarded 15.5% fewer fouls against home players. Thus, this study allowed to conclude that rather than penalising away players more, the dominant effect of crowd noise influenced referees to penalise home players less. Additionally, it was also found out that the referee's experience also plays an important role in the course of the game, as it can positively affect the number of fouls awarded. Despite the apparent insignificance of this factor, the reality is that this kind of 'refereeing edge' can be crucial in sports, as simple decisions as awarding a penalty can significantly alter the course of a match [62].

Another factor that was pointed out as a great influencer of teams' performance, and thus in results forecasting, is the team's participation in external cups, being one of the explanations for this based on the existence of financial incentives. In order to understand why external cups can influence teams' performance, Goddard and Asimakopoulos [66] evaluated the Football Association Challenge Cup's effect on the top four English football leagues, demonstrating that the possibility of having success in external cups provides a significant morale boost, having by consequence a positive effect on the overall league's performance. In this study, the authors also identified that other factors, such as championship, promotion and relegation issues, are determinants of the match results in football.

In addition to that, football experts argue that models for football prediction should not only take into consideration the team's previous matches and results, but also the past results between opponents, since it is possible that some teams may be more motivated to win against a particular opponent [58].

Finally, another factor that may influence teams' ability is the occurrence of managerial changes, new transfers or depreciation of current team fundamentals. In the work presented by Audas et al. [67], the real impact of these factors on team performance is shown, concluding that, on average, teams that changed their manager during an ongoing season performed worst over the following three months of games. This could be explained by the fact that new managers can sometimes adopt playing styles that are inconsistent with the current playing staff and also because it is common sense that when a team is underperforming, there is usually a change in the manager, which may also affect the state of mind of the players.

3.1.1.2 On-the-fly Match Analysis

According to Rusling [68], the epidemiological research in English football has shown that each professional footballer experience an average of 1.3 injuries per season. Of all the injuries sustained, 38% are the result of contact mechanisms and 58% of non-contact mechanisms. In line with this, a 7-year European study found that teams suffered an average of 2 injuries per player per season, which equates to an injury incidence of 8 injuries for each 1000 training and match hours. Furthermore, injury incidence for English youth academy footballers was initially indicated to be lower, with data indicating that teams suffer around 0.4 injuries per player per season, wherein each player would sustain about 2.23 injuries per 1000 hours of total exposure. Contact based injuries resulted from uncontrollable extrinsic factors, such as tackles, and were then considered to be unpredictable. Conversely, non-contact injuries, such as low back pain or injuries due to running, were considered to be theoretically predictable [69], [70].

Nevertheless, there are many factors that may cause injuries during a football match, being the most common ones poor training methods, structural abnormalities, weakness in muscles, tendons, ligaments and unsafe exercising environments [71]. Moreover, sports injuries are associated with a wide range of hardly quantified parameters/situations, namely the quality of sports and training equipment, the prior years of sports training and experience under certain competition, the inadequate warm-up (whether insufficient or excessive), the nutritional and hydroelectric imbalances, etc. [72], being that a significant percentage of injuries (9%-34%) happens due to overuse [68]–[70]. The majority of these injuries happen during competition, being most of them described as traumatic, with 29% being due to foul play [9]. Additionally, fatigue greatly influences the occurrence of injuries. Although localized muscle fatigue (LMF) can sometimes contribute to good results, for example, it can promote muscle growth (as observed in bodybuilders), most of the times fatigue can cause serious injuries, especially when the level of fatigue is high, being this due to the fact that fatigued muscles are not capable of absorbing so much energy since they are stretched [73].

A challenge that medical teams face is their inability to successfully predict the occurrence of injuries and the recovery time of an injured player. Undoubtedly, providing an accurate prediction would be helpful for the coach and technical teams [70]. Hence, a number of different methodological approaches have been used to describe the reasons for sports injuries. These include interviews of injured athletes, analysis of video recordings of actual injuries, clinical studies, in vivo studies and simulation of injury situations. On the other hand, some authors, such as Drawer and Fuller [74], went beyond these traditional methods by studying the effect that injuries have on the collective performance of the football team. By measuring the team quality as a sum of player quality scores from a sample of five seasons of the English Premier League, the authors showed that the injuries of key players were significant in explaining team performance. Concerning this problem, Rusling et al. [68] highlighted the use of tools, as the Functional Movement Screen (FMS), to predict individuals at heightened risk of injury within collective sports. However, the current literature does not allude to whether or not FMS has a role in predicting injuries on-the-fly, i.e., during a match. Moreover, the number of evidences assessing the FMS in determining injury risk within individuals specifically for football is rather limited. Thus, further investigations need to be conducted to obtain more solid and relevant results.

As it is shown, injuries play a crucial role in the overall collective performance of the team, and studying them may allow to go beyond the classical predictive analysis, as one may be able to not only predict the outcome of the football match, but also when a given player might be susceptible to get injured, for example, by implementing an automated system capable of predicting and detecting fatigue in football players and, consequently, acting as a warning device, thus promoting a better performance and avoiding unnecessary injuries [75]. However, to do this, a real-time measurement of changes in LMF is required [75].

Although the occurrence of injuries greatly influences the final result of a football match, this is not the only factor that contributes to the overall performance of the team. In fact, many other on-the-fly factors have been discussed in the literature. In this context, in order to evaluate the influence of physiological factors in player’s performance, Faude et al. [76] proceeded to an extensive analysis of 360 videos of goals occurred in the first German National League, detecting that in the majority of goals scored anaerobic actions were observed, both for the scoring player and the assisting one. Moreover, in the work presented by Rampinini et al. [77], a significant decline in Italian Series A player’s performance between the first and second half was observed, being these demonstrated to be a consequence of poor physical preparation. In fact, several game analysis have already revealed the existence of a reverse relationship between ball possession and distance covered, meaning that in the majority of situations, teams that possess superior technical skills perform the games at lower relative work rates when compared to their opponents. Such analysis led several authors to conclude that besides of increasing the probability of a player developing an injury, fatigue also affects scoring frequency [73], [78], [79], which ultimately influences football prediction.

Furthermore, usually athletes are under great pressure during a match, exhibiting various emotional reactions not only during [80] but also before a competition [81]. Although some of the emotions felt can have positive effects on the athletes, such as confidence and motivation, others, including tension, anger, fatigue and depression can negatively affect players' performance. In fact, several studies regarding the mood states of athletes – estimated by EEG [82] or event-related potentials [83] - have proven that there are differences in the characteristics of information processing, brain function and the brain nerve cell metabolic node of athletes between different sport events and between athletes with different technical skills, confirming the influence of emotional state on players' performance.

In Table 3.1, we summarize the main factors that influence the outcome of a football match and, by consequence, football prediction.

Table 3.1 - Summary regarding the factors that influence Football Prediction.

TYPE OF FACTOR	
MATCH STATISTICS AND SITUATIONAL VARIABLES	**ON-THE-FLY MATCH ANALYSIS**
Home or away game Previous Results Scoring Frequency Number of matches in a week Weather Teams' players Number of injured key players Bookings (yellow and red cards) Age of players Players' Managers	Players' performance Teams' performance Injuries Motion Physiological State

3.1.2 - Football Prediction: a scientific overview

For many authors, such as Hale [84], the result of a football match is considerably difficult to predict due to its inherent variability, which raises the following question: *How can we make predictions under such variability?*

Igiri & Okechukwu [2] highlight the many techniques that have been used to predict the result of the football match, such as ANN, NB, KNN, SVM, among others. Adducing further contributions to these techniques, in the early 2000, Koning [85], started to explore a Bayesian approach with Markov chains and the Monte-Carlo method, estimating the quality of football teams. Similarly, Rue et al. [9] suggested a Bayesian dynamic generalized linear model to estimate the time-dependent skills of football teams and to predict football matches. Crowder et al. [86] modeled, to some extent, the 92 teams of the English Football Association League using refinements of the independent Poisson model to predict the probabilities of home win, draw or away win. Andersson et al. [87] estimated the expected number of goals in a football match based on the scoring intensity to predict the probability that a team has to win a tournament using the estimated scoring intensities. Goddard et al. (2004) [88] proposed a regression model to forecast the English Football Association League results. Halicioglu [89] statistically analyzed football matches and recommended a method to predict the winner of the Euro 2000 football tournament. Rothstein et al. [90], combined fuzzy logic with genetic and neural optimization techniques to formalize football predictions.

In 2011, Buursma attempted to ascertain the most relevant features necessary to predict the football match end-result for betting purposes, exclusively considering previous results [91]. The author employed seven different machine learning algorithms to classify the matches into home win, draw or away win. The results demonstrated that any of the classifiers was able to present an accuracy above 55% and, because of that, the author suggested that an improved system should be considered to include additional features besides the results from previous matches, such as bookings (yellow and red cards), the teams' players, their managers and so on. In Nivard and Mei's report [92], five predictive models of the football end-result are presented: i) toto-models; ii) multi-independent score model; iii) single-independent model; iv) dependent score model; and v) pseudo least-square estimator score model. Although those models are named as match models, as the previous works, they are built to predict only the end-result of football matches. The author applied each of these models considering the number of goals of each opposing team and the number of goals scored by the home and away teams, tested in the English Premier League between 2007-2008 and 2010-2011. Unfortunately, although the models, in general, were able to predict the winner correctly in over 50% of the matches during a season, they were unable to accurately predict the final score in more than 15%.

In a more comprehensive fashion, Min et al. [1] proposed a framework for sports prediction using Bayesian inference and rule-based reasoning, together with an in-game time-series approach. These authors developed a football result predictor called Football Result Expert System. This was based on a Bayesian hierarchical model and depicted reasonable and stable predictions. The authors applied a procedure to estimate the value of the main effect, which was used to explain the scoring rate. Although their predictions were 95% accurate, their work only highlighted the teams with the highest propensity to score or concede goals, which is a major limitation of this study.

Later on, Farzin et al. [7] followed a Bayesian-based approach to predict the results of football matches, considering the influence of physiological and non-physiological factors in Futbol Club Barcelona (FCB) in the 2008-2009 Spanish League. Consequently, authors divided the dataset into two dimensions: i) non-physiological factors (e.g., weather, history of five previous matches, results against / for team, home game and players' psychological state); and ii) physiological factors (e.g., average age of the players, the number of injured key players, average number of matches in a week, key players' performance, team performance, and average number of goals for all home and away matches). The NETICA software was used to build the model, which yielded values for average age of the players as a medium, history of the last five games to win, injured main players, psychological state of players and weather conditions during the match. According to the authors, the obtained prediction accuracy to predict the 2008-2009 season FCB results was 92%. Nevertheless, the whole model was tuned for the specific case of the FCB, without describing a solid methodology, both technologically and scientifically, to assess the main factors that affect the final result, namely the physiological factors.

Alternatively to the previous authors, Igiri and Okechukwu [2] used the prediction system for football match results using both ANN and LR techniques, with the Rapid Miner Predictive Analytics Platform as a data mining tool. The authors highlighted several factors impacting the result of a football match, including home advantage and injuries, showing that the latter, especially in key players, play a significant role on the team's performance. This approach yielded 85% and 93% prediction accuracy for ANN and LR methods, respectively. However, not only the authors exclusively focused in the qualitative end result of each match (i.e., win, draw, loss), as it is unknown how the authors compute the learning rate and weights.

All these works put together allowed to confirm that many factors, such as state of mind, player's performance and physical capacity play an important role in algorithms for football prediction. Despite that, from all the previously presented variables the literature has been focused at to model the football game, physiological factors have been the least explored. Concerning this, Kramer et al. [93] proposed a muscle fatigue indicating parameter, regarding the changes that occur in surface electromyography (sEMG), that could be computed in real time by a simple analogue device, allowing real-time fatigue measurement for example for monitoring players during a football match.

Wavelet coefficients can be used in non-stationary and time-varying signal processing. Since EMG contains electrical signals related to muscle activity and the amplitude of these wavelet coefficients coincides with muscle fatigue development, these can also be proposed as features for identifying muscle fatigue for both static and dynamic contractions. As a result, Moshou et al. [94] proposed the implementation of an automatic method for muscle fatigue detection by using neural networks, in which a self-organizing map (SOM) was used to visualize the variation on the approximation wavelet coefficients, enabling the detection of muscle fatigue over time, by separating EMG signals from fresh and fatigued muscles.

Still around the physiological aspect of the game, but in a completely different perspective, Serfntein [95] proposed a statistical predictive equation comprising biomechanics, balance and proprioception, namely plyometric strength ratios of non-dominant leg plyometrics / bilateral plyometrics, dominant leg plyometrics / bilateral plyometrics, and non-dominant leg + dominant leg plyometrics / bilateral plyometrics, as well as previous injuries, all combined to determine a youth soccer player's risk of injury occurrence in the lower extremity. This study showed that it was possible to create a predictive model for noncontact youth soccer injuries based on a pre-season biomechanical, plyometric and proprioceptive evaluation, along with a previous injury history questionnaire.

These investigations allowed researchers to understand that prediction in football is more complex than as it seemed at first, depending not only on non-physiological factors, such as the history of team's results and players' performance in the field, but also depending on physiological factors, such as player's muscular activity [96]. Nevertheless, being able to identify the adequate data to be extracted and processed is of the outmost importance, namely to avoid falling within the Big Data phenomenon and move from a data-driven focus to a data-informed approach to improve athletic performance [11]. Next section highlights the most relevant signals for football prediction.

3.1.3 - Relevant Signals for Football Prediction

Due to the recent developments in sensor technologies, sport teams have now the possibility of using wearable support and monitoring tools in order to improve their results. From these, different time-varying signals, such as physiological and kinematic data, are acquired providing essential tips to physiologists, coaches and players about team's performance [97]. Hence, due to the recognition of the importance of these signals in the outcome of a football match, in the next subsections we will discuss in more detail their acquisition, analysis and classification. However, since researches directly applied to football are still much reduced, we will start by performing a more comprehensive discussion in which several works using these biosignals are presented, regardless of its applications.

3.1.3.1 Kinematic Signals

The study of the human body kinematics relies on the analysis of a wide range of parameters extracted from the movement, wherein the most important ones are the acceleration, the velocity and the position of the body joints, which can be measured using a wide range of technologies available in the market [98]. From these technologies, one can highlight time-of-flight cameras, infrared cameras with active and passive markers, and wearable sensors. In terms of ecological validity, wearable sensors tend to be an appropriate technology for the kinematic analysis of human movement over infrastructures of cameras that can only be of use if the person is constrained to a specific location [99].

3.1.3.1.1 Trajectory's Estimation

Although there are several studies regarding kinematics analysis, only a few have been reported within sport sciences literature [100]–[103]. In fact, it was only in a more recent study presented by Couceiro et al. [100] that a method to overcome automatic tracking problems of football players was proposed. In here, an adaptive FC approach was used to improve the accuracy of tracking methods by estimating the position of players based on their trajectories so far. The accuracy of the proposed approach was evaluated under different sampling periods of 250, 500 and 1000 ms, during one half-time of an official football match. Moreover, the authors studied the influence of using adaptive fractional coefficients $\alpha_n[t]$ when compared with predefined α_n values, demonstrating not only that adaptive $\alpha_n[t]$ generally results in a higher performance but also that smaller sampling periods result in better outcomes (Figure 3.2). Thus, they found out that dynamic FC is a viable approach for the study of football players' trajectories, increasing also the autonomy of tracking systems.

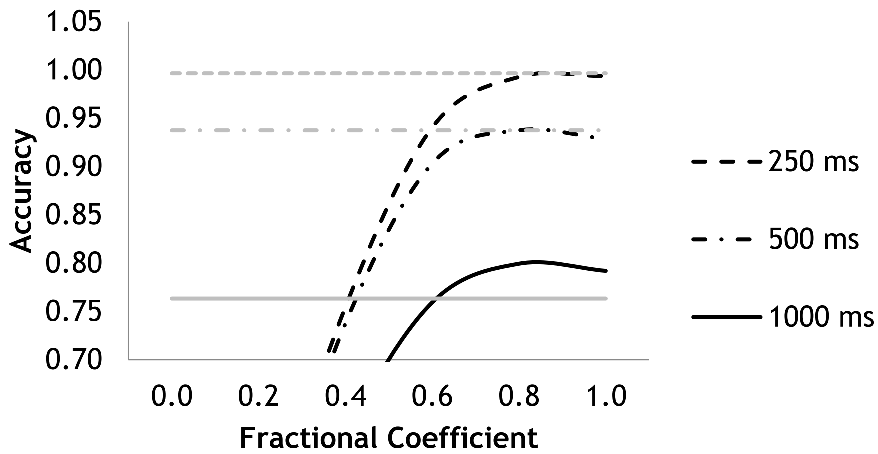


Figure 3.2 - Comparing adaptive fractional coefficients $\boldsymbol{\alpha_n[t]}$ with predefined values $\boldsymbol{\alpha_n}$ [100].

Meanwhile, Couceiro et al. [55] studied the predictability and stability levels of players during an official football match, since these are important concepts to fully characterize the variability of the whole team. As such, considering that FC can be very useful to understand player's motion, since it is a prediction method based on the memory of past events, a FC approach to define player's trajectory was considered. As such, to compare the variability that is inherent to the player's process variables (e.g., distance covered) and to assess his predictability and stability, entropy measures were considered. For this aim, the variability of the fractional coefficient over time was used in order to provide relevant information about a player's predictability whereas the stability was obtained by assessing the number of times that a player left his TR. Experimental results showed that despite being the most unstable player, the goalkeeper is the most predictable one. Moreover, they found out that the midfielders are the most unpredictable players and that lateral defenders are the most stable ones. This observations allowed to conclude that it is possible to observe that one player can be highly predictable (in terms of trajectory) while unstable (going outside of his TR) and vice-versa. As such, the authors further studied the relationship between predictability and stability (Figure 3.3), concluding that the level of predictability varies significantly with the positional main role of players. Note that the points are divided into four cluster, where the red circle corresponds to the goalkeeper, the blue triangles the defenders, the green lozenges the central players and the purple squares correspond to the forwards.

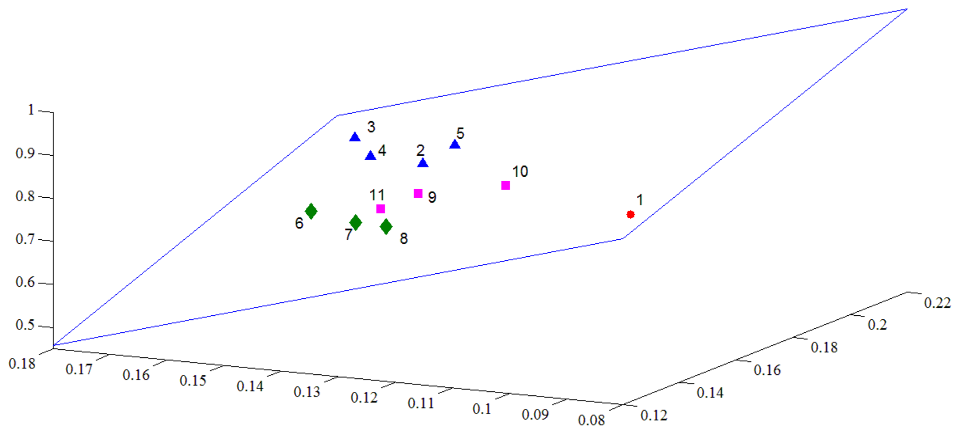


Figure 3.3 - Relationship between predictability α_n and stability β_n^{τ} (Red circle – goalkeeper; Blue triangles – defenders; Green lozenges - central players; Purple squares - forwards) [55].

Later on, in the study presented by Copete et al. [104] a method for estimating the trajectories of soccer players during ball motion and the direction of the ball after being kicked was proposed, by using dynamical prediction to estimate the future position of both players and ball. For this aim, the authors employed deep neural networks in order to represent the complex dynamics of temporal sequences of player positions and a one-hidden-layer neural network for estimating the orientation of the ball movement. To evaluate the proposed model, the authors used data from the RoboCup 2D Soccer Simulation League, concluding that the future movement directions of the players and of the ball can be successfully estimated.

Finally, in a more recent work [105], Lee and Kitani proposed a model for predicting the trajectories of the wide receiver in the game of American football. For this aim, the authors built a computational model which took into consideration not only prior knowledge about the game (e.g., route trees, defensive formations) but also short-term predictive models of how the environment changes over time. Since prior knowledge about the game is an information easy to access, the main goal here was to build predictive models of the environment. To do this, several models were proposed being the wide receiver modelled with a Markov Decision Process, in which the reward function is a linear combination of static features (prior knowledge) and dynamic features (short-term prediction of opponent players). Their results demonstrated that is possible to achieve better trajectories' estimation when we use more informed predictive models.

3.1.3.2 Physiological Signals

Due to the high physical demands of a match, the optimal physical and mental preparation of football players is now becoming an essential part of the game, being the focus of interest of many professional football associations. Besides of depending on player's motion, the performance of players in football matches also depends on other factors, such as player's technical and tactical skills. However, accordingly to several studies, when trying to improve teams' performance, physiological capabilities of players should also be considered [106], [107]. As a result, in the last few years, the analysis of physiological signals from players as a way to improve teams' results has gained a particular interest.

Physiological signals, like the heart rate (HR) or the galvanic skin response (GSR), are generated by the human body and they can be used to assess mental and physical stress of a person while (s)he is performing tasks or is involved in specific situations as they are often not consciously perceived and cannot be controlled by the person [108]. In this work, we will focus on the EMG signals. As such, in the following subsection a brief introduction to this type of signal is made.

3.1.3.2.1 – Electromyography

EMG is a technique used for recording and analyzing myoelectric signals. These correspond to the electrical potential (motor unit action potential) which is generated by the nerve cells that control muscle cells when they are electrically or neurologically activated [109]. There are several techniques for extracting the electrical signals from muscles, being these divided in surface EMG and intramuscular EMG, which is an invasive technique that measures the muscle activity by using a needle electrode directly into the muscle. Thus, EMG signals basically reflect the muscle contractions over time. Due to its simplicity, the EMG activity can be used in several areas and for various purposes. For example, it can be used to provide information about muscle health and diagnosis or even to study the functional movements, work conditions and postural tasks [110].

3.1.3.2.1.1 – Signal Characteristics

The typical amplitude for sEMG signals lies between 1-10 mV. This is a very low value, meaning that, usually, in order to collect data it is necessary to implement an amplifier. Regarding its frequency, the signal is between 0-500 Hz, being dominant mainly between 50-150 Hz. The peak frequency is typically located between 50 and 80 Hz. From that point the spectrum curves decreases and reaches zero between 200 and 250 Hz. Figure 3.4 illustrates the typical frequency power spectrum for EMG signals.

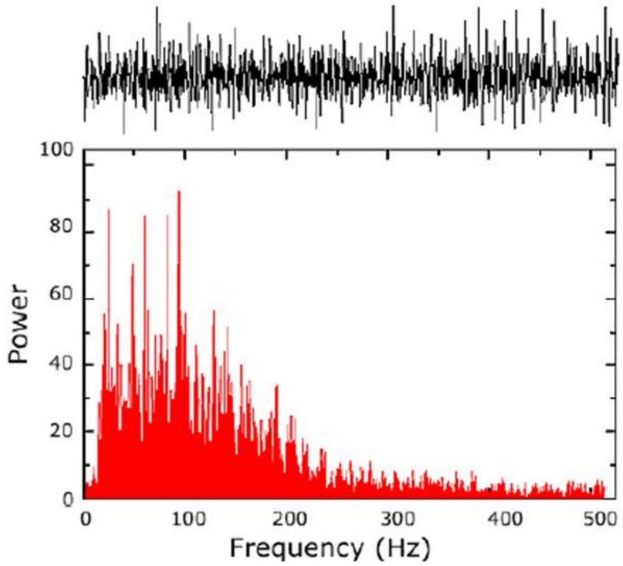


Figure 3.4 – Example of a typical Frequency spectrum for an EMG signal [111].

The sEMG is a signal highly influenced by noise, which can come from two main sources. Firstly, from ambient noise that is caused by electromagnetic radiation sources, like radio transmission devices. Secondly, EMG noise can be caused by movement artifacts. Because of that, the phase of filtering in EMG signals is of extreme importance [112].

3.1.3.2.2 – Actions' Classification

Over the past decade, the possibility of human movements detection and recognition has triggered an increased interest from many authors [113]–[115]. The analysis of human movement is of high practical interest in several areas, such as in sports and healthcare. In this context, motion recognition can be used to provide feedback to the user about his/her physical activity, thus, promoting a more active life [116]. Moreover, motion recognition based on wearable technology can be used to achieve optimal physical preparation of elite athletes [96], being that real-time monitoring can provide even more valuable information, allowing a higher performance and consequently increasing the chances of having good results. Yet, as every person performs movements differently, mainly due to their experience and ecological dynamics at stake, the recognition rate of human activities highly depends upon multiple causes [117], being that these recognition rates can be improved by adopting a careful selection of descriptive features.

Although the majority of research regarding actions' classification and motion recognition has been based on kinematics and kinetics data, the use of physiological data for this type of classification is now emerging as a viable alternative.

Regarding this, Chan et al. [118] proposed a fuzzy approach as a method of EMG signals' classification for a multifunctional prosthesis control. For this aim, time segmented features were clustered in an unsupervised method accordingly to the *Basic Isodata* algorithm in the training phase, being these results used to initialize the fuzzy system parameters. Then, fuzzy rules were trained by using the back-propagation algorithm. By comparing the fuzzy approach with an ANN method on four subjects, authors were able to conclude that although the classification results were quite similar, ranging from 70% to 90% depending on the subject, the fuzzy approach was considered superior to the latter in some points, as it demonstrated a slightly higher recognition rate, insensitivity to overtraining and higher reliability.

In another work, Jeong et al. [119] recognized EMG signal patterns of lower limb muscles by implementing neural networks during the recovery of postural balance of human body. Thus, features such as zero crossing, integral absolute value, spectral energy, central frequency and variance of central frequency were extracted from EMG signals and then used to classify the signals into five different categories, including forward perturbation, backward perturbation, lateral perturbation and two oblique perturbations. The motions were recognized with mean success rates of 75%.

Meantime, in the work presented by Joshi et al. [120], EMG from lower limbs was used in order to classify the eight different gate phases of a person. In here, four time domain features, namely, the mean absolute value (MAV), WL and variance slope sign changes, were used as well as the 4th order Autoregressive model in order to get the feature vector. The Linear Discriminant Analysis (LDA) was used for classification and a 50 ms second window was chosen. The results regarding the classification from both legs demonstrated a very low mean accuracy of 52.62%. Thus, in order to improve the classification results, time-synchronization and the Bayesian Information Criteria segmentation algorithm were applied, having resulted in an overall increase of the mean accuracy (75.32%).

In the context of football players, motion recognition usually involves a range of data collection techniques that comprise live observation and post-event video analysis, in which player's performance is manually assessed through its movement patterns. Due to the considerable time required to manually collect and analyse this data, the use of automated tracking technology in football teams is now becoming a more popular tool. However, inadequate video and computational facilities available at sports venues represent major challenges for the successful implementation of such technology. Moreover, since athletes are more quick and agile they tend to exhibit more complex movements with many unpredictable changes in direction as well as frequent collisions with other players. This type of behaviour violates the assumptions of smooth movement on which computer tracking algorithms are typically based.

Despite there are still few research for actions' classification based on physiological data, this is becoming an emerging area able to provide the basic tools to develop several potential applications for football matches, such as measuring team organization, planning tactics and strategies, providing objective measures of intervention effectiveness and providing meaningful physiological feedback, being this beneficial to the entire football club [121].

3.1.3.2.3 – Fatigue Detection

In the last three decades it has become quite common to evaluate LMF by means of sEMG signal processing [110], [122], [123]. Indeed, neuromuscular fatigue can be induced by sustained muscular contractions, being these usually accompanied by external manifestations such as the inability to maintain a desired force output, muscular tremors and localized pain [110].

When the muscle is fatigued the EMG signal displays two typical characteristics. Firstly, a change in amplitude, illustrated in Figure 3.5, due to the recruitment of additional motor units by the central nervous system, in order to maintain the required power output.

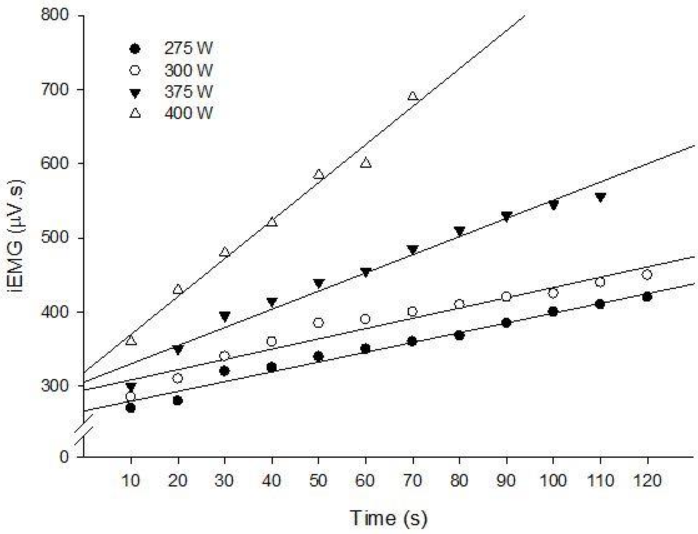


Figure 3.5 – Typical increase in EMG amplitude (represented on the y axis as integrated EMG (IEMG), μV.s) with time (s) at constant intensities (275, 300, 350 and 400 watts) [110].

Additionally, the power frequency spectrum suffers a leftward shift, i.e., the frequency properties decrease, as it is illustrated in Figure 3.6. This change in the frequency spectrum indicates the recruitment of more fatigue-resistant motor units to cope with the task constraint.

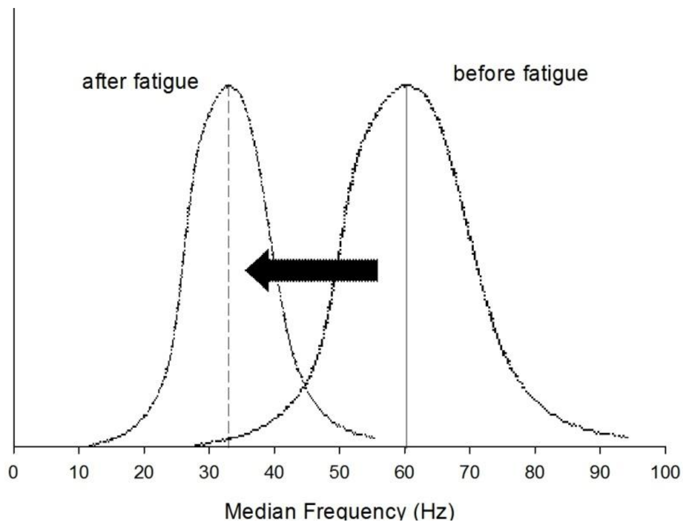


Figure 3.6 – Typical changes in EMG power frequency spectrum (Hz) of a muscle during a fatiguing contraction [110].

3.1.3.3 - Features and Classifiers for Biosignals

Modelling and selecting the features are steps of major importance that allow to provide accurate and opportune information about human motion, being through these steps that one can optimize classifiers for improved results. Although different classification methods may provide different levels of accuracy and precision, the choice upon the most adequate features may significantly propel the classification result to a completely different level. In this context, Avci et al. [124] presented a review regarding the most common techniques used for activity recognition using inertial sensors, exploring not only the main features used but also the most common classifiers. Regarding this, they divided the features in five types: i) Time Domain, ii) Frequency Domain, iii) Time-Frequency Domain, iv) Heuristic Features and v) Domain Specific, concluding that features such as mean, variance and spectral energy are very commonly used. In a similar work, Lara and Labrador [125] presented a review about different features extraction methods, divided into time domain, frequency domain and others. In the time domain, these authors used the most traditional features, such as mean, standard deviation, variance, interquartile range, mean absolute deviation, and others, such as the correlation between axes, entropy, and kurtosis. In the frequency domain, they referred the FFT and discrete cosine transform, though others were also discussed, such as the principal component analysis, LDA, autoregressive model, and Haar-like features. In another work presented by Bao and Intille [113], several features such as mean, energy, entropy, and correlation were used to assess which are the best methods of classification. To do this, the authors tested multiple classifiers, such as decision trees, instance-based learning, DT, and NB classifiers, concluding that DTs offer a higher classification accuracy for recognizing the chosen human activities. Similarly, Hyunn [114] explored the use of the same features, but with the addition of frequency-based features, such as the FFT, showing that the latter are enough to recognize some basic activities, like walking, standing, and sitting. To classify the activities, Hyunn chose SVMs, HMMs and NB. Furthermore, Zhang and Sawchuk [126] used a mobile phone with a sensory technology, considering features such as mean, variance, correlation, and entropy. Besides concluding that the worst classification methods were the KNN and the NB classifier, they also demonstrated that SVM was the method that depicted a superior performance. The work of Faria et al. refers other features extracted from a depth camera (RGB-D), such as the well-known Microsoft Kinect, which are modelled in such a way to characterize daily activities [52]. The authors considered energy-based features using the joint velocities, log-energy entropy-based features using skeleton poses, and sample autocorrelation-based features using the distances of skeleton poses in different time instants.

In another work, Jensen et al. [127] studied the performance of a set of generic features in two classification problems regarding biosignal analysis. To do this, two datasets were used, being that the first was built from physiological electrocardiography (ECG) data in order to assess activity levels of an athlete, whereas the second one was built from kinematic sensor data for the purpose of assessing experience level of a golf player. The feature set consisted of statistical moments, including Mean, Standard Deviation, Variance, Kurtosis and Skewness, and additional simple signal characteristics, such as Minimum, Maximum, Energy and Median. Then, several methods were used to classify the feature set, being these *AdaBoost*, LDA, NB, KNN and SVM with a linear kernel. The results obtained, showed LDA demonstrated the highest performance in both ECG data (88.8%) and kinematic data (90.2%) when using a generic set of features. Despite that, it was also visible a good performance by the other classifiers for the first dataset, since they all presented a classification rate equal or superior to 80%. Meanwhile, Mohd et al. [128] evaluated the performance of three different features of the Time Domain in the classification of EMG signals, concluding that standard deviation has the best overall performance when compared to the maximum amplitude of the signal and the root mean square (RMS). Meantime, Sharma et al. [129] presented a review concerning the definition of several methods and approaches suitable for extracting the features from EMG signals. The authors concluded that there are various efficient methods, such as Wavelet Transform approach or

Autoregressive methods. Moreover, they concluded that in the time-frequency domain, the most common features used are the Power Spectral Density, the spectral magnitude averages, the Thompson transform and the Short Time Fourier Transform.

Accordingly to Raez et al. [122], one of the ways to classify sEMG signals is by measuring the Euclidean distance between the motor unit action potentials (MUAPs) – electrical activity measured by the EMG [111] – waveform, in which a shimmer is generated in the representation of time-triggered and non-overlapping MUAPs. In another work, Boca and Park [130] suggested the implementation of a real-time application of an ANN, capable of accurately recognize the myoelectric signal signature. In here, signal's features were first extracted through Fourier analysis and then clustered by using the fuzzy c-means algorithm. Then, data was automatically targeted and sent to a multilayer perceptron (MLP) type neural network. Finally, a digital signal processor operated over the resulting set of weights, allowing the mapping of the incoming signal in real-time. The experimental results demonstrated that this approach produces highly accurate discrimination of the control signal over interference patterns.

Based on the literature, the use of ANN to classify EMG signals is one of the most popular methods within the scientific community. According to this, Al-Mulla and Sepulveda [131] suggested an evolved feature to predict the time to LMF by using supervised ANN, being this algorithm composed of five training inputs and one testing signal. To calculate the inputs' rate of change the first 20% of the evolved feature signal were used, enabling the simplification of this rate and, thus, promoting a faster ANN training. To adjust its training weights, the ANN used time to fatigue for the five training signals, allowing it to predict the time to fatigue by using only 20% of the total sEMG signal. The results presented in this study revealed an average prediction error of 9.22% for time prediction.

Concerning the use of EMG signals to detect muscle fatigue, the literature shows that the parameters normally used are the amplitude and the frequency of the signal [123], [132]. On the one hand, the amplitude of EMG signals increases progressively as a function of time when the fatigue increases. On the other hand, mean frequency (MF) and median frequency (MEDF) decrease with fatigue. Thus, muscle fatigue can be monitored by analysing changes in the EMG frequency properties [133]. As such, MF and MEDF are currently the most useful and popular frequency-domain features and frequently used for the assessment of muscle fatigue in surface EMG signals [123].

A summary of the most used features and classifiers is presented on Table 3.2 and 3.3.

Table 3.2 - Most common used features for biosignals.

TYPE	FEATURE	REFERENCES
Time Domain	Mean, RMS	[113]-[116], [126]-[128], [133]-[137]
	Standard Deviation, Mean Absolute Deviation and Variance	[113]-[116], [126]-[128], [135]-[137]
	Skewness, Entropy and Kurtosis	[113], [114], [125]-[127]
Frequency Domain	Fourier Transform	[114], [126], [129], [137], [138]
	Power Spectral Estimation	[113], [114], [136]-[138]
	Power Spectral Density	[129]
Time-Frequency Domain	Wavelet Transform	[129], [138]-[140]
	Autoregressive Models	[129], [138]
Signals	EMG	[128], [130], [131], [133]
	ECG	[127], [141]-[143]
	HR	[144], [145]
	GSR	[144]

Table 3.3 - Most common classifiers used for biosignals.

CLASSIFIER	REFERENCES	
KNN	[113], [114], [124], [126], [127], [140], [142], [146]	
Bayesian Models	[113], [114], [124], [126], [127], [147]	
DTs	[113], [141], [142], [148], [149]	
SVM	[114], [115], [124], [126], [127], [141], [142], [145], [146], [150]	
ANNs	[116]	
HMMs	[114], [115], [124], [126], [146], [151]	
DBMM	[152]	
LDA	[127], [141]-[143], [150]	
ANNs	Traditional	[130], [131], [143], [153]
	MLP	[130], [150]
	SOM	[141]

As it is possible to visualize in the previous tables, there is a set of features and classifiers that is more commonly used for algorithms' implementations with biosignals. However, it is important to realize that each problem is different and that what works well in one condition may not work under different conditions. As such, and taking into consideration the importance of features and classifiers' selection in the final result, in order to obtain optimized results it is crucial to experiment different combinations of features and/or classifiers. Bellow the definition of the most common features is summarized.

- TIME DOMAIN
  - Root Mean Square

$$RMS = \sum_{n=1}^{N} \sqrt{\frac{1}{N}\sum_{n=1}^{N} x_n^2} \tag{3.1}$$

  - Integrated EMG

$$IEMG = \sum_{n=1}^{N} |x_n| \tag{3.2}$$

  - Mean Absolute Value

$$MAV = \frac{1}{N}\sum_{n=1}^{N} |x_n| \tag{3.3}$$

  - Waveform Length

$$WL = \sum_{n=1}^{N-1} |x_{n+1} - x_n| \tag{3.4}$$

- FREQUENCY DOMAIN
  - Mean Frequency

$$MF = \frac{\sum^{M} f_j P_j}{\sum^{M} P_j} \tag{3.5}$$

  - Median Frequency

$$MEDF = \frac{1}{2}\sum^{M} P_j \tag{3.6}$$

  - Autoregressive Features

$$x_n = -\sum_{i=1}^{p} a_i x_{n-i} + w_n \tag{3.7}$$

Where x_n denotes the n^{th} sEMG signal sample, N is the length of the signal, f_j the signal frequency spectrum, P_j the sEMG power spectrum, a_i the AR coefficients, w_n is the white noise or error sequence and p represents the order of the AR model [154].

3.2 - Current Technologies for Football

Currently, extrinsic feedback information can be provided to coaches and athletes by several technologies, being some examples presented in this section.

3.2.1 - FIRSTBEAT SPORTS

FIRSTBEAT SPORTS is a platform designed for professional use that not only allows to monitor training loads and recoveries but also to increase team's performance. This technology allows coaches to effectively collect, analyse and interpret several performance data from the players, being this analysis based on the identification of individualized patterns of player's HR. Besides of providing a physiological analysis from all the players, FIRSTBEAT also contributes to injury prevention, being capable of assessing if they are overtraining. Moreover, it also allows to see the training progress as well as understand player's recovery and readiness [155].

Regarding its technical information, as it is depicted in Figure 3.7, FIRSTBEAT SPORTS is composed by several components: an interface software capable of operating in Windows 7, 8, 10 and Vista; a water-proof FIRSTBEAT Team Receiver, capable of collecting data with a maximum range of 400 m and a maximum of 80 players at the same time up to 700 hours (battery lifetime); a FIRSTBEAT Textile HR Belt, used for measuring player's HR and a High Performance HR variability recorder (Bodyguard) that can be used to record several physiological data during a match in order to help coaches to increase team's performance later [156].

Currently, this technology is already being used by over 600 professional teams, including FC Porto, PSG, Atletico Madrid and Sevilla FC [157].

Figure 3.7 - FIRSTBEAT Components [155].

3.2.2 - Catapult

Catapult is currently one of the global leaders on the field of athlete analytics, being the only system that is capable of measuring collisions. Besides of effectively monitor athlete's condition, Catapult also enables insight into athlete risk, readiness and return to play, giving a reliable tool for performance improvement. Moreover, this technology also offers the possibility of a more specialized monitoring regarding the goal keeper, quantifying with precision the dives (direction and intensity), jumps, accelerations and decelerations, changes in directions and recovery time, allowing a more personalized approach.

Catapult operates with a Global Navigation Satellite System, which allows higher accuracy rates, a wireless local positioning system to capture movement inside and out and one Inertial Movement Analysis for micro-movement analysis [158].

Due to its great popularity, this system has already been used by various international teams, such as SC Braga, Marítimo, Real Madrid and Toronto FC [159]. In Figure 3.8 the components of Catapult’s kit are presented.

Figure 3.8 - Catapult's Components [158].

3.2.3 - GPSports

GPSports is a sophisticated performance monitoring device that incorporates advanced global positioning system (GPS) tracking with HR and accelerometer monitoring, objectively assessing athlete’s and team’s loading across a range of intensities for distance, speed, acceleration, HR and impacts. As the previous one, GPSports is able to provide accurate feedback in real time, allowing coaches to confidently plan trainings and games, regarding speed, conditioning, strength and power in order to maximise athletic performance and minimise injuries. Moreover, as each session is automatically compared to the athlete’s training history, it provides more reliable information to coaches since the load of previous sessions can influence athlete’s future performance. Finally, this is capable of determining the team’s pre-match condition, by using market-leading chronic and acute load analysis, which helps coaches to objectively decide whether an athlete is fit or fatigued for a match [160]. GPSports is composed by a software that acts as an interface for coaches and by a hardware - SPI HPU -, which is currently the smallest and most powerful GPS on the market, delivering accurate and critical data from elite athletes. This has a battery life of over 8 hours, being capable of operate during 2 sessions per day with one charge [161]. Currently, this technology is been used by several teams around the world such as Chelsea Football Club and Valencia C.F.

3.2.4 - Mbody Pro

Mbody Pro is another system that can be used for football monitoring. Unlike the previous, besides of taking into consideration HR variability it also focus on muscle measurement, having as assumption that muscle’s status can have a great impact on training and by consequence, athletes' performance. Hence, by combining these two outputs, Mbody Pro opens up a new dimension in objective muscular performance monitoring, allowing to not only to understand what is happening with muscles during a match or a training but also to assess how muscles behave under different conditions in order to prevent injuries.

This technology allows both real-time and post exercise analysis of several parameters such as EMG, HR, cadence, distance, balance, stability, relaxation, speed, and coordination. As a result, Mbody Pro provides coaches and players a very useful tool to measure, monitor and analyse performance, allowing to not only target and optimize their technique, but also to define and gain more control over recovery processes and to detect and prevent problems, such as imbalances and deviations in the muscular system [162]. In Figure 3.9 a representation of the components of Mbody Pro System is presented.

Figure 3.9 - Mbody Pro System [163].

3.2.5 - STATS

STATS is one of the world's leading sports data and technology company, offering various solutions according to the needs and budgets of each football club. As such, depending on the goal of the team STATS has four different options: Live Video Analysis, Post-Match Analysis, Capture Tracking Data and Athlete Monitoring. Regarding the latter, they have two platforms: STATS Kinetic and STATS Dynamix. The first one was designed to better prepare players, decrease the injury risk and speed up recovery times, by integrating raw spatial, inertial, collision and cardio data. The second one is a more affordable online analysis platform designed for helping coaching staff optimize athlete's performance and minimize injury risk. For this aim, STATS Dynamix uses spatial data and integrates player tracking data and GPS information to allow coaches to compare and analyse data from different training sessions and matches [164].

3.2.6 - TraXports

TraXports is a wearable device manufactured by Ingeniarius, Lda. Due to its simplicity and comfort, this can be used to study athlete's performance in their natural competitive atmosphere in order to improve team progression. By combining information from an ultra-wideband (UWB) wireless technology with an inertial measurement unit (IMU) *TraXports* is able to provide players' position in real-time and over the internet. Figure 3.10 illustrates *TraXport's* equipment.

Figure 3.10 - *TraXport's* Equipment [165].

To estimate player's position, only four anchors are required (Figure 3.11), which not only allows the reduction of time setup but also the complexity of the overall system. The distances between anchors and wearable devices are estimated by using multiple wireless measures, discarding the need of any GPS. *TraXports* provides an high positional accuracy (±10 centimeters), under both indoor and outdoor scenarios, with a frequency update of 30Hz [165].

Figure 3.11 - Anchors of *TraXports* [165].

3.3 - Summary

As presented in this chapter, the state of the art shows that many authors have been presenting methodologies to predict the football match outcomes, though none explores all relevant features that influence athletes' performance and that might determine success, or lack of it, during a match. Indeed, there is a great potential in using Human-Machine interfaces to improve player's performance and therefore increase football prediction. However, defining a proper representation for all the input data is not always an easy task, since it often requires complex transformations. Besides that, there are still no established standard feature sets for low-dimensional biosignals, which increases the difficulty in this field [166]. Although this seems irrelevant, the fact is that the literature demonstrates that the choice of features and classifiers has a great influence in the final outcome. As such, one can easily understand the importance of having an optimization phase within the algorithm's architecture.

Moreover, one can conclude that there is a wide range of variables that might influence the football game outcome, and all those need to be studied as a whole and not in parts (*e.g.*, result of the match over time, individual performance of players, injuries, playing home or away, among others). Furthermore, the dynamic interplay of risk factors during sport activity and their relationship with injuries needs a comprehensive investigation. In this sense, an evaluation of each isolated risk factor does not take into consideration how the athlete performs the required functional movement patterns [167]. With a proper comprehensive investigation, one may go beyond forecasting the game outcome and be able to prevent sport injuries by developing the athletes coping skills and assist coaches in their decision-making [168].

Furthermore, by analysing the literature regarding the main technologies used for the context of football matches, one can notice that the majority of them is focused on the acquisition and analysis of physiological signals, mainly ECG, and that although there are some technologies that consider kinematic data, at the best of the author's knowledge there is still none which considers both kinematic and physiological signals, especially the EMG data from athletes. Additionally, none of the current technologies is capable of predicting variables, such as, position, actions and injuries, being this the main reason for the implementation of ARCANE. As such, taking into consideration ARCANE project, we proposed, herein, to contribute to the development of a predictive architecture for football, by estimating athlete's position, actions and health status during a football match.

Thus, in the following chapters the methodology concerning the architecture for position, action and health status recognition is described.

Chapter 4

Project Overview

This chapter first presents the overall approach for this master thesis. Afterwards, the technology used for collecting both kinematic and physiological data will be addressed. Regarding the latter, the construction of the wearable device will be explained in detail. At last, a brief description of the software adopted in this work will be also presented.

4.1 - Overall Methodology

As previously described, in this thesis we will work with two different signals used for different purposes: kinematic and physiological signals. More specifically, kinematical signals will be used for estimating the trajectories of football players during a match, whereas the physiological signals will be used not only for actions' classification but also for fatigue detection. As such, one can easily understand that it is necessary to establish different approaches for each type of signal, since they have different characteristics, thus, requiring different treatments. As such, in order to better understand the adopted approach, Figure 4.1 illustrates the overall methodology.

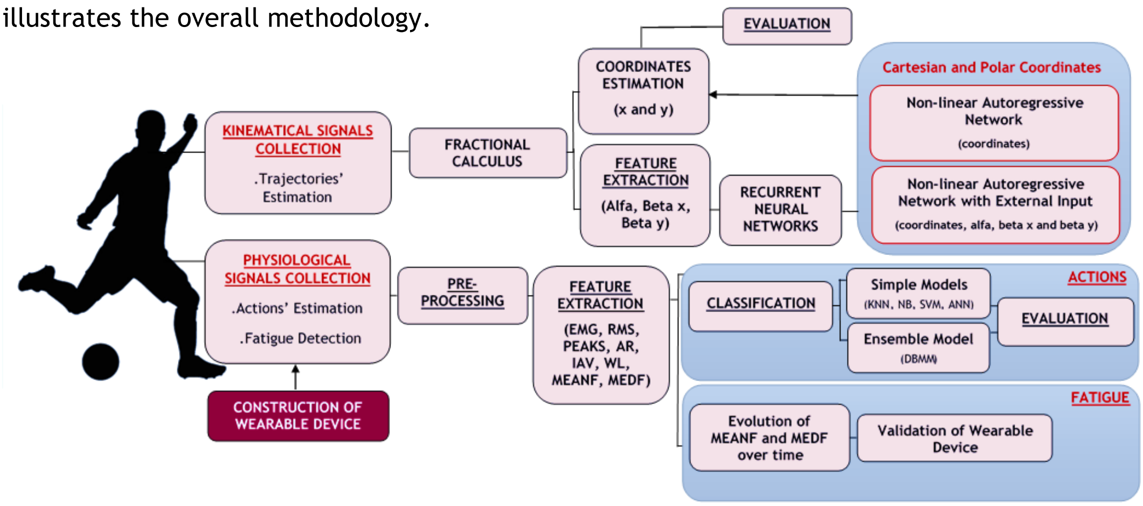


Figure 4.1 - Overview of the methodology adopted.

As depicted in the previous figure, two different approaches were established, one for each type of signal. Regarding the first one (*kinematic signals collection*), two different methods will be used for predicting the x and y coordinates of an athlete over time, namely the FC and RNNs. Concerning the latter, we will use two different types of RNNs, which only differ in the feature array that is fed to the network. Note that since kinematic data will be obtained by *TraXports*, which has already a Kalman filter implemented and, thus, gives as outputs signals already pre-processed, the phase of pre-processing will not be necessary for those signals.

Nevertheless, for physiological signals it will be necessary to build a wearable device, which will be explained in detail in subsection 4.2.2. After pre-processing and feature extraction, two different approaches will be tested: one for actions' classification and other for fatigue detection. Regarding the first one, two different sets of classifiers will be tested: four traditional models, namely, NB, SVM, KNN and ANN and an ensemble model, the DBMM.

Although, in general, different methodologies were established for each signal, one can easily notice that there is a common step between the both signals that is collecting data from the athlete. This is a very important step that should be performed with special attention, since obtaining a "clean" signal with little to no noise and representative of the data, be it kinematic but especially the muscle activity, can be quite challenging. As such, taking into consideration the importance of data collection in this project, in the next subsections, we will present in more detail the equipment used for collecting each type of data.

4.2 – Technology for Data Acquisition

4.2.1 - TraXports

In this work, data regarding human movement will be provided by *TraXports*, which was previously presented in section 3.2.6.

TraXports is a "pervasive" positioning system, with the potential to be used in any collective sports, under both outdoor and indoor environments. The proposed system overcomes the need for GPS signal and any complex infrastructure of cameras, requiring only the initial deployment of four stations, externally to the field, just before the match starts, thus resulting in a pre-installation time inferior to 15 minutes.

TraXports technology is based on 6.5 GHz UWB communication, compliant with the IEEE standard 802.15.4-2011 UWB, inertial sensors (IMU), and ECG monitor, presenting itself as a robust, precise and accurate state of the art positioning system. As it is demonstrated in Figure 4.2, the system comprises four stations (static nodes placed externally to the field) and a countless number of beacons (i.e., mobile nodes worn by the athletes), depending on the number of athletes one wishes to track.

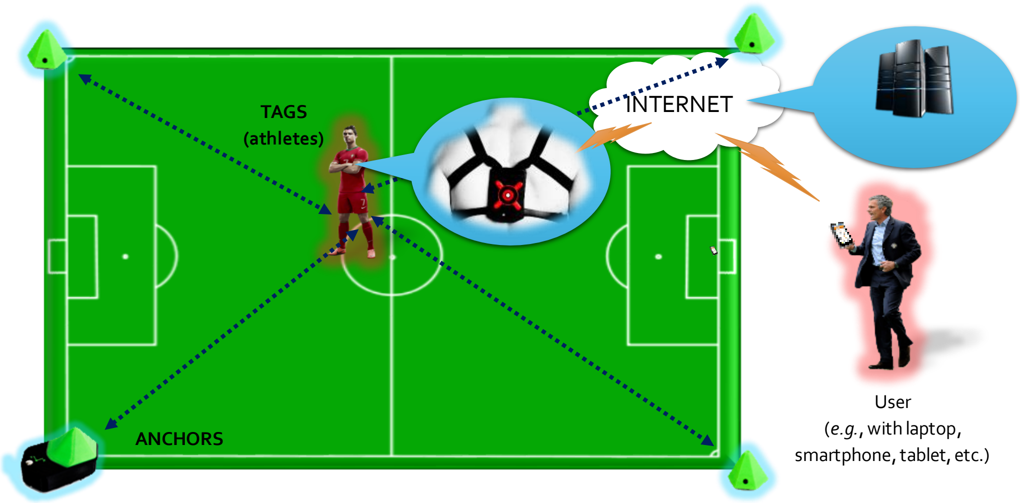


Figure 4.2 - *TraXports* positioning system [169].

This system is able to estimate the position of a particular athlete, virtually in any location due to the multi-sensor fusion between the UWB system, the gyroscope, the magnetometer and the accelerometer.

Each *TraXports* beacon comes fully integrated in a reduced size and volume wearable vest, placed on the athlete's back, presenting itself as a considerably lower cost solution when compared to all alternatives available in the market and previously presented. As the system benefits from wireless localization technology, this additionally allows to automatically associate *TraXports* to a cloud service, providing the ability to monitor, in real time and through the internet, athletes' performance during the match. All monitoring is performed intuitively, using web applications, that allow to store and analyse the data without any incompatibility issues associated with the platform (i.e., computer, tablet or smartphone) nor the operating system (i.e., Windows, Linux or MAC OC) of the operator (e.g., coach, sport analyst, researcher or spectator) [165].

4.2.2 – Conductive Shorts

In this project, in order to obtain the data regarding the physiological signals, namely the EMG signals, it was necessary to build a new wearable device, which was set up to measure the electrical activity from the muscles from the right thigh as a proof-of-concept. In the future, this wearable device will be extended to also measure the electrical activity from the muscles from the left thigh.

4.2.2.1 – Acquisition System

The muscles' electrical activity was acquired by integrating the MyoWare™ Muscle Sensor (AT-04-001), from Advancer Technologies, into the wearable solution (See Appendix A.1). This can be directly plugged into 3.3V-5V development boards and allows to measure two different outputs, the amplified raw EMG and the EMG linear envelope, being possible to adjust the gain for the latter. Additionally, due to its wearable design, Myoware Muscle Sensor can be used in two different ways. Either the electrodes can directly snap to the sensor, without the need of using extra cables and wires, or the sensor can be combined with other shields to display the signal or even to easily supply energy to the sensor. In the case of the latter, a Cable Shield, as well as a set of 3 electrode cables were also used in order to connect the sensor to the electrodes that were further away from the main circuit. Note that, since the electrodes herein used were made from conductive fabric, the electrodes' cable was cut to use only the part of the cable with the jack entry instead of the heads of the electrodes. Figure 4.3 demonstrates the components used with the Myoware Muscle Sensor.

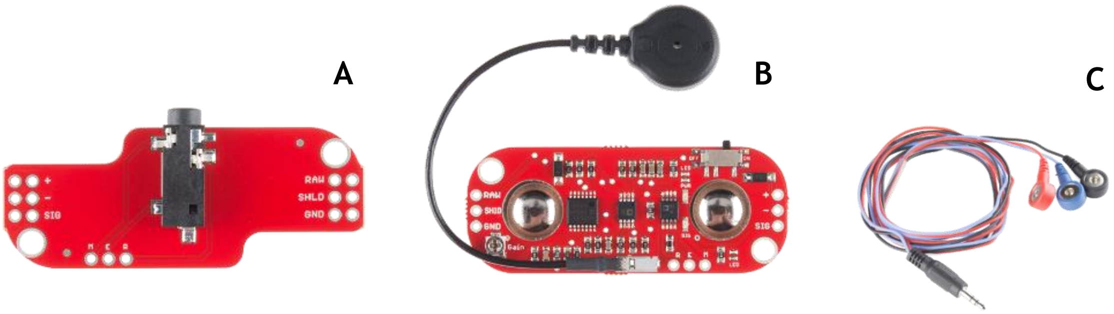


Figure 4.3 - MyoWare Muscle Sensor components. **A)** Cable Shield; **B)** EMG sensor; **C)** Electrodes Cable.

Since this solution had the requirement of being wearable, it was necessary to use a microcontroller that allowed to transmit data without using cables. To do so, the Particle's Photon was used (See Appendix A.2). This is a Wi-Fi connected microcontroller that combines a STM32 ARM Cortex M3 microcontroller with a Cypress Wi-Fi chip. The power to the Photon can be supplied via on-board USB Micro B connector or directly via the VIN pin. Due to the need already addressed before, in this work, the power was directly supplied to the VIN pin by a 550 mA 3.7 V LiPo rechargeable battery with a typical average current consumption of 80mA and a usual time duration of 2 hours. Figure 4.4 illustrates the Photon and the battery herein used.

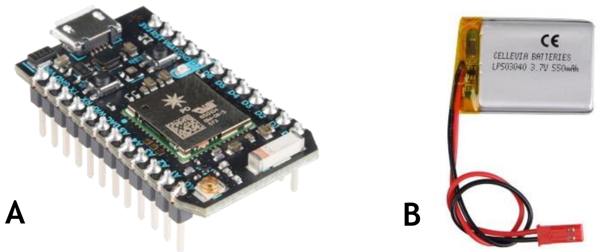


Figure 4.4 - Components of the wearable shorts. **A)** Photon **B)** Li-Po Battery.

In order to illustrate the system's electronics, the circuits' schematics is presented in Figure 4.5.

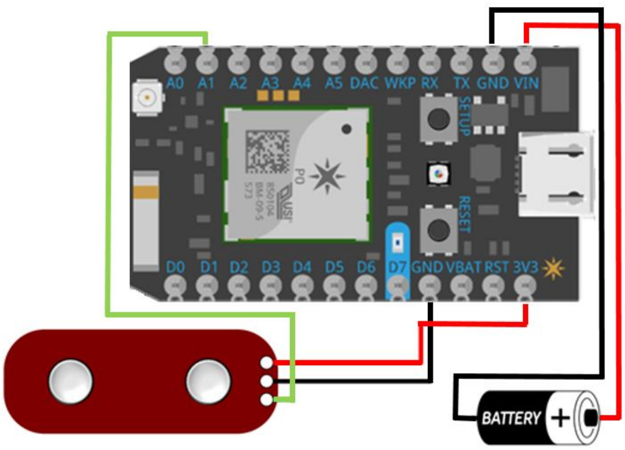


Figure 4.5 - Circuits' Schematics of the wearable shorts.

It is noteworthy that the EMG sensor outputs analogue values, being this connected to an analogue pin in the microcontroller. As such, to receive as output the real voltage values at each instance, the following equation was adopted.

$$V = \frac{S}{2^b} V_{cc}\ [Volts] \tag{4.1}$$

Where S are the sensor analogue readings, V_{cc} is the operating voltage (3.3 Volts for Photon) and *b* is the number of bits regarding the resolution of the Analog-to-Digital Converter (12 bits for Photon).

Figure 4.6 shows the integration of components addressed above (Photon, Myoware Muscle Sensor, Cable Shield and Li-Po Battery).

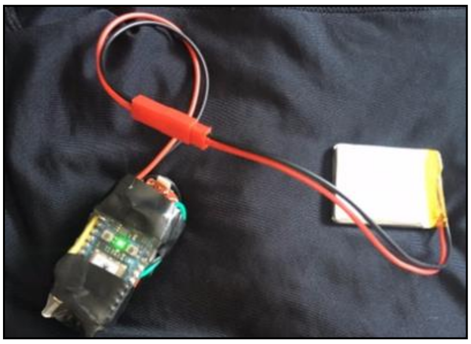


Figure 4.6 - Demonstration of the main electronic components of the wearable shorts.

As it was already mentioned, the electrodes used for detecting the electrical activity of the muscle were made from a conductive fabric, the MedTex130, which is commonly used for conducting electrical current in wearable devices (See Appendix A.3). This is formed by silver-plated nylon that is capable of stretch in both direction, being highly conductive with a surface resistivity of < 1 ohm/sq. Figure 4.7 illustrates the conductive fabric used in this project.

Figure 4.7 - Conductive Fabric MedTex130.

Thus, in order to implement the electronic circuit on the shorts, three rectangular strips of conductive fabric were sewed in three specific points, according to the electrodes' placement illustrated in Figure 4.8.

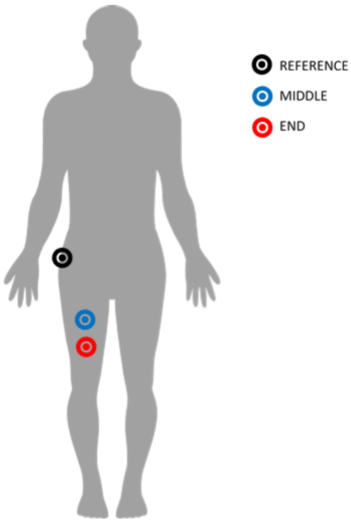


Figure 4.8 - Electrodes' placement for the right lower limb muscle.

It is noteworthy that the strips were strategically placed to analyse the muscle activity from the right tight. One strip was placed in the area above the end of the muscle (in red), the other was sewed approximately 1 cm above the former, in the area that corresponds to the middle zone of the muscle (in blue), and the final strip was used as reference, being sewed distantly from the other two, in this case, above the iliac bone.

A zigzag stitch was used in order to maintain the fabric's characteristics, namely its stretchiness and flexibility. This is basically an overcast stitch in both directions that allows the conductive fabric to stretch in both directions without destroying the fabric.

Afterwards, the electrode cable snaps were added to the shorts by attaching one male snap to each of the conductive fabric strips. This step allows to perform the connection between the EMG sensor and the electrical activity of the muscle.

Note that in order to have a wearable device that could fit the majority of people, medium-sized shorts were bought. Although they possess electronic components, these can easily be washed in a regular washing machine, since the conductive fabric is also washable.

Figure 4.9 shows the wearable solution developed in the context of this thesis, with the conductive fabric sewed and the electronic components connected.

Figure 4.9 - Conductive wearable shorts used for EMG signal collection. **A)** Illustration of the electrical components. **B)** Illustration of the electrical components stored in a pocket.

Finally, as previously stated, the EMG data was transmitted through wireless communication. Particle's Photon can be programmed in an Arduino-based Particle Integrated Development Environment available on Particle's site [170]. This interface allows to subscribe and publish messages, in this case, the EMG voltage values to their Cloud, allowing to store data on the internet. However, due to its low capacity for receiving data (255 bytes per second) and due to necessity of sending a huge amount of data per second (5000 bytes per second for a sampling frequency of 1000 Hz), it was decided to use an external Cloud for sending data. In order to do this, the Photon was reprogrammed to connect to the external company's Cloud, namely Ingeniarius Cloud[1], which was partially replicated in a Raspberry Pi communicating with the Photon over Wi-fi. Thus, instead of transmitting the EMG data to Particle's Cloud, the Photon transmitted the data to Ingeniarius Cloud, which was already prepared for extracting data in the .csv format file. However, despite the use of Raspberry Pi for storing the EMG data, it was not possible to increase the signal's sampling frequency to 1000 Hz (the recommended frequency for EMG signal's acquisition), due to the limitations of Wi-fi connectivity and storage capability of the Photon. Therefore, it was only possible to ensure a sampling frequency of 100 Hz. Figure 4.10 illustrates how data circulates over the multiple components of the system.

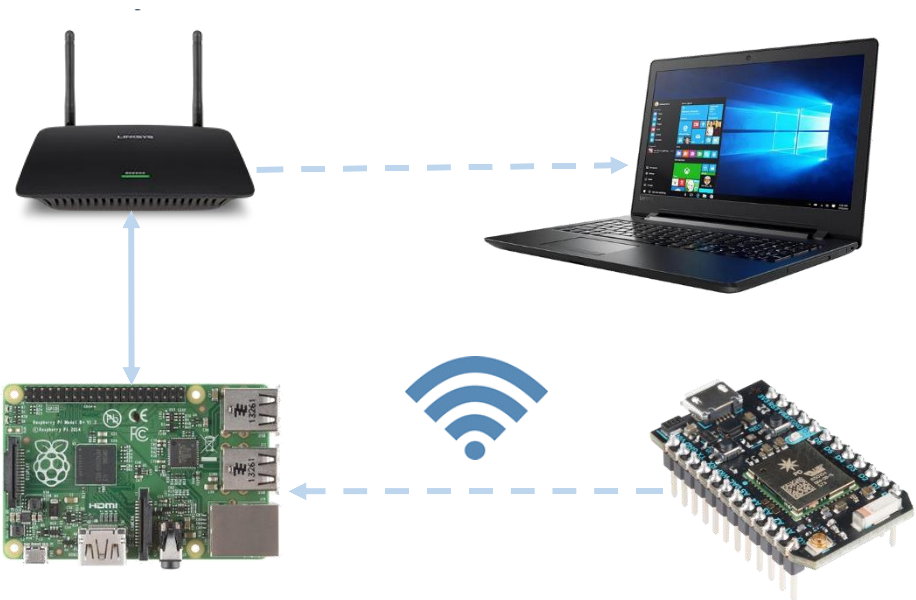


Figure 4.10 - Illustration of EMG data transmission.

To visualize the data, a laptop was also connected to the same Wi-Fi network, thus allowing to observe the incoming data while the user performs the experiment with the shorts. Figure 4.11 shows two different participants wearing the wearable shorts, in which henceforward we will call *TraXports V2*.

[1] https://cloud.ingeniarius.pt/

Figure 4.11 – Example of two participants wearing TraXports V2.

4.2.2.2 – Ecological Validation

To assess the viability of *TraXports V2*, for usage in future works, a group of 20 people was asked to use this device during a small period (t=5 minutes) while doing physical activities (walk, run, jump, etc) and to rate from 1 to 5 several aspects of the device, being 1 very bad and 5 excellent. As such, the following factors were assessed:

1. Ease of use: In here, the participants reflected about the overall usage of the conductive shorts, for example, if it was a difficult task to put on the shorts (due to the electronic components) and even if it was difficult to perform the several different activities while using the shorts.
2. Comfort: In here, the participants assessed the shorts in terms of comfort, i.e., if they felt comfortable while using the device or if they felt discomfort.
3. Aesthetics: This factor reflects about the overall appearance of the wearable device.
4. Durability: This factor is related to the participants' point of view regarding the overall system capacity of operating with batteries during a long time. Regarding the durability of the wearable shorts over time, i.e., after different usages, this can only be assessed in a future work.
5. Material/Fabric: In here, the participants expressed their opinion regarding the fabric of the shorts. In this case, it was an elastic fabric, suitable for sports. Thus, their answer is related to the questions such as: is the fabric so thick that it does not let the sweat evaporate? Or is the fabric so thin that it is possible to see the underwear?
6. Suitable for sports: This factor express the participants' opinion regarding the suitability of this wearable device for sports context. For example, if they feel that this is a device with utility or not, if they imagine this kind of device being dressed by football players, among others.

Figure 4.12 illustrates the results.

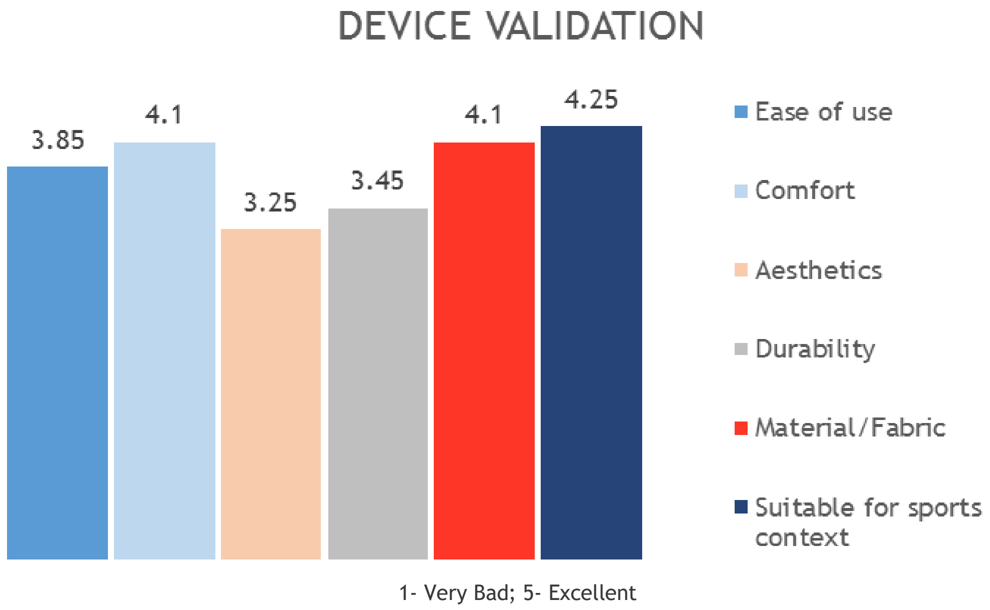


Figure 4.12 - Results regarding *TraXports V2* viability.

As it is possible to observe, the overall opinion is that this is indeed a viable device, especially in terms of being suitable for sports context, which was the aspect that obtained higher score. On the other hand, we can also notice that, from the users' point of view, this is a device that still needs to be improved, mainly in terms of aesthetics and durability. Regarding the fabric used, we can conclude that it is a good choice for this type of project, since the participants felt that it was a comfortable and good material. Nevertheless, it is unsure how its conductive capabilities will change with use, namely with sweat. As this is out of the scope of this work, a future exhaustive evaluation of its durability should be carried out.

4.2.3 – Software

Matlab is a software developed by MathWorks for engineers and scientists, that provides an environment in which it is possible to perform several mathematical computations, such as matrix manipulations, functions and data plotting, algorithm's implementation and interfaces' creation, being also able of interfacing with other programs written in various languages, including C, C^{++}, Java and Phyton.

In this master thesis, all the main work will be performed using Matlab, since this will be necessary not only for receiving the signals (kinematic and physiological), but also for analysing and classifying them, and for constructing the proposed architecture [171].

4.3 - Summary

In this chapter, the overall methodology for this master thesis was presented, in which it was explained that we will work with two different type of signals, namely the kinematic and physiological signals. Considering this and the different kind of characteristics that each signal has, it was necessary to define two different approaches, one for each type of data. As such, regarding kinematic data, we will first use two different methodologies for estimating the players' coordinates, FC and RNNs. However, one should point out that the FC approach is not our main target of study, since this has already demonstrated to perform well in this type of studies. In fact, our main aim is to use this method to extract suitable features, including the predictability and stability coefficient, in order to use them during the implementation of several RNNs. On the other hand, for physiological data we will have two main objectives, the classification of several athletes' actions and detection of fatigue in athletes.

Since the phase of data collection was a big part of this thesis, we then presented the two devices that were used to collect both kinematic and physiological data. Note that, for the latter, there was no device available. As such, it was necessary to construct a new wearable device capable of measuring EMG signals from the right thigh of subjects, being this process clearly explained. This was our first contribution to this thesis.

Finally, the software used for the analysis of all the collected data was presented.

Chapter 5

Kinematic Data

This chapter addresses the architecture for trajectories' estimation. Here, we will present two main subjects. Firstly, we will describe in detail the data collection process, as well as features' selection, being the latter based on FC. Afterwards, we will implement several RNNs, based on the features extracted with the FC method, and we will present a comprehensive study in order to compare the performance of FC and RNNs.

5.1 – Data Collection

The kinematic signals, namely pose and orientation, were provided by Ingeniarius since there was no possibility of using *TraXports* during the period of this project. As such, Ingeniarius provided a dataset acquired with students from the Faculty of Sport Sciences and Physical Education of Coimbra during a handball match. The acquisition had a time duration of approximately two hours, taking into consideration the setup time. Data collection was divided in several periods of 8 to 16 minutes, in which the players successively switched positions between each other (e.g., from defence to offensive position). Since the aim here was to predict the coordinate of each player in the following second, the *x* and *y* coordinates as well as the orientation of the player at each second were collected. The kinematic data was collected with a sampling frequency of 33 Hz.

5.2 - Feature Extraction

The choice of the features for kinematic data was made taking into consideration the state of the art presented in subsection 3.1.3.1 and the output given by *TraXports*. As such, we defined the following features:

1. Pose:
    a. Coordinate *x* of the player.
    b. Coordinate *y* of the player.
2. Predictability (fractional) coefficient, which shows if the player's trajectory is more or less predictable.
3. Stability Coefficient, which informs about the stability of the player.

It is noteworthy that two different sets of features were used. The first one, only uses the *x* and *y* coordinates and is aimed for the FC approach whereas the other uses all of the features and is intended for the RNN approach. Table 5.1 summarizes the choice of features.

Table 5.1 - Summary of features for kinematic data.

DOMAIN	FEATURES	SET	
Kinematic Signal	Coordinate x	1	2
	Coordinate y		
	Predictability Coefficient		
	Stability Coefficient		

5.3 - Estimation of athletes' trajectories

In the context of ARCANE, during a real time football match, predicting the position of all the players, including the opposite team, with a relative advanced time (5-15 seconds) can be seen as an advantage. Thus, by receiving this information in real time, the coaches are capable of taking more conscious decisions about the flow of the game, being, thus, more susceptible to win.

5.3.1 – Fractional Calculus

The estimation of players' trajectories over time through the FC method was performed by considering the literature presented in Section 2.2, in which the method of alfa calculation (fractional coefficient) and its further optimization is explained in detail. As such, the x and y coordinates of each player were calculated for each iteration, as well as the alfa and beta (along x and y coordinate) to each player. Note that the FC approach was used, herein, with the main goal of extracting meaningful features from the kinematical signals to be used with the RNNs, as it is represented in Figure 4.1, since FC is a method that already proved to be efficient in trajectories' estimation [100], [101]. As such, in order to obtain an optimized training dataset with the best choice of features, we first analysed the influence of different sampling frequencies on coordinates' estimation with the FC approach. To do that, the original dataset was downsampled for the frequencies of 30, 10, 2, 1 and 0.5 Hz and the error in terms of mean Euclidean distance (MED) was calculated, according to equation 5.1.

$$MED = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \tag{5.1}$$

The results regarding the error between the real and the estimated coordinates for each iteration over time and for Player 1 are illustrated in Figure 5.1.

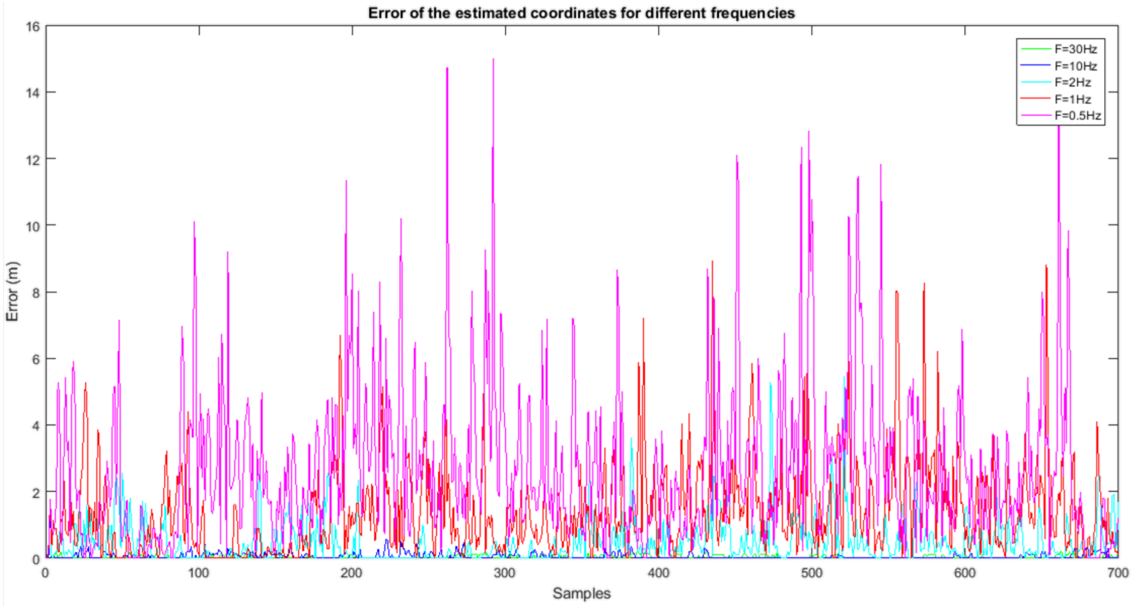


Figure 5.1 - Influence of different sampling frequencies on coordinates' estimation.

By analysing the graph presented above, it is possible to clearly see that the Euclidean distance (error) increases with the decrease of frequency. To assess the overall error for each frequency, the mean error of all players during the entire period of time was also calculated, being these illustrated in Figure 5.2.

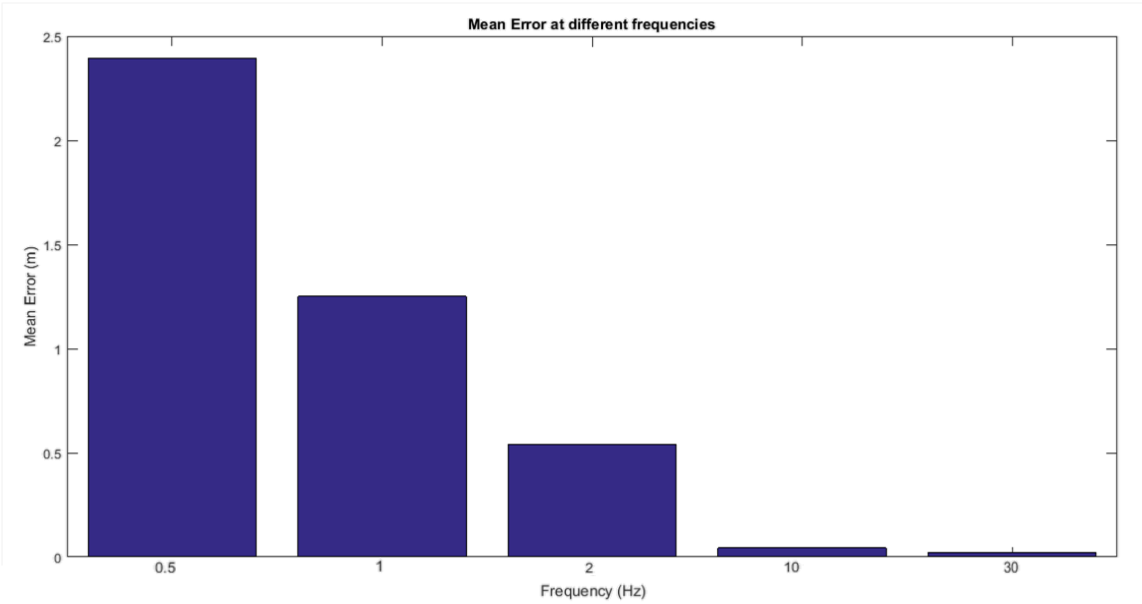


Figure 5.2 - Mean error at different sampling frequencies.

Similarly to the results presented before, in here, one can easily identify a general increase in the error for lower frequencies. For example, for frequencies higher than 2 Hz, the mean error is less than 0.5 meters. On the other hand, for lower frequencies, such as for 0.5 Hz, the overall error is greater than 2 meters, which is not acceptable in a real time football game situation. Indeed, for the FC approach in which the only inputs used are the x and y coordinates, one can state that it is better to use higher frequencies in order to obtain better results. However, since we are assessing which is the best frequency to use for the RNNs approach, we have to take into consideration the variation of alfa and beta over time as well. As such, by analysing the definition of alfa and beta, it is possible to infer about its variation with different frequencies, taking into consideration that, for higher frequencies, it is likely to have more than two consecutive samples with the exact same value (coordinate). For example, with a frequency of 30 Hz, the system is acquiring data from the players at each 33 ms, meaning that the players would have to move to different positions at each 33 ms in order to have consecutive samples with different coordinates, being this not representative of a real match situation, meaning that although lower frequencies produce higher errors, these are more useful for the context of a football match. From equation 2.24, despite the optimization of alfa being dependent on the coordinate at time t-1, if the results between two consecutive samples are equal, this will not affect negatively the value of alfa, meaning that there will not be values of alfa that are not considered a number (NaN) or that produce infinite values (Inf). On the other hand, by analysing the definition of beta (equation 2.25), one can infer that for higher frequencies there is a higher chance of getting poor results for beta. As such, although higher frequencies proved to produce smaller errors between the real and the estimated coordinates, it is also noticeable that they also produce poorest results regarding the stability coefficient, since there are several consecutive samples which have the same value. Thus, in order to use the whole set of features in the RNNs algorithm it is important to choose a frequency that is able to deliver an equilibrium between the error and the beta values. After analysing all the data, we chose the dataset that was built at a sampling frequency of 1 Hz, not only because it produces acceptable errors and beta values, but also because this is the most viable frequency to be used, if we consider a real game situation, i.e., the coach will receive the coordinate estimation for each player for the following second. This is a viable period of time which is not affected by delays that may happen during the game.

After assessing the error with the FC approach and defining which was the most viable frequency to be used with the RNNs algorithm, the next step was to test the chosen dataset with different RNNs. Hence, in the following subsection, an extensive study regarding the ability of RNNs to estimate trajectories is presented.

5.3.2 – Recurrent Neural Networks

To evaluate the performance of RNNs on coordinates' estimation during a football game, several RNNs were tested, being these divided in Non-linear Autoregressive Network (NAR) and Non-Linear Autoregressive Network with External Input (NARX), which only differ on the type of data that is fed to the network to estimate the coordinates. As such, for the first case, the network only receives as inputs the x and y coordinates of the player over time with the intent of estimating the subsequent position of the player, whereas, for the other, besides receiving the player's coordinates, the network is also fed with an external input, which in this case were the predictability and stability coefficients. Additionally, since the x and y coordinates are not fully representative of the player's movement, the RNNs were also tested with polar coordinates, according to the following equations:

$$r = \sqrt{x^2 + y^2} \tag{5.2}$$

$$\theta = \arctan(\frac{y}{x}) \tag{5.3}$$

In order to implement the RNNs, the Matlab Neural Network toolbox was used[2], defining networks with 10 hidden layers and without performing multi-step prediction, i.e., the estimation of the coordinates at t+1 took into consideration the previous real coordinates instead of the previous estimated coordinates. Note that to simulate the prediction in real time, the algorithm was set to successively and cumulatively add more data to the training and testing set at each iteration, starting with a training set composed by the first 80 samples and the testing set composed by the following 20 samples. At each iteration, 20 new samples were added to the training set and the following 20 samples were tested. Moreover, in the case of the polar coordinates, the RNNs were fed with the players' polar coordinates, which were converted again to Cartesian ones and compared to the real Cartesian coordinates after the RNNs predictions were finished. Thus, we implemented several RNNs, having first performed a study with the player' coordinates in the Cartesian form (N1-N4) and, then, a study with the polar coordinates of the player (N5-N8).

Thus, several RNNs were implemented being these detailed below:

- N1 refers to the implementation of a network that only uses as inputs the variables that we want to estimate. In this case, the variables were the real Cartesian coordinates from player 1 over time (*x* and *y* coordinates).
- N2 refers to the implementation of a network that learns to predict the time series (*x* and *y* coordinates) given past values of the same time series, *i.e.*, the feedback input, and another time series, in this case, the predictability and stability coefficients (alfa and beta).
- N3 refers to the implementation of a network that is similar to the previous, only differing in the external input given. In this case, in order to understand if the external input that was being given was the most accurate, we tested these variables (alfa and beta) separately. This networks has only as external input the predictability coefficient alfa.
- N4 refers to the implementation of a network that for x coordinate estimation uses the real x coordinate as internal input and the predictability coefficient (alfa) and stability coefficient (beta x), whereas for the estimation of y coordinate uses the real y coordinate as internal input and the predictability coefficient (alfa) and stability coefficient (beta y).
- N5 refers to the implementation of a network that only receives internal input. In this case, since the input were the real polar coordinates of the player, the network estimated the polar coordinates as well. Thus, in order to make a meaningful comparison, these outputs were then transformed in Cartesian coordinates and the respective errors were calculated.
- N6 refers to the implementation of a network with polar coordinates as internal inputs, and the predictability (alfa) and stability coefficients (beta x and beta y) as external inputs.
- N7 refers to the implementation of a network similar to N3, in which the polar coordinates are used as internal inputs and the predictability coefficient (alfa) is used as external information.
- N8 refers to the implementation of a network in which the polar coordinate *r* was used as internal input. Regarding the external input, beta x and beta y were first calculated

---

[2] http://www.mathworks.com/products/neural-network/

in terms of polar coordinates (See equations 5.2 and 5.3) and then used along alfa as external inputs.

Table 5.2 presents the results obtained regarding the prediction of player's 1 coordinates with different RNNs, in terms of MAV, RMS and the MED.

Table 5.2 - Comparison between the performance of NAR and NARX under different constraints.

COORDINATES	NETWORK			MAV (m)	RMS (m)	MED (m)
CARTESIAN	NAR (x, y)	N1	x	0.93	1.48	1.17
			y	0.52	0.95	
	NARX (x, y, alfa, beta x and beta y)	N2	x	0.86	1.53	1.11
			y	0.50	0.76	
	NARX (x, y, alfa)	N3	x	0.10	1.70	1.10
			y	0.45	0.74	
	NARX (x, alfa, beta x) (y, alfa, beta y)	N4	x	0.83	1.60	0.92
			y	0.38	0.66	
POLAR	NAR (r, teta)	N5	x	0.80	1.78	0.87
			y	0.12	0.21	
	NARX (r, teta, alfa, beta x and beta y)	N6	x	0.73	1.22	0.76
			y	0.12	0.19	
	NARX (r, teta, alfa)	N7	x	1.87	0.84	0.83
			y	0.21	0.13	
	NARX (r, alfa, sqrt (beta x,beta y), atan (beta y/beta x))	N8	x	0.78	1.22	0.81
			y	0.13	0.22	

Besides of testing with the usual combination of features (N1, N2, N5 and N6) we also tested different combinations of features to obtain the optimized feature vector. As such, the analysis of the results presented on Table 5.2 allows us to better understand the performance of different RNNs on coordinates' estimation. Indeed, as it was expected the networks that do not use external inputs produce worst results (MED= 1.17 m for Cartesian coordinates and MED=0.87 m for polar coordinates). This results confirm that the predictability and stability coefficients are good parameters to be used for football players' coordinates prediction for RNNs approaches. Moreover, when comparing the RMS obtained for the x and y coordinates, it is possible to identify that y coordinates always present lower values. This can be explained by that fact that the x coordinates presented a higher variability between samples, which leads to a more difficult training during the RNNs' implementation and, by consequence, increases the error.

Moreover, when comparing the error obtained with the RNNs that were implemented with Cartesian coordinates (N1-N4) and the error obtained with the RNNs that were implemented with polar ones (N5-N8), one can easily observe that there is a significant decrease of error when the RNNs are implemented with polar coordinates, as it is illustrated in Figure 5.3.

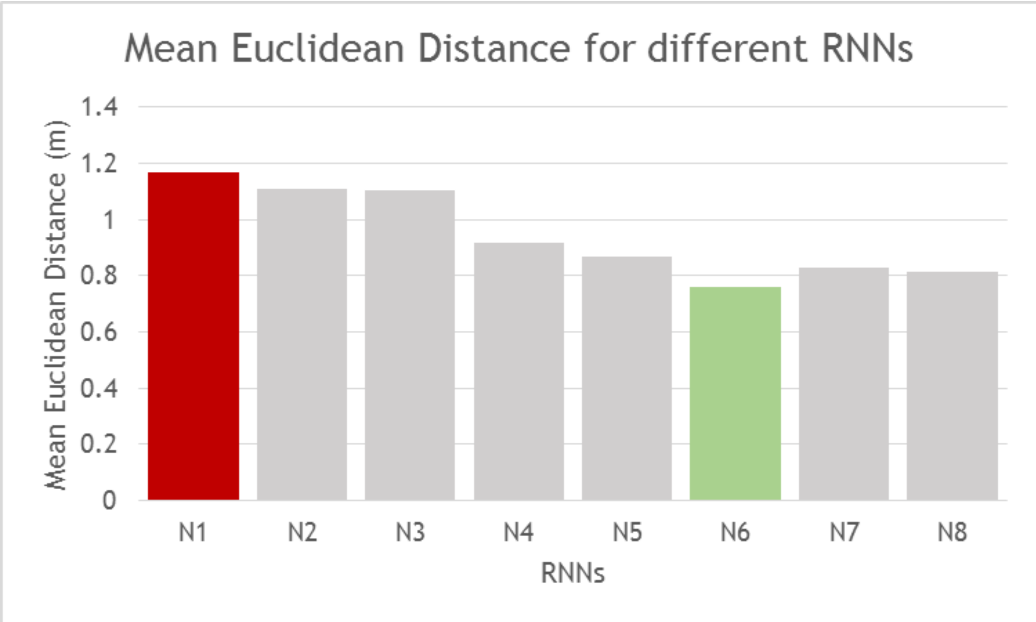


Figure 5.3 - Comparison of the MED between different RNNs.

From this graph, it is also possible to clearly visualize that the network N6 (in green), i.e. the NARX that used as input the polar coordinates (r and teta) as well as the fractional and stability coefficients (alfa, beta x and beta y), is the best one to estimate the trajectory of football players, since it produces the lowest error (MED= 0.76 m). On the other hand, one can also see that the network that produced the worst results (MED= 1.17 m) was N1 (in red), which corresponds to a NAR that uses as inputs only x and y coordinates.

Finally, we should also point out that, when comparing these results with the ones obtained through the FC approach, which for a sampling frequency of 1 Hz produced a mean error of 1.26 m (Figure 5.2), one can conclude that the RNNs can still produce better results, since the worst result obtained for this method was still slightly lower than the mean error obtained with the FC approach.

5.4 - Summary

In this chapter, an extensive study regarding the prediction of football players' coordinates during a football game was presented.

The FC was first used for estimating the athletes' trajectories during a match, having confirmed its efficacy in works under this scope. Additionally, it was also possible to identify that the sampling frequency highly influences the final error, being that higher frequencies lead to lower errors and vice-versa. This first study was performed to find out the frequency that could lead to better results during the RNNs implementation. As such, taking into consideration that the stability coefficient (Beta) is very sensitive to samples that are consecutively equal, we chose the sampling frequency of 1 Hz. After having settled the sampling frequency, the predictability and stability coefficients were calculated adopting the FC approach, previously described in Section 2.2. Then, the x and y coordinates as well as these two coefficients were used as feature vector for the implementation of several RNNs, in order to discover which type of RNNs could be more viable for this type of work.

The results demonstrated that in order to have estimations with lower errors, the best RNN to be implemented is the one that uses as internal inputs the polar coordinates and as external inputs the fractional and stability coefficients. Indeed, this was already expected since polar coordinates can represent more accurately the movement of the athlete over time, since it correlates each players' position on plane with a reference point, calculating that distance as well as the respective angle. Additionally, with this system, the velocity vector can be directly related with the amplitude of the polar coordinates [172]. By that, the RNNs are capable of getting more information about the athletes' motion, which, by consequence leads to better results. Due to the better performances demonstrated by the technique of polar coordinates, several works have already implemented it in the context of sports [173], [174], for example, in Penas and Anguera [174] the polar coordinates technique allowed the authors to produce a conceptual map of the motor relations among the members of a soccer team by comparing the different associations among players.

Finally, it was also possible to conclude that, in general, the RNNs perform better than the FC approach.

Chapter 6

Physiological Data

This chapter presents the methodology and results concerning the physiological data. As such, we will first explain the process of data collection and signal processing. Then, we will use the processed EMG signals for the classification of different athletes' actions. Finally, we will evaluate the viability of *TraXports V2* by performing muscle fatigue detection.

6.1 – Data Collection

In this work, EMG data was provided by *TraXports V2* (see Section 4.2.2). This is a wearable device developed under the scope of this thesis with the aim of further developing *TraXports*, by allowing not only to measure the pose and the orientation of the player, to perform gait analysis and, most importantly, to measure muscle activity. However, since this is still an ongoing work, at the moment it is still in a primary phase, being only able to give as outputs the raw EMG signal and a processed EMG signal concerning the right thigh of the subject. It is noteworthy that since we are using the Myoware Muscle Sensor, which outputs analogue values, and, therefore, cannot express the negative voltages of the EMG signal, the sensor is programmed to output the EMG signal centered on the $V_+/2$ (3.3V/2) voltage, meaning that, as Figure 6.1 illustrates, the raw EMG has its baseline located at 1.57V. In here, it is possible to visualize two raw concatenated EMG signals regarding the action walking for the participant P1.

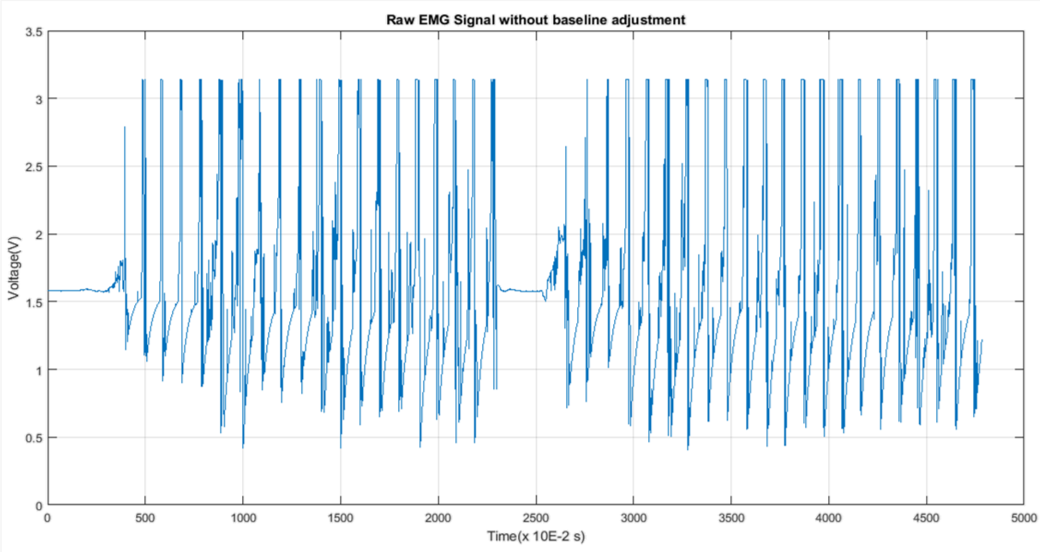


Figure 6.1 – Example of a raw EMG signal for the action walking without baseline adjustment.

As such, it was first necessary to adjust the baseline of the EMG signal to 0 Volts. For that aim, a Savitzky-Golay smoothing filter of polynomial order 3 was implemented, being the result demonstrated in Figure 6.2.

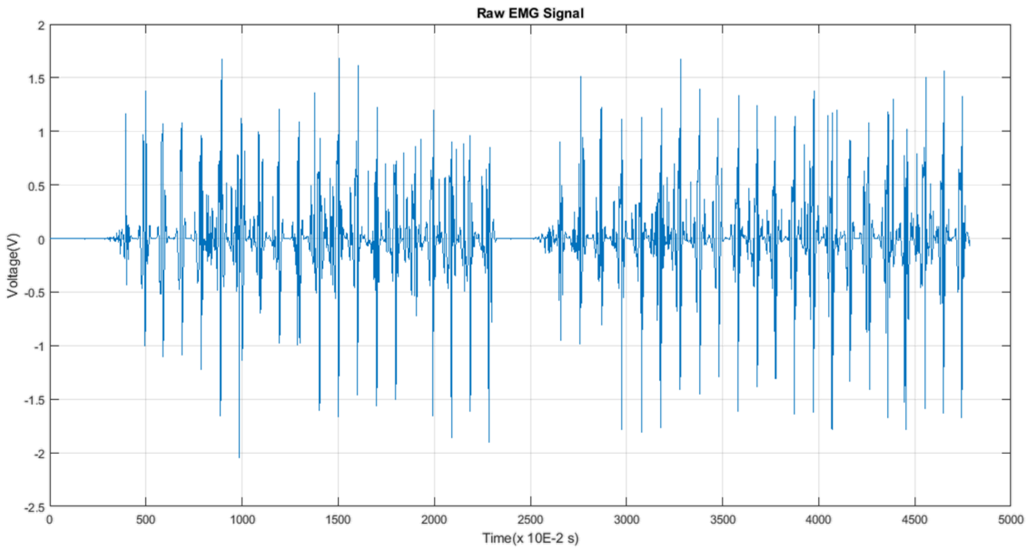


Figure 6.2 – Example of a raw EMG signal for the action walking after baseline adjustment.

It is noteworthy that after performing the baseline adjustment, the signal becomes immediately “cleaner”, being possible to easily distinguish different muscle contractions, which, in this case, happen each time the participant walks. However, we can also notice that the muscle contractions happen very close from each other, being this due to the fact that the signal is being sampled at a frequency of 100 Hz, which for EMG signals, is very low (see Section 4.2.2.1).

6.1.1 – Sample Description

The acquisition was performed with a sample of 3 participants: one female subject and 2 male subjects (age: 31.33 ± 1.25 year old; height: 170 ± 4.08 cm; weight: 72.67 ± 8.99 kg). All the participants are personal trainers, being considered healthy, highly physically active and with no medical history on nerve damage or motor impairment on any of the lower limbs.

6.2 - Pre-processing

According to the literature [111], [122], [175], EMG signals pre-processing comprises three main steps: i) Filtering, ii) Rectification and iii) Smoothing. The last two steps consist in finding the linear envelope of the EMG signal and can be replaced by using the RMS approach or even the Moving Average. However, in this thesis, we will use the first approach (Rectification and Smoothing). A more detailed explanation regarding the proposed approach for EMG pre-processing is presented below:

6.2.1 - Filtering

In this phase, the aim is to remove unwanted noise from the original signal. Note that filtering is not always necessary. For example, if the device of EMG acquisition already provides clear signals, there is no need of performing additional filtering. However, in this case, since we chose to use the raw EMG output from the Myoware Muscle Sensor, there is the need of performing a pre-filtering. According to the literature [175], usually, a band pass filter is first applied to the raw signal, allowing to remove both low and high frequencies from the signal. The low frequency cut-off of the band pass filter removes the baseline drift that is usually associated with movement, breathing, etc., and the Direct Current offset, being the typical values for the low frequency cut-off 5 to 20 Hz. On the other hand, the high frequency cut-off of the band pass filter removes high frequency noise and prevents aliasing from occurring in the sampled signal, being the values typically between 200 Hz and 1 kHz.

In spite of this, taking into consideration the sampling frequency allowed by this device (Fs=100 Hz) we applied a band pass filter with a low frequency cut-off of 4 Hz and a high frequency cut-off of 50 Hz. These frequencies were chosen after observing the signals behavior with different frequencies. In Figure 6.3, it is possible to visualize the signal before (raw EMG signal) and after being processed (filtered EMG signal).

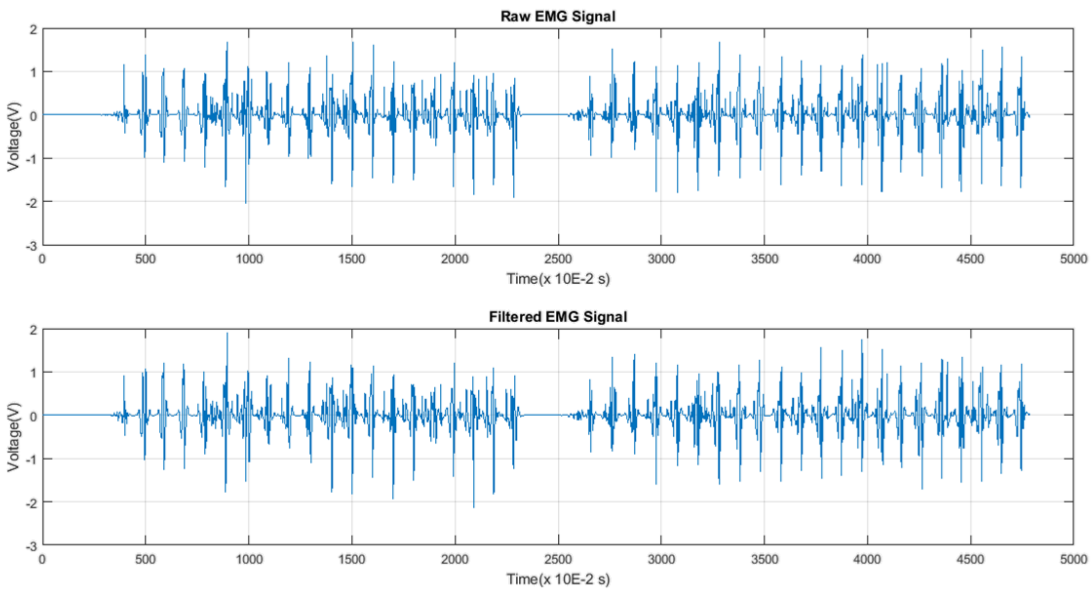


Figure 6.3 - Comparison between the raw and filtered EMG signal.

6.2.1.1 - Linear Envelope of the EMG signal

Afterwards, we proceeded to the acquisition of the linear envelope of the EMG signal, which comprises two main phases, Rectification and Smoothing. These will be explained in more detail hereafter.

a. Full-Wave Rectification

The main objective of rectification is to reorganize the signal in a way that standard amplitude parameters, such as mean, peak/maximum value and area can be calculated. For that aim, this operation converts all the negative values into positive ones [111]. Note that, contrary to filtering and smoothing, this procedure does not affect the signal noises, which explains the need of the following step. In the following figure (Figure 6.4) it is possible to visualize the EMG signal after rectification.

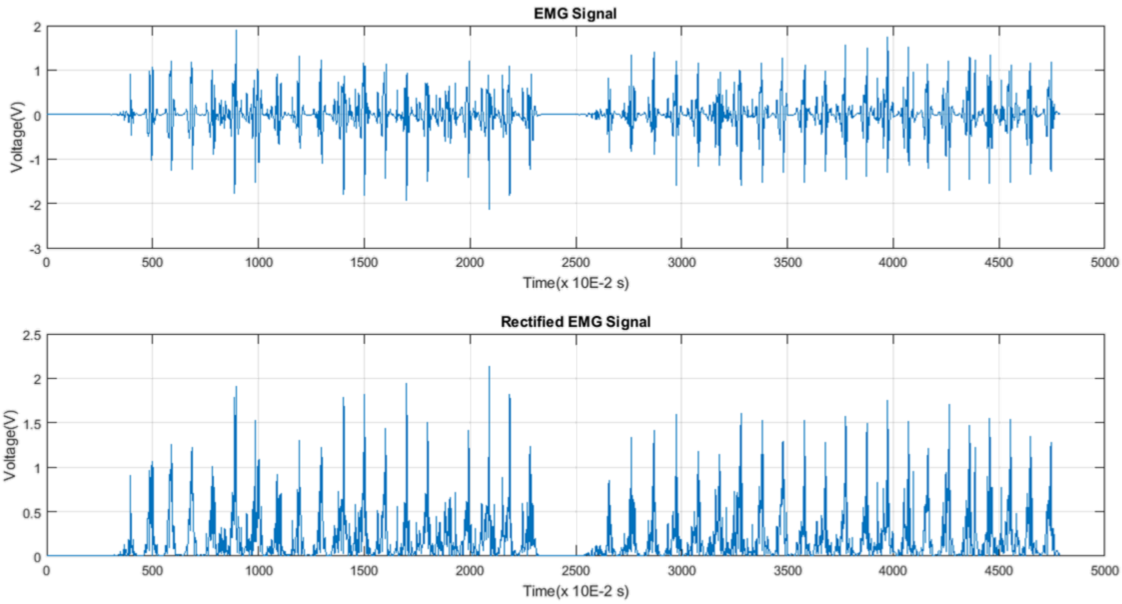


Figure 6.4 - Comparison between the filtered and rectified EMG signal.

b. Smoothing

This operation allows the creation of a linear envelope in the signal, leaving only the center part of it. Like filtering, smoothing also takes out outliers, i.e., data that is considered noise. However, while filtering takes into account the muscle activation range, the smoothing considers the signal itself [176]. In order to perform this operation we will implement a low pass filter, namely a Butterworth Filter, since, accordingly to the literature, this is one of the most used digital filters for decreasing the ratio between the signal and the noise [177], and, thus, to smooth the signal. After analysing the behaviour of the signal according to different low pass filters a 5 Hz frequency cut-off and 4 order polynomial was implemented. Figure 6.5 illustrates the EMG signal after smoothing.

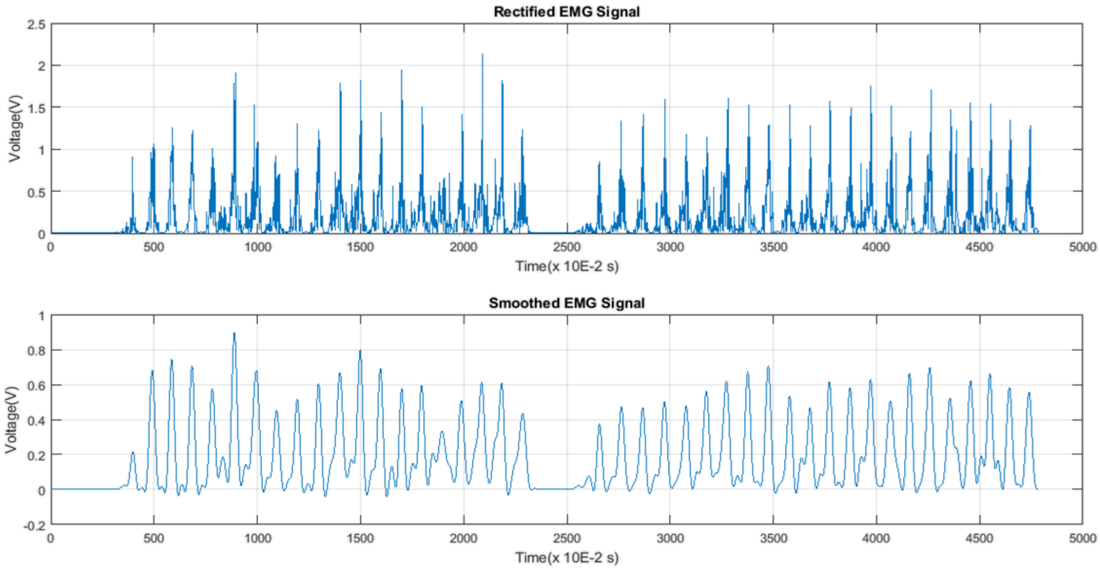


Figure 6.5 - Comparison between the rectified and smoothed EMG signal.

6.3 - Actions' Classification

In the context of a football match, having a system capable of predicting players' actions during the game may be very useful, since, by that, the coach can understand how each opposite player acts during specific situations and teach their players how to behave according to each situation, leading to better chances of winning. Thus, the following subsections present a detailed description regarding the classification of different activities performed by athletes, in which a comprehensive comparison between different classification's approaches is also illustrated.

6.3.2 - Procedure Description

Each participant performed 4 different activities, repeatedly, during approximately 20 seconds each, except in the fourth activity where the participants only performed the respective action (ball kicking) once at each time. This procedure was repeated twice. It is noteworthy that the activities chosen are actions usually dominant in football matches, namely: walk, run, jump and ball kicking, being each of them classified with a numeric label, as represented in Table 6.1. For the jump activity, the participants were asked to perform several types of jumps, including lateral, backward and forward jumps. All these activities highlight the movement of the lower limbs, in which the EMG signal corresponds to the muscle activity of the right thigh of each subject. As such, in the end of data collection, for each participant four pairs of different signals were acquired, i.e., two signals for each activity, meaning that each signal corresponded to the repeatedly execution of one activity during 20 seconds, except from the last action. Figures 6.6, 6.7, 6.8 and 6.9 show different participants executing the activities under study.

Table 6.1 - Labels for the activities performed by each subject.

LABEL	1	2	3	4
ACTIVITY	Walk	Run	Jump	Ball Kicking

Figure 6.6 – Example of one participant executing the first action (walking).

Figure 6.7 – Example of one participant executing the second action (running).

Figure 6.8 – Example of one participant executing the third action (jumping).

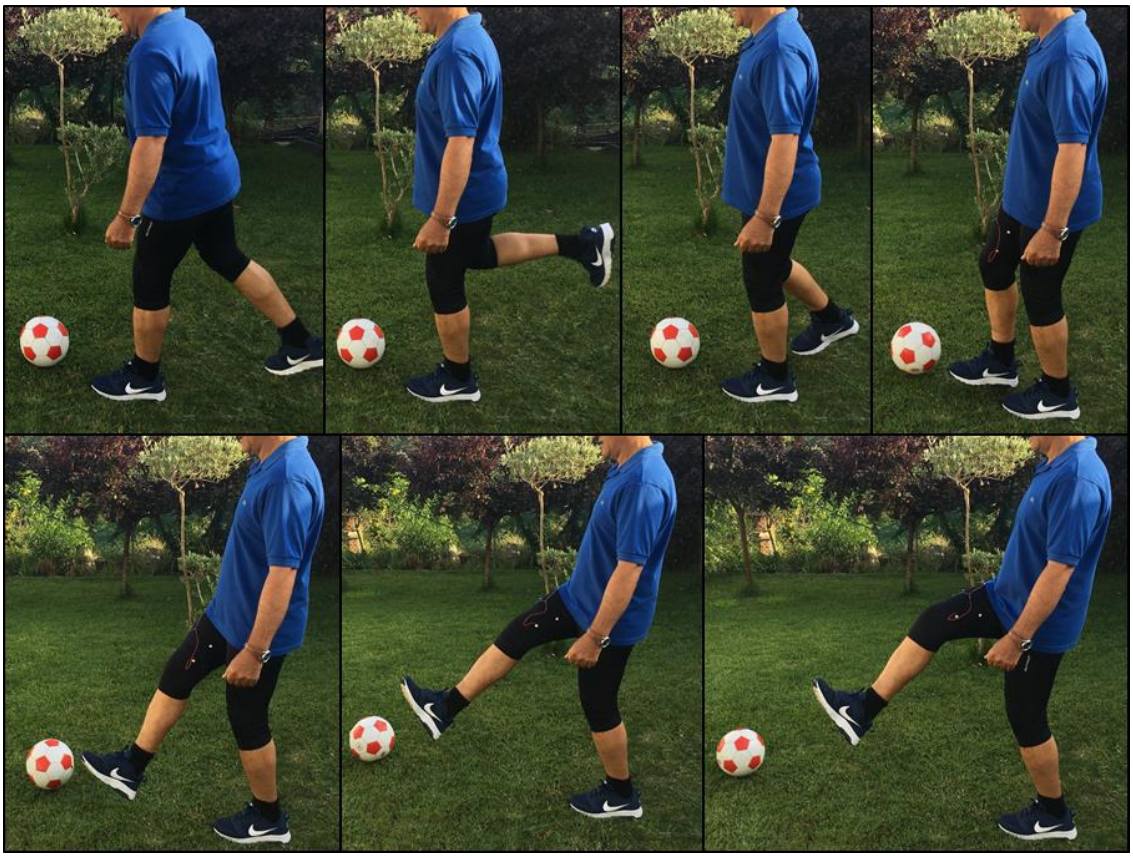


Figure 6.9 – Example of one participant executing the fourth action (ball kicking).

As it is possible to visualize in the various figures previously presented, the data collection took place in several different locations, including inside and outside facilities, demonstrating the viability of *TraXports V2* as wearable device. In general, during data collection all the participants felt comfortable and did not express any discomfort regarding the electronic components.

6.3.3 - Feature Extraction

After preparing the EMG signal by following the several processing operations previously described, the next step was to perform feature extraction to form the feature vector that was going to be used to classify the activities.

Thus, feature selection for EMG signal was made according to the state of the art presented in the subsection 3.1.3.2.

The following features, already formulated in section 3.1.3.3, were then defined:

1. The EMG signal:
    a. Raw EMG signal.
    b. RMS, defined as the square root of the mean over time of the square of the vertical distance of the graph from the rest state, related to the constant force and non-fatiguing contraction of the muscle.
    c. Number of peaks in the processed EMG signal, with a condition of minimum peak height of 0.2 V. This reference was chosen after visually inspecting the signals referent to different activities.

d. Integrated Absolute Value, defined as the area under the curve of the rectified EMG signal, i.e., the mathematical integral of the absolute value of the raw EMG signal since the beginning of the activity.

2. Autoregressive feature, described as a linear combination of previous samples plus an error term that is independent of past samples.
3. Waveform Length, defined as the distance over which the wave's shape repeats.
4. Power Spectrum (Fourier Transform):

   a. MF, defined as the average frequency value that is computed as the sum of the product of the EMG power spectrum and frequency, divided by a total sum of spectrum intensity.

   b. MEDF, is the frequency value at which the EMG power spectrum is divided into two regions with an equal integrated power.

Note that the feature vector was calculated after signal processing and by following an accumulative approach, i.e., at each iteration the feature was calculated by taking into consideration all the previous signal values. Note that, in this work, each signal corresponds to the execution of one of the activities during approximately 20 seconds. Thus, during this step, the extraction of features was performed to each signal separately, meaning that in the end we produced four different feature vectors (one for each activity) for each participant. Table 6.2 summarizes the choice of features for physiological data.

Table 6.2 - Summary of features for physiological data.

DOMAIN	FEATURE	
Time	EMG Signal	Raw EMG signal
		RMS
		Integrated Absolute Value
		Number of Signal Peaks
Frequency	Power Spectrum (Fourier Transform)	MF
		MEDF
Time-Frequency	Autoregressive Features	
	Waveform Length	

According to the literature, feature extraction is of high importance since it can greatly affect the classification outcome [178]. As such, it is important to test different sets of features and to establish an optimized set. Thus, we first started to test the data with several combinations of features, having reached to the conclusion that the set of features presented in Table 6.2 is the best choice. This choice was made by performing a visual inspection to a larger set of features and by understanding which were the ones that differed more from each other. In the following figures, two examples of features are illustrated, in which it is possible to visualize the behaviour of the Raw EMG (Figure 6.10) and a set of two different features, namely the Raw EMG and the RMS (Figure 6.11), according to the four different activities performed, being visible in the second figure that features of the same class cluster together nicely. This heterogeneity in feature selection (distinct clusters for different activities) will allow us to identify more accurately the activity being performed.

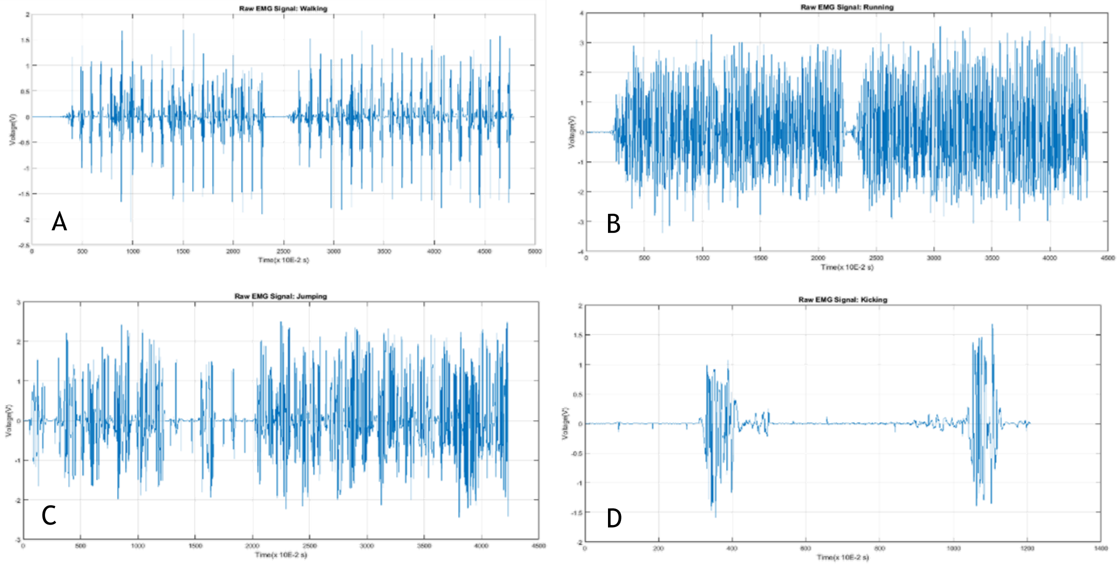


Figure 6.10 – Raw EMG for different activities. **A)** Walking; **B)** Running; **C)** Jumping; **D)** Kicking.

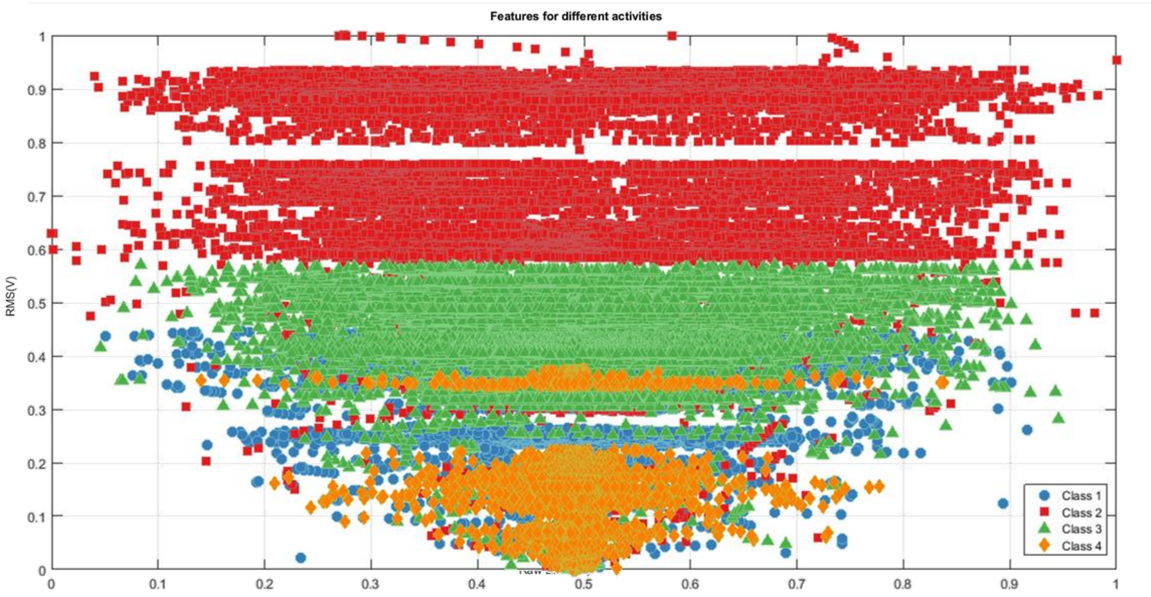


Figure 6.11 – Raw EMG versus RMS for different activities.

Once the features for all the activities were correctly stored in the data array, it was noticeable that there were some features, namely the IEMG, the MF and the MEDF, which were on a relatively larger scale than the others. Since this directly affects the capability of the algorithm to accurately classifiy each, we performed a normalization of all of the features' values to confine the data between the range [0,1]. Figure 6.12 demonstrates an example of feature normalization.

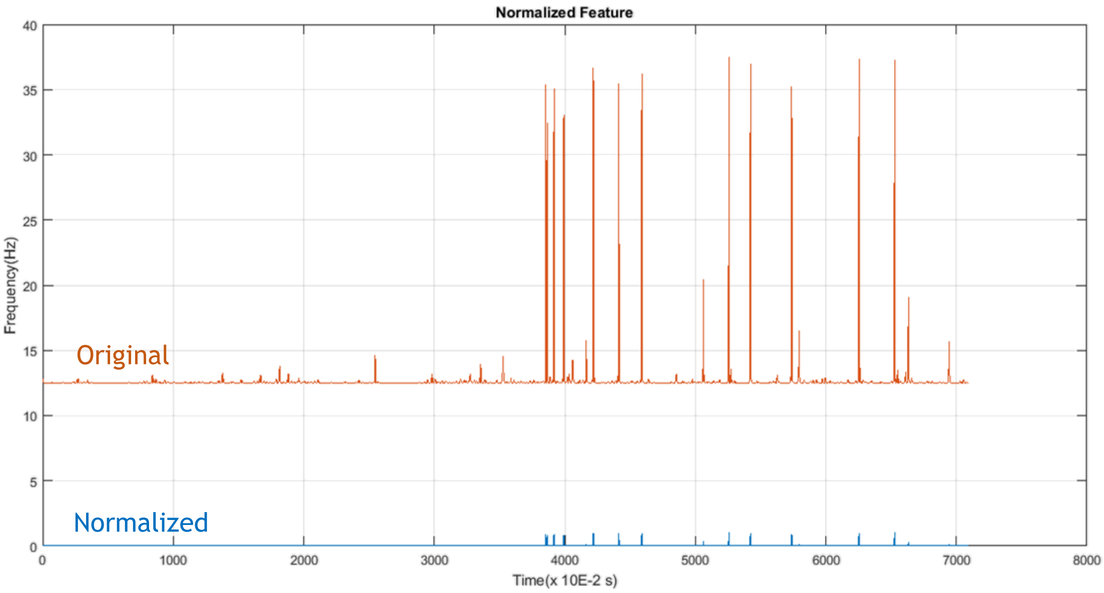


Figure 6.12 - Normalization of the feature MF (red: original feature, blue: normalized feature).

6.3.4 – Classification and Evaluation

This subsection presents the results regarding actions' classification by testing several classification algorithms. To evaluate the performance of the proposed classifiers, we will first use the concepts of precision and sensitivity (recall). The first one is related with reproducibility and repeatability, measuring the fraction classified as positive that is truly positive.

$$Precision = \frac{TP}{TP + FP} \tag{6.1}$$

On the other hand, recall measures the fraction of positive examples that are correctly labeled [179].

$$Recall = \frac{TP}{TP + FN} \tag{6.2}$$

Where, TP comes from True Positive, FP from False Positive, FN from False Negative and FP from False Positive.

Additionally, to further evaluate the performance of the classifiers, we will also use the confusion matrix, which allows to clearly visualize the performance of the algorithm for each activity (class). Figure 6.13 illustrates a confusion matrix.

		PREDICTED	
		+	-
REAL	+	**TP**	**FN**
	-	**FP**	**TN**

Figure 6.13 - Example of a confusion matrix.

6.3.4.1 – Simple Classifiers: NB, ANN, SVM and KNN

In this work, in order to decide about the best choice of classifiers for this type of data and to optimize the final classification architecture, we tested several algorithms under different conditions. Thus, during the first phase of testing, several simple classifiers were implemented, namely NB, ANN, SVM and KNN, firstly, under a k-folds cross validation approach and, afterwards, under the leave-one-out approach.

The ANN was implemented with 30 hidden layers and SVM was implemented according to the strategy one versus all. For KNN, 3 neighbours were used accordingly to the Euclidean Distance.

Hence, the feature vector previously defined on Table 6.2 was used to classify the different activities under study, being the results demonstrated on the following tables. The classification algorithms were implemented by using the Matlab toolbox available on [180].

6.3.4.1.1 – Cross Validation

Firstly, the method of k-folds cross validation was implemented, in which the dataset, formed by the features of all 3 participants, is divided in k segments. By that, the algorithm uses k-1 segments for the training phase and 1 segment for testing. Three types of cross validation were implemented: 5, 10 and 30 folds. The results of precision (Prec) and recall (Rec) are presented in terms of percentage in Table 6.3. The column "Total" represents the mean of precision and recall for each classifier used.

Table 6.3 - Overall Precision and Recall obtained for activities' classification for NB, ANN, SVM and KNN according to different k-folds.

		30-FOLDS	10-FOLDS	5-FOLDS	TOTAL
NB	Prec	62.46	62.59	62.68	62.58
	Rec	64.63	64.72	64.80	64.72
ANN	Prec	53.12	54.98	55.17	54.42
	Rec	42.24	47.53	48.15	45.97
SVM	Prec	61.79	71.75	64.87	66.14
	Rec	61.98	71.85	63.73	65.85
KNN	Prec	68.46	68.39	68.24	68.36
	Rec	68.45	68.35	68.22	68.34

Regarding the performance of the classifiers, it is possible to verify that the best classifier is KNN, with a total precision of 68.36% and a total recall of 68.34%. On the other hand, the classifier that presented a lower performance was the ANN, with a total precision of 54.42% and a total recall of 45.97%. In fact, this was not expected, since according to the literature (See section 3.1.3.3) the ANNs are considered one of the best classifiers for EMG data. Thus, there are two possible explanations for these results: overfitting or underfitting. The first one happens when the model describes error or noise instead of the underlying relationship between data and usually happens when the model is excessively complex, i.e., has more parameters than number of observations, which makes it to overreact to minor fluctuations in the training data, and, thus, leading to worst performances. However, since in this case, the ANNs were set up to have a validation phase, exactly to prevent overfitting, the results obtained might not be due to overfitting. Instead, the most possible explanation is the occurrence of underfitting. This happens when the statistical model is not capable of capturing the underlying trend of data. In fact, by analysing our data one can infer that the most plausible reason for these low results is that underfitting occurred, not only because we fitted a linear-model to non-linear data but also because of the reduced number of features used. Thus, a good strategy to overcome this problem would be to add more hidden layers, increase the number of neurons in the hidden layers or add more features to the input space.

6.3.4.1.2 – Leave-one-person-out

Despite the performance obtained with the k-folds cross validation approach, one should point out that this does not represent a real situation, in which we have the algorithm trained for a specific set of people and we want to classify the activities performed by a new person that the algorithm does not have knowledge about. As such, in order to represent a real classification and to have more suitable data for posterior comparison, we also tested EMG dataset by using the leave-one-out approach. In here, the algorithm uses always two participants to train the classifier and the other to test it. For example, in the classification associated to the first participant, the data used for training were only from the other two participants (P2 and P3). The results are demonstrated on Table 6.4.

Table 6.4 - Overall Precision and Recall obtained for activities' classification for NB, ANN, SVM and KNN for the different participants.

		P1	P2	P3	TOTAL
NB	Prec	55.88	44.17	70.45	56.83
	Rec	54.85	49.60	74.32	59.59
ANN	Prec	50.32	34.34	64.39	49.68
	Rec	47.73	31.60	69.68	49.67
SVM	Prec	58.59	44.39	65.08	55.21
	Rec	55.63	49.25	68.98	57.60
KNN	Prec	61.24	39.08	68.85	56.39
	Rec	59.38	41.03	70.08	56.83
SUBJECTS' TOTAL	**Prec**	56.51	40.50	67.19	
	Rec	54.40	42.87	70.77	

As it was expected, the overall precision and recall of the different classifiers decreased, being this mainly due to the fact that the classifier is using data from a new person. Indeed, although, in general, the features do not vary that much for the same person, when we compare between different people it is possible to observe greater differences between the same features, being these highly dependent on several factors, such as, the gender of the person, as well as its age and physical activity.

We can also observe that NB, SVM and KNN produced similar classification results, all with precision above 55%. On the other hand, similarly to the results shown in Table 6.3, the ANN was the classifier that produced worst results. Moreover, it is also possible to conclude that participant 3 (P3) was the one that produced better results, with a total precision of 67.19% and a total recall of 70.77%, and participant 2 (P2) was the one who produced the worst results, with a total precision of 40.50% and a total recall of 42.87%. These results highlight the influence of different people with different characteristics on the classification, since, for example, in this case, P2 and P3 are both male persons, however, P2 is an older participant, when compared to P3, meaning that he is probably not as fit as the other participant.

Finally, one should also point out that, when implementing these type of classification algorithms, it is very important to also analyse the resultant confusion matrix, to understand which are the classes that are being better classified and vice-versa. For example, for a final precision of 50%, this could mean that all the classes are being half well-classified, or that two of the four classes are being 100% well-classified and the other two completely misclassified. By that, and taking into consideration that the DBMM approach basically classifies each sample according to the best results produced by each individual classifier, we also built the confusion matrix for each classifier (Figure 6.14), in order to infer about the possibility of having better results with the DBMM approach. Since P2 was the participant which produced worst results, in the following figure we present the confusion matrix for each classifier with P2 data used as test.

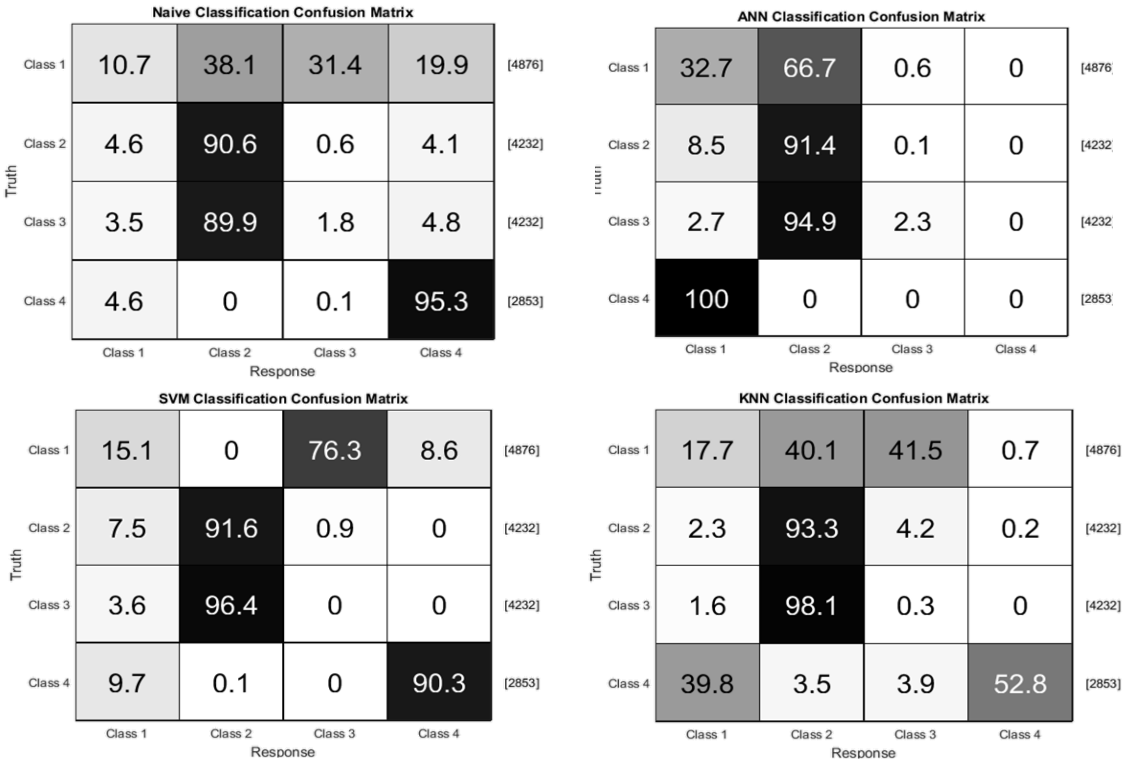


Figure 6.14 - Confusion Matrix for each individual classifier.

Black squares mean that the algorithm attributed the majority of the samples to that class, whereas white squares mean that the algorithm did not attributed much samples to that specific class.

Indeed, as it is possible to observe, there are some classes that are very well-classified with all the algorithms, such as class 2 which refers to the action of running and others that do not have that good results, such as class 3 which refers to the action of jumping.

Overall, it is noticeable that none of these individual classifiers are able to correctly classify all of the actions, being that, in general, they can only succeed in two out of the four classes herein presented. As such, it is possible to infer that by combining all these classifiers, which have good performances under different situations (e.g., different classes), the DBMM approach will be a good architecture for this type of data, and therefore, will produce higher precision and recall results.

6.3.4.2– Ensemble Classifier: DBMM

Taking into consideration that the results obtained with the individual classifiers are still very low, i.e., they are not sufficient for a real-time game situation, it was necessary to implement a more complex classification algorithm, namely one that could pick up the best results of each classifier and combine them with the intent of producing better classification results.

6.3.4.2.1 – Base Classifiers for DBMM fusion

The DBMM approach can be used with several different classifiers, since all the outputs are converted in probabilities (posterior probabilities). As such, considering this and the results obtained in the previous subsection, in this work, we used four different base classifiers, namely, the NB, the ANNs, the SVM and the KNN.

The first classifier considers that the features are independent between each other, given the variable class, being the probability density function for each model of features expressed as follows:

$$P(C_i|A) = \alpha P(C_i) \prod_{j=1}^{m} P(A_j|C_i) \tag{6.3}$$

Where α is the normalization factor and m the number of independent models of features.

Regarding the ANNs implementation, this was performed by using the Matlab Neural Network toolbox. In this case, the activation function of the hidden layer was set to be tangent sigmoid *(transig)* whereas the activation function regarding the output was set as exponential normalized *(softmax)*. Thus, the outputs can be interpret as estimations of the class posterior probability.

Then, for the implementation of the linear kernel in a multi-class SVM, a classifier named error-correcting output codes (ECOC), available on the regular Matlab's toolboxes, which allows to fit multi-class models for SVMs, was implemented. A Gaussian kernel was defined and the SVMs were trained according to the strategy one-vs-one.

Finally, for the KNN classifier, which predicts labels according to the neighbouring true labels, the posterior probability is given by the following equation.

$$p(C_J|X_{new}) = \frac{\sum_{i\epsilon nbd} W(i) 1_{Y(X(i)=j)}}{\sum_{i\epsilon nbd} W(i)} \tag{6.4}$$

Where W= weights for each point. In here, the number of neighbours was set to 50.

Thus, the DBMM approach was implemented with the EMG dataset, being the results presented in Table 6.5.

Table 6.5 - Overall precision and recall of the different activities and participants with the DBMM approach.

		P1	P2	P3	ACTIVITIES' TOTAL
WALK	Prec	60.18	51.78	94.28	68.75
	Rec	83.36	69.32	66.17	72.95
RUN	Prec	82.34	85.58	91.47	86.46
	Rec	92.39	91.82	99.98	94.73
JUMP	Prec	53.25	77.30	71.84	67.46
	Rec	66.02	75.54	79.69	73.75
BALL KICKING	Prec	91.23	68.20	86.56	82.00
	Rec	22.85	41.68	94.86	53.13
SUBJECTS' TOTAL	Prec	71.72	70.72	86.04	**76.16**
	Rec	66.16	69.59	85.17	**73.64**

The results obtained show a clear increase in the classification performance. Similarly to the classification results presented in Table 6.4, participant 3 was the one who yielded a higher overall classification performance, with a total precision of 86.04% and a total recall of 85.15%. Regarding the other two participants, these presented similar classification performances. Additionally, it is also possible to conclude that the running activity was the one which demonstrated better classification results, with an overall precision of 86.46% and an overall recall of 94.73%. On the other hand, the ball kicking activity was the one that produced worst classification results, having an overall precision of 82% and an overall recall of 53.13%. Concerning this, although the value for precision is relatively high, the value of recall is very low, meaning that only 53.13% of the data that was referent to the ball kicking activity was, indeed, classified as ball kicking. This can be explained by the fact that the classes were not balanced, *i.e.*, the first 3 classes were more populated than the last one, being this decision made because of the existent similarity between the signal walking and the more populated signal kicking. Basically, since the sensor only detects the electrical activity of the muscle and, therefore, is activated by muscular contractions, the action of giving one step becomes quite similar to the action of kicking the ball once, being these represented by a peak in the EMG signal. Thus, by having a more populated Class 4, this starts to become very similar to the data that represents Class 1, leading to underperforming classification results. On the other hand, if we choose to have a less populated Class 4, with only two concatenated EMG signals, the algorithm does not possess sufficient amount of information regarding this class in order to produce good classification results. As such, during the choice of participants’ data, an equilibrium between the amount of data present in Class 4 and the similarity between Class 1 and 4 was made. To visualize the overall performance of the algorithm, the confusion matrix regarding the classification results are presented in Figure 6.15.

REAL \ PREDICTED	WALK	RUN	JUMP	KICK
WALK	**72.95**	2.24	17.77	7.04
RUN	1.16	**94.73**	3.43	0.67
JUMP	12.23	9.61	**73.75**	4.40
KICK	27.85	3.02	16.01	**53.13**

Figure 6.15 - Confusion matrix for the classifier DBMM and EMG data.

In here, it is possible to visualize more clearly the performance of the proposed approach and to better understand why there were some classes that did not perform that well. Indeed, by analysing the results obtained in the confusion matrix, it is possible to understand the reason behind the low TPs percentage value for the Kick activity (53.13%). Firstly, one can conclude that this is a class that is not very well identified since the values for the FNs are also very low. One of the reasons for this is the lack of data regarding this class for DBMM implementation. Additionally, it is also possible to conclude that the Kick class was often mistaken with two other classes, mainly the Walk class, in which 27.85% of data regarding the Kick class was classified as Walk. Indeed, this was already expected, due to the abovementioned reasons.

Regarding the other classes, it is also possible to understand why the Walk class is commonly mistaken with the Jump class, and the other way around as well. This can be explained by the existent differences between participants, meaning that, for one participant that is more muscled, the EMG signal becomes more well-defined, whereas for participants which are less muscled, the sensor can sometimes fail to detect the muscle contraction. Additionally, we also have to take into consideration the sampling frequency herein used (Fs=100 Hz), which is much lower than the advised sampling frequency for EMG data (Fs=1000 Hz) [111]. This means that, for example, in the case of the Jump class, since the activity is more intense and quicker, the several muscle contractions may not be completely detected. Instead, only half of the muscle contractions are detected (by visual inspection), leading the method to misclassify it over the Walk class, in which the muscle contractions appear successively one after the other.

It is noteworthy that, during the classification phase, the DBMM always considers all the past classification results in its decision, meaning that its performance for a specific class increases over time, leading to a convergence over time towards the correct class. As such, in order to better visualize this convergence, Figure 6.16 is presented.

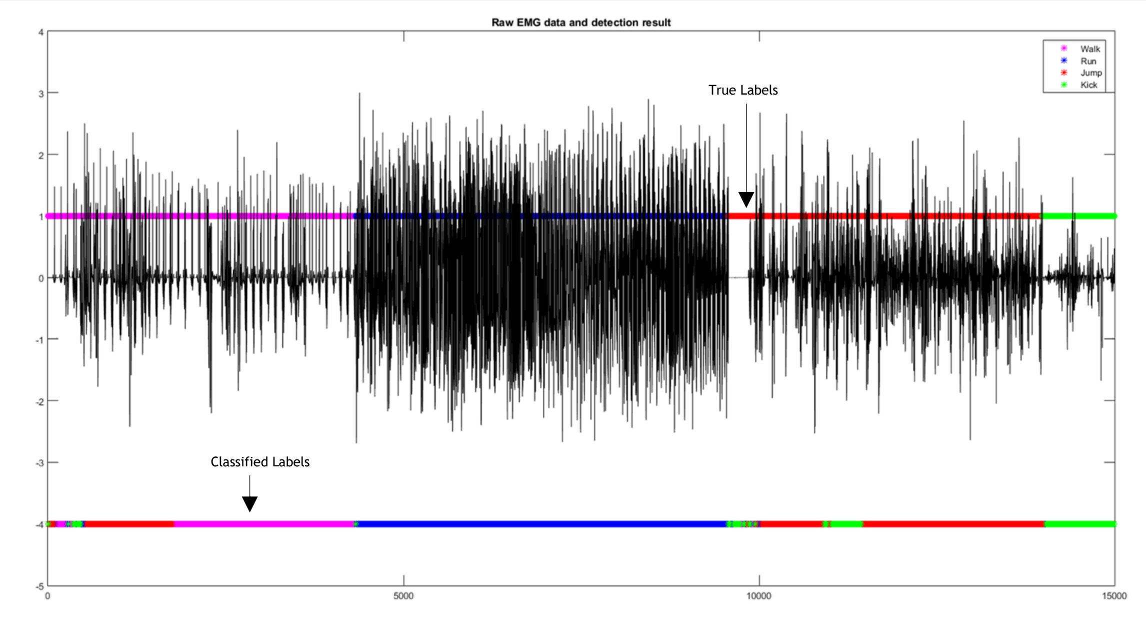


Figure 6.16 - Comparison between the true and the classified labels for each activity.

Note that in this figure the raw signal referent to the four activities successively performed is represented at black, being noticeable four sets of different signals in the figure. Moreover, each activity is represented by one colour: violet for Walk, blue for Run, red for Jump and green for Kick, being also represented the true labels for each activity in the top line and the classified labels for each sample in the bottom line.

Finally, as it is shown in Figures 6.17 and 6.18, it is possible to detect an overall increase in the performance of activities' classification, being that all the participants had their activities better classified when using the DBMM approach, opposed to the individual classifiers. The participant which benefited more from this approach was P2, having increased 30% in its overall precision and 27% in its overall recall. On the other hand, P1 was the participant who had its results less increased, having increased only 15% in its overall precision and 12% in its overall recall. Once again, this highlights the influence of having people that differ very much from each other.

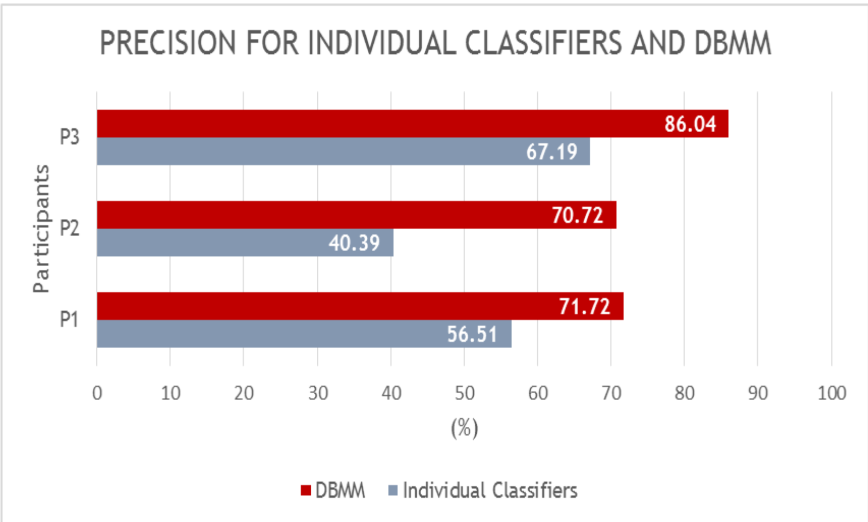


Figure 6.17 - Comparison of the precision obtained with simple classifiers and DBMM.

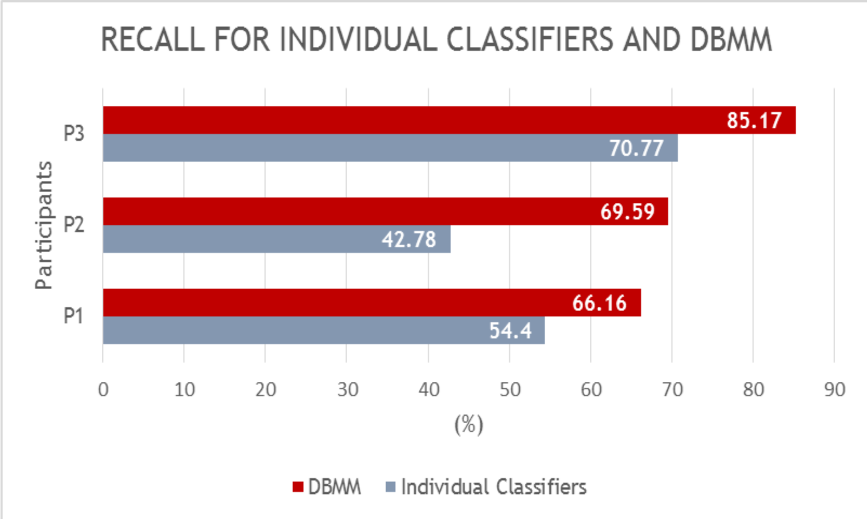


Figure 6.18 - Comparison of the recall obtained with simple classifiers and DBMM.

Indeed, the DBMM approach produces an overall precision of 76.16% and an overall recall of 74.64%, which is by far the best result obtained for this EMG data. Although these results seem apparently low, these are concordant with the results encountered in the literature for similar works [118]-[120]. Indeed, there are not much literature regarding actions' classification by using only EMG signals from the lower limbs, since besides being a signal that does not provide that much significant information for different actions, it is also a place where the muscles are more hidden, and therefore, it is more difficult to obtain signals that in fact do represent the action. As such, considering that in this work it was only used the EMG data for actions' classification, one can conclude that this in indeed a good result, proving the efficiency of the DBMM approach.

6.4 - Fatigue Detection

6.4.1 - Procedure Description

In order to detect muscle fatigue, the EMG signals were collected from each participant while they were walking in two different times, once before the execution of an intensive physical activity and another time after the physical activity, which, in this case, was a CrossFit class. Each data collection had an approximate duration of 20 seconds.

6.4.2 – Feature Extraction

According to the findings presented in subsection 3.1.3.2.3, the EMG signal displays two different characteristics during muscle fatigue. Firstly, there is a change in the amplitude of the EMG signal and, secondly, there is a shift of the EMG power frequency spectrum, i.e., the frequency decreases. As such, in order to detect muscle fatigue we will use three features, being these detailed in Table 6.6.

Table 6.6 - Summary of the features used for fatigue detection.

DOMAIN	FEATURE	
Time	EMG Signal	Integrated Absolute Value
Frequency	Power Spectrum (Fourier Transform)	MF
		MEDF

6.4.3 – Fatigue Detection

In this subsection, several characteristic features of fatigue will be calculated in order to further validate the utility of the wearable device *TraXports V2*.

The features were calculated taking into consideration the raw EMG signal. The results concerning the IEMG value, the MF and MEDF of the signal before and after the execution of the intense physical activity are presented in Table 6.7.

Table 6.7 - Results obtained for the different features before and after the execution of an intense physical activity.

		BEFORE FATIGUE	AFTER FATIGUE
P1	IEMG (V)	0.31	0.88
	MF (Hz)	15.80	11.71
	MEDF (Hz)	10.90	6.58
P2	IEMG (V)	0.28	1.03
	MF (Hz)	13.79	12.90
	MEDF (Hz)	9.02	6.96
P3	IEMG (V)	0.34	0.59
	MF (Hz)	15.30	12.58
	MEDF (Hz)	10.88	7.92

As it is possible to observe from Table 6.7, all of the features behaved according to the literature encountered, i.e., the amplitude of the signal suffered a small increase whereas the frequency features decreased. Note that the values herein obtain for frequency are lower than the typical values for EMG signals (60-100 Hz), being this explained by the fact that our EMG data was acquired with a sampling frequency 10 times lower than the normally used (100 Hz instead of 1000 Hz). As such, taking that into consideration, one can conclude that these frequency values are still within the expected values. In order to better visualize the existent difference between the EMG features, namely the MF, from a non-fatigued muscle and from a fatigued-muscle, Figures 6.19-21 are presented.

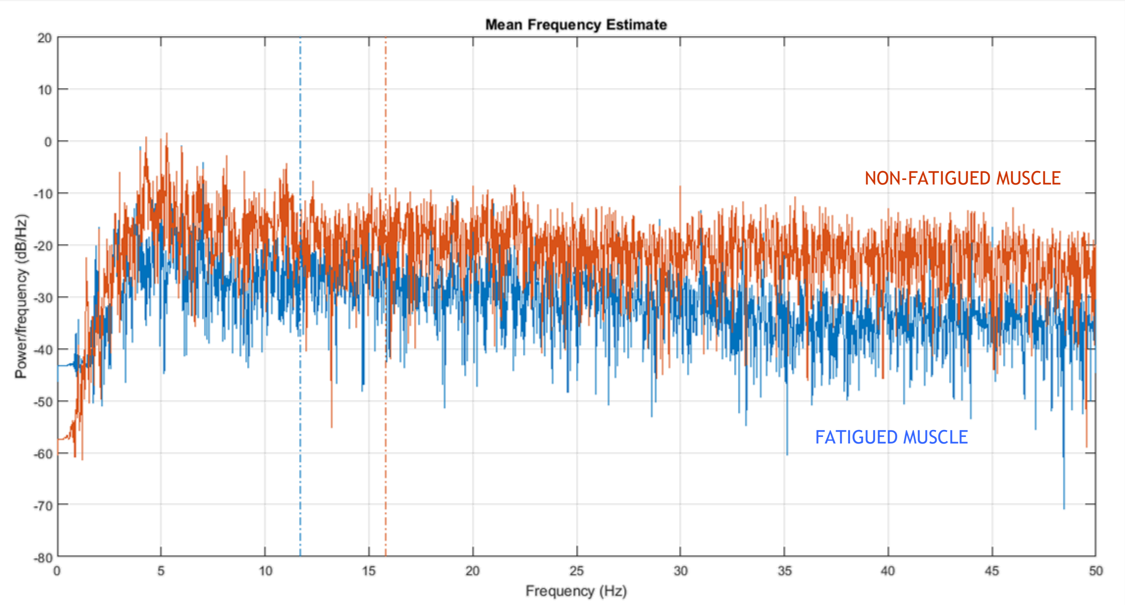


Figure 6.19 – MF for the participant 1 before and after the execution of an intensive physical activity.

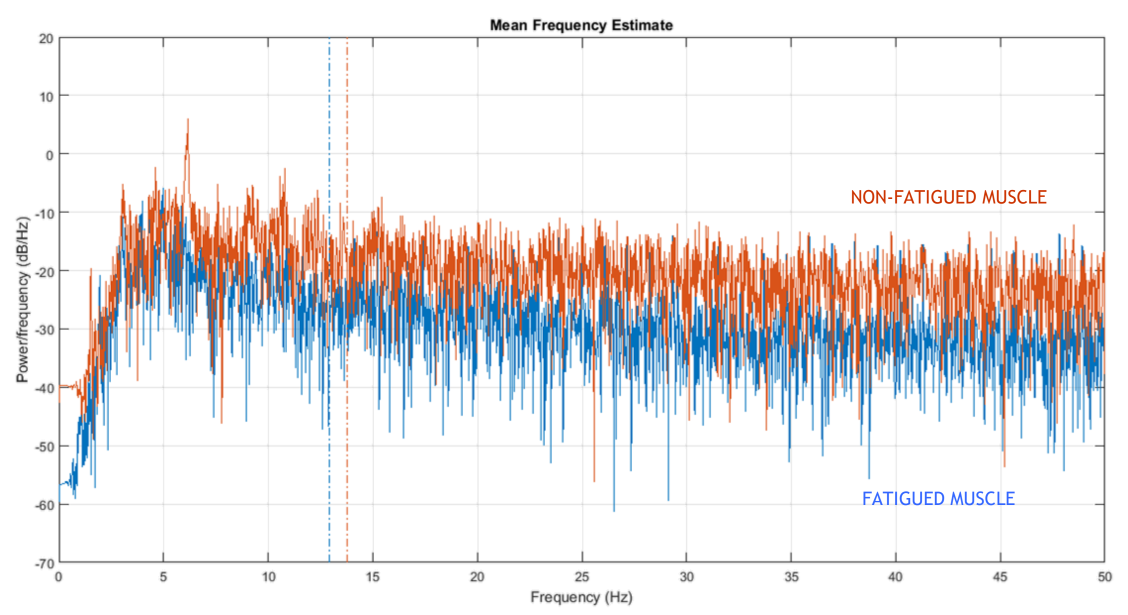


Figure 6.20 – MF for the participant 2 before and after the execution of an intensive physical activity.

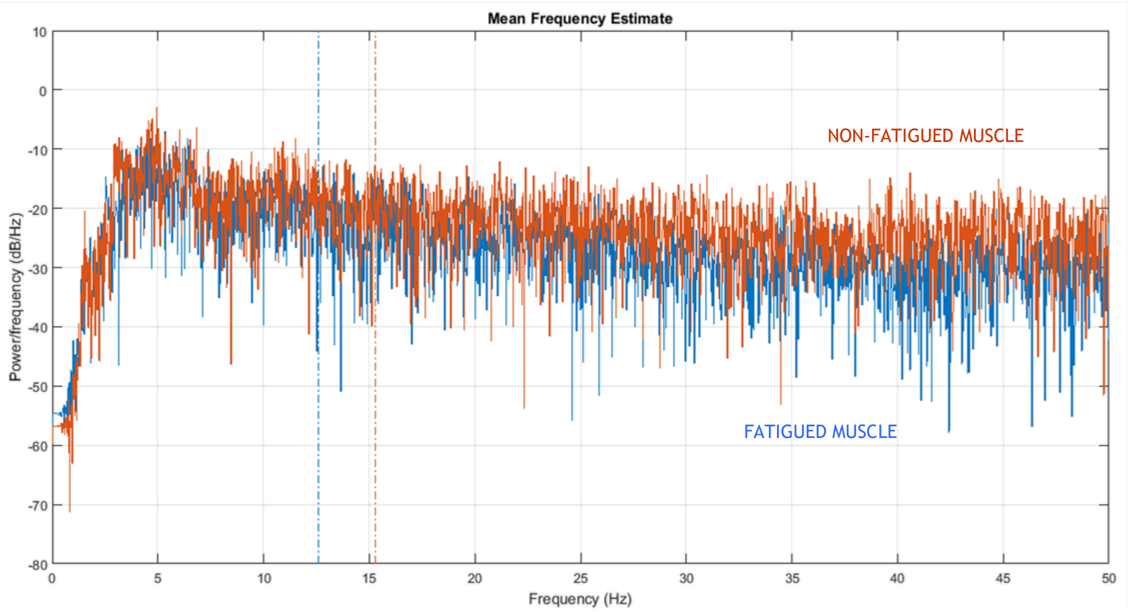


Figure 6.21 – MF for the participant 3 before and after the execution of an intensive physical activity.

Moreover, by analysing the previous figures, it is also possible to notice that the participant P3 was the less affected by the execution of the CrossFit class. Indeed, this corroborates the results obtained in section 6.3, in which, it was demonstrated that this was the participant more physically well prepared.

Finally, with this study, it was possible to validate the utility and success of *TraXports V2* in detecting muscle fatigue.

6.5 - Summary

In this chapter, an extensive study regarding the use of physiological data in the context of a football match was presented, being this divided in two main parts.

Firstly, we used EMG data from the right lower limb of several participants in order to classify four different actions, including walking, running, jumping and ball kicking. In order to compare the performance of the herein proposed architecture, DBMM, we first performed a study with four individual classifiers, namely NB, ANN, SVM and KNN, and compared their performance under different constraints. Since this approach did not represent a situation of real time classification, we then tested the same classifiers under the leave-one-out approach, in which for each participant that was tested, the other remaining participants were used as training. In here, a clear decrease of the overall precision and recall was noticed, with the worst classification performance from the ANNs. Afterwards, the EMG data was used to implement the DBMM approach, being demonstrated a clear increase in the overall classification performance. Additionally, it was possible to visualize performance differences between participants, with participant 3 with the best overall results. This is explained by the fact different people have different characteristics (gender, age, musculature) and although these participants were all personnel trainers, we can still see that there are some that are more physically prepared than others. As such, one should point out that this type of work is highly influenced by the choice of participants. Moreover, it was clearly demonstrated that the activity running was the one which produced better classification results. On the contrary, the activity ball kicking did not performed well, being this mainly explained by the existent similarity between the walk activity and kick activity. In fact, by analysing the overall confusion matrix we can easily visualize that there was an evident confusion between these two activities. Additionally, we studied the capacity of DBMM to classify actions over time, concluding that it mainly fails to classify the first samples of each class, meaning that its performance increases at each iteration (Figure 6.15). Furthermore, when comparing the performance obtained with the first implemented methods (four individual classifiers) and the DBMM approach it was possible to detect an overall increase in the performance of activities' classification, being also noticeable an increase in the overall classification for each participant. Indeed, the implementation of DBMM allowed to produce an overall precision of 76%, being this concordant with the literature.

Finally, in the second part of this chapter a study regarding fatigue detection was performed. In here, the features that were demonstrated to better indicate fatigue in EMG signals, namely the IEMG, the MF and the MEDF were calculated from the raw EMG signals under two different situations. Firstly, while performing the walk activity and before an intense physical activity. Secondly, while performing the walk activity after performing an intense physical activity. The results presented allowed to corroborate the observations described in the literature, since there was a clear increase in the amplitude of the EMG signal and a clear decrease in the frequency properties, after the muscle being fatigued. Additionally, it was also possible to visualize that the participant 3 was the one who was more resistant to fatigue, since there was a lower variation between the features before and after fatigue. This is concordant to the results obtained for actions' classification, which indicated that participant 3 was the most well physically prepared.

Although these were good results, we have to take into consideration the final aim of this thesis that is to contribute to the implementation of a real time framework for football prediction. As such, one should point out that the strategy herein presented, of receiving as input the whole signal and classify it accordingly to the accumulative features' extraction method, would not be suitable for a real time classification, despite the nature of DBMM to convergence to the right class over time. Thus, for implementation in real time, an approach of temporal segmentation should be implemented, i.e., as the signal is received by the architecture, this should be able to detect each muscle contraction, extract the corresponding features to each muscle contraction, and attribute an action to that muscle contraction accordingly to the muscle contractions previously trained.

This work allowed to confirm the viability and utility of *TraXports V2* as a wearable device for the context of a football match.

Chapter 7

Conclusion

In this chapter, first, the overall conclusions for this master thesis are summarized. Then some suggestions regarding future works are presented.

7.1 – Final Remarks

The ability of predicting the final outcome of a football match, decoding the teams' performance, and additionally preventing any potential athletes' injuries, is becoming more and more a theme of subject within the scientific community. Indeed, football is a very competitive sport, in which managers, coaches and even football players try to be always on advantage. This can be done by having better players, better conditions or even by owning better technology. Thus, having a device capable of providing objective answers to coaches regarding the physical and psychological state of each player during a football match is seen by the community as a great asset.

According to the literature, defining an architecture capable of taking into consideration not only non-physiological parameters but also physiological ones is of extreme importance. In fact, the state of the art shows that many authors have been presenting methodologies to predict the football match outcomes, though none explores all the relevant features that influence athletes' performance, being this one of the main explanations for the lack of successful machine learning algorithms for football prediction. In line with this, we concluded that is crucial to take into consideration both kinematic and physiological signals while developing an architecture for football prediction.

As such, with this project it was intended to contribute to the undergoing project ARCANE by developing an architecture for football prediction, capable of estimating both athletes' position and actions as well as to detect fatigue over time. For that aim, three main objectives were defined. Firstly, to estimate trajectories of football players, then, to recognize several actions performed by a football player, including, walking, running, jumping and kicking, and, finally, to detect fatigue. In order to do that, two different types of data were used, kinematic and physiological signals. Regarding the latter, it was necessary to build a new wearable device, termed *TraXports V2*, capable of measuring the electrical activity of the right lower limb muscles. This was adequately validated within a group of 20 people, being concluded that it was indeed a suitable device for the context of a football match.

After establishing the most relevant signals, we then discussed about the most used classifiers and features for each type of signal, concluding that there is a specific set of features and classifiers that produces better results for each type of signal. For example, for kinematic data we concluded that the best classifiers were KNN, SVM, HMMs, BM and DBMM and the best features where the mean, the variance, the Fourier Transform, the Wavelet Transform and the Spectral Entropy, whereas for physiological data we concluded that the best classifiers were the ANNs and the best features were the Fourier Transform, the Wavelet Transform and the Autoregressive Models. Finally, we also discussed about the importance of features and classifiers' selection in the final result, concluding that it highly influences the final result. As such, we reached to the conclusion that it is of extreme importance to have a phase of features and classifiers' optimization within the algorithm's architecture.

Since different signals have different characteristics, one can easily understand the necessity of establishing different approaches for each type of signal. As such, we first used kinematic data for the estimation of athletes' coordinates over time, being implemented two different methodologies, FC and RNNs. The FC was used with the main objective of providing suitable features for the implementation of the RNNs. For that aim, we first defined the more appropriated frequency by analysing the Euclidean distance errors for different frequencies, having concluded that the best frequency was 1 Hz. After having defined the optimized frequency, the feature vector containing the real x and y coordinates was created and used for the implementation of several RNNs, having conclude that the best RNN to be implemented is the one that uses as internal inputs the polar coordinates and as external inputs the fractional and predictability coefficients. The results also demonstrated that overall RNNs perform better than FC.

Finally, an extensive study regarding the use of physiological data in the context of a football match was performed. Concerning actions' classification, we found out that the use of DBMM with EMG data clearly increases the overall classification performance, when compared to the classification results of individual classifiers. Additionally, it was also possible to conclude that this type of work is highly influenced by the choice of participants. Indeed, different people have different characteristics (gender, age, musculature) and although these participants were all personnel trainers, we can still see that there are some that are more physically prepared than others. Furthermore, we concluded that the running activity was the one which produced better results, whereas the ball kicking activity was the one which performed worst. Indeed, taking the consideration that in this work we are only using EMG signals and that at the sampling frequency herein used (Fs=100Hz) it is very difficult to produce adequate signals, one can conclude that the overall precision of 76% and recall of 74% is actually a good result and within the expectations. Nevertheless, this system requires to be further improved in order to be adapted to a football match, since it was noticeable that it had difficulty in correctly classify these simple classes. Thus, when adding more classes, such as, different types of shootings and kicking, this architecture will not be capable of accurately classify all the activities due to their similar nature. Thus, when the wearable device is improved, namely, extract also kinematic data, this architecture will become more suitable for being applied in a real time football match.

On the other hand, the work related with fatigue detection was performed in order to further validate the wearable device *TraXports V2*. Indeed, there was a clear increase in the amplitude of the EMG signal and a clear decrease in the frequency properties, after the muscle being fatigued. Additionally, we can also point out that it was possible to understand that the participant 3 was the one who was more physically well prepared, since not only presented the best classification results but also it was the one who suffered a lower variation between the features before and after fatigue.

In conclusion, we considered that all the proposed objectives were successfully fulfilled, having made a significant contribution to the development of a new architecture for football prediction.

7.2 - Future Works

Although this master thesis has achieved its aims and objectives, we feel that there is still space for improvements. As such, the suggestions regarding possible future work will be presented hereinafter.

Firstly, we felt that the wearable device herein produced, *TraXports V2*, had some limitations, mainly in terms of the frequency of data acquisition. As such, in a posterior work, the sampling frequency should be increased at a value between the range of 500 and 1000 Hz, preferentially 1000 Hz. This can be achieved by using a more powerful Wi-Fi transmitor, for example. Still regarding the improvement of *TraXports V2*, we suggest the implementation of more conductive fabric in the shorts, specifically in the left leg and in the area corresponding to the calf muscles, in order to have more physiological information regarding the muscles' activity. Furthermore, one of the points that should be further improved is the connection between the kinematic and physiological data. As such, we highly suggest the implementation of an inertial sensor, for example an accelerometer, in TraXports V2 in order to be able of collecting both signals at once. That way, it will be possible to have more accurate data about players' motion and physical state, meaning that the classifiers' performance will definitely increase.

Regarding the architecture itself, this still needs to be further improved. Firstly, during the step of feature extraction the EMG signal should be temporally segmented in order to localize each muscle contraction and to only extract features on these locations. Another suggestion that should be considered, is the addition of more features and activities for classification, for example, different types of shoots, or ball receiving. The use of more base classifiers in the DBMM, such as the HMM and RNNs, is also another way of improvement.

Furthermore, the architecture should be implemented in real time in order to produce a viable architecture for real time football prediction. Thus, after having all of this data combined, a software operating the algorithm of actions' classification, trajectories' estimation and fatigue detection should be implemented and tested during a football match.

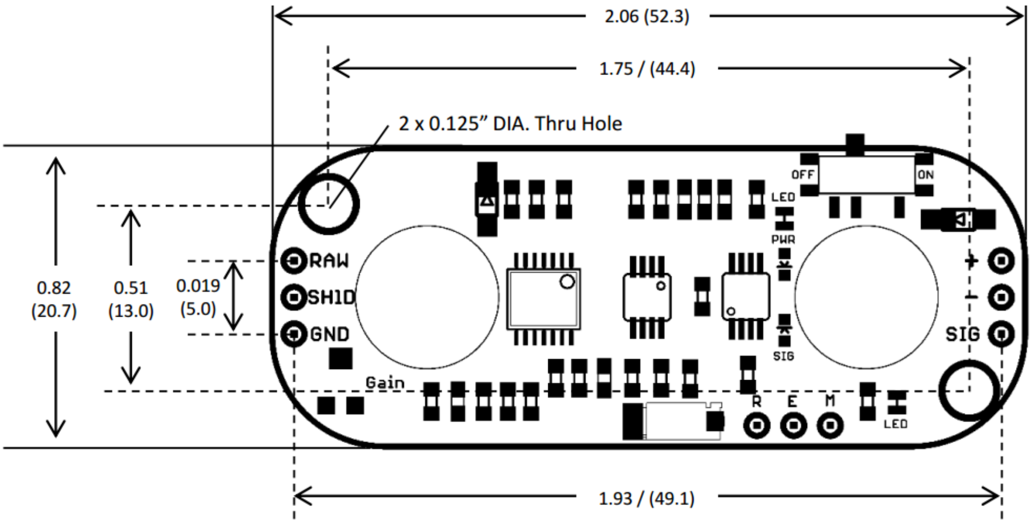
Appendices

A.1 – MyoWare Muscle Sensor (AT-04-001) Datasheet

A.1.1 – Electrical Specifications

Parameter	Min	TYP	Max
Supply Voltage	+2.9V	+3.3V or +5V	+5.7V
Adjustable Gain Potentiometer	0.01 Ω	50 kΩ	100 kΩ
Output Signal Voltage EMG Envelope Raw EMG (centered about +Vs/2)	 0V 0V	 -- --	 +Vs +Vs
Input Impedance	--	110 GΩ	--
Supply Current	--	9 mA	14 mA
Common Mode Rejection Ratio (CMRR)	--	110	--
Input Bias	--	1 pA	--

A.1.2 – Dimensions



A.2 – Phtoton Particle Datasheet

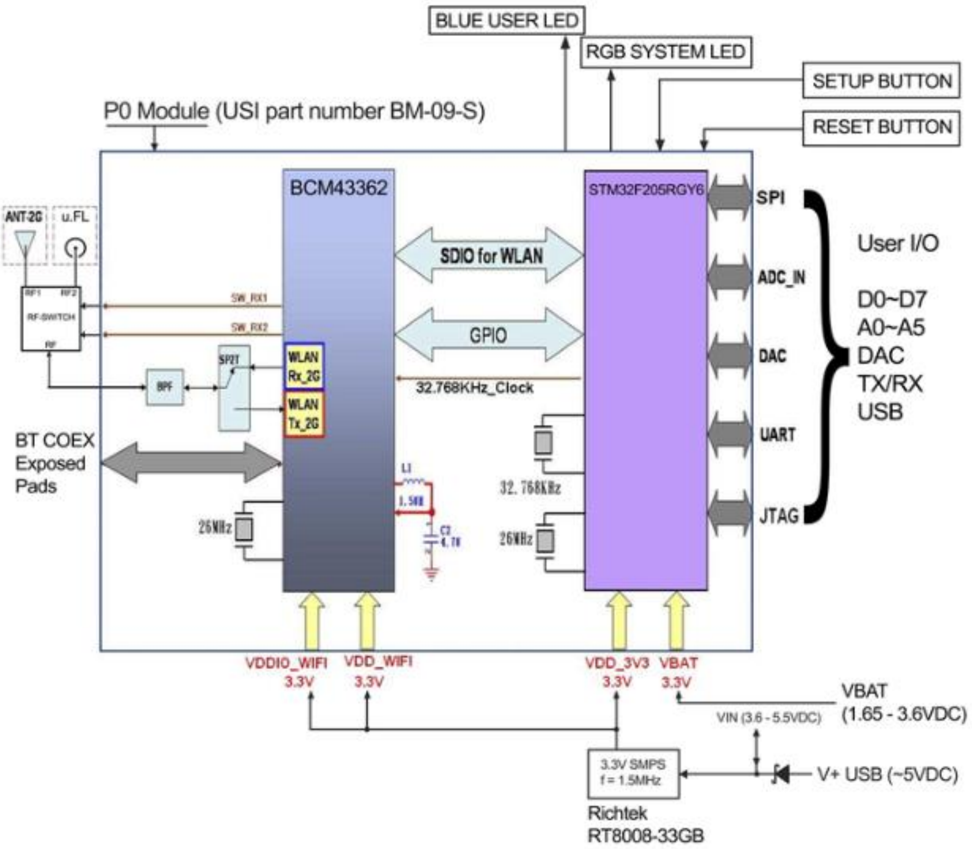
A.2.1 – Recommended operating conditions

Parameter	Symbol	Min	Typ	Max	Unit
Supply Input Voltage	V_{VIN}	+3.6		+5.5	V
Supply Input Voltage	V_{3V3}	+3.0	+3.3	+3.6	V
Supply Output Voltage	V_{VIN}		+4.8		V
Supply Output Voltage	V_{3V3}		+3.3		V
Supply Input Voltage	V_{VBAT}	+1.65		+3.6	V
Supply Input Current (VBAT)	I_{VBAT}			19	uA
Operating Current (Wi-Fi on)	$I_{VIN\ avg}$		80	100	mA
Operating Current (Wi-Fi on)	$I_{VIN\ pk}$	235[1]		430[1]	mA
Operating Current (Wi-Fi on, w/powersave)	$I_{VIN\ avg}$		18	100[2]	mA
Operating Current (Wi-Fi off)	$I_{VIN\ avg}$		30	40	mA
Sleep Current (5V @ VIN)	I_{Qs}		1	2	mA
Deep Sleep Current (5V @ VIN)	I_{Qds}		80	100	uA
Operating Temperature	T_{op}	-20		+60	°C

A.2.2 – Wi-Fi Specifications

Feature	Description
WLAN Standards	IEEE 802 11b/g/n
Antenna Port	Single Antenna
Frequency Band	2.412GHz -- 2.462GHz (United States of America and Canada)
	2.412GHz -- 2.472GHz (EU/Japan)
Sub Channels	1 -- 11 (United States of America and Canada)
	1 -- 13 (EU/Japan)
Modulation	DSSS, CCK, OFDM, BPSK, QPSK, 16QAM, 64QAM

A.2.3 – Block Diagram



A.3 - MedTex130 Datasheet

A.3.1 – Technical Specifications

SURFACE RESISTANCE	< 5 Ohms/
PLATING	99.9 % pure silver
ABRASION RESISTANCE	10,0000 cycles
TEMPERATURE RANGE	-30 to 90ºC
TOTAL THICKNESS	0.45 mm
WEIGHT	140 g/m^2
STRETCH	Double stretch direction
ROLL LENGTHS	50 LY average
ROLL WIDTH	135 cm
MATERIAL	78% Nylon + 22% elastomer

References

[1] B. Min, J. Kim, C. Choe, H. Eom, and R. B. McKay, "A compound framework for sports results prediction : A football case study," *Knowledge-Based Syst.*, vol. 21, pp. 551–562, 2008.

[2] C. P. Igiri and E. Okechukwu, "An Improved Prediction System for Football a Match Result," vol. 04, no. 12, pp. 12–20, 2014.

[3] M. Glickman and H. Stern, "A State-Space Model for National Football League Scores," *J. Am. Stat. Assoc.*, no. 93, pp. 25–35, 1998.

[4] M. Svensson and B. Drust, "Testing soccer players.," *J. Sports Sci.*, vol. 23, no. 6, pp. 601–618, Jun. 2005.

[5] B. Drust, G. Atkinson, and T. Reilly, "Future perspectives in the evaluation of the physiological demands of soccer.," *Sports Med.*, vol. 37, no. 9, pp. 783–805, 2007.

[6] M. A. Cengiz, N. Murat, H. Koc, and T. Senel, "A Bayesian Computation for The Prediction Of Football Match Results Using Artificial Neural Network," *Int. J. Sci. Knowl.*, vol. 1, no. 2, pp. 31–37, 2012.

[7] O. Farzin, E. Parinaz, and M. Faezeh, "Football Result Prediction with Bayesian Network in Spanish League-Barcelona Team," *Int. J. Comput. Theory Eng.*, vol. 5, no. 5, pp. 812-815, 2013.

[8] Z. Fan, Y. Kuang, and X. Lin, "Chess game results prediction system.," 2013.

[9] Rue and Salvesen, "Prediction and retrospective analysis of soccer matches in a league.," *Stat.*, vol. 49, no. 3, pp. 399–418, 2000.

[10] Ingeniarius, "About Us," 2016. [Online]. Available: http://ingeniarius.pt/?page=aboutus. [Accessed: 02-Jan-2017].

[11] M. S. Couceiro, G. Dias, D. Araujo, and K. Davids, "The ARCANE Project: How an Ecological Dynamics Framework Can Enhance Performance Assessment and Prediction in Football.," *Sports Med.*, vol. 46, no. 12, pp. 1781–1786, Dec. 2016.

[12] H. Hill, V. Bruce, and S. Akamatsu, "Perceiving the sex and race of faces: the role of shape and colour.," *Proc. Biol. Sci.*, vol. 261, no. 1362, pp. 367–373, Sep. 1995.

[13] N. Davidenko, "Silhouetted face profiles: a new methodology for face perception research.," *J. Vis.*, vol. 7, no. 4, p. 6, 2007.

[14] P. Sharma and M. Kaur, "Classification in Pattern Recognition : A Review," *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 3, no. 4, pp. 298–306, 2013.

[15] A. K. Jain, R. P. W. Duin, and J. Mao, "Statistical Pattern Recognition : A Review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 1, pp. 4-37, 2000.

[16] M. Paracher, S. Sharma, A. K. Sharma, and J. P. Gupta, "Anatomy on pattern recognition," *Indian J. Comput. Sci. Eng.*, vol. 2, no. 3, pp. 371–378, 2011.

[17] S. Asht and R. Dass, "Pattern Recognition Techniques : A Review," *Int. J. Comput. Sci. Telecommun.*, vol. 3, no. 8, 2012.

[18] C. M. Bishop, *Pattern recognition and machine learning*. Springer, 2006.

[19] P. Cunningham and S. J. Delany, "K-Nearest Neighbour Classifiers," *Mult. Classif. Syst.*, pp. 1–17, 2007.

[20] K. Q. Weinberger and L. K. Saul, "Distance Metric Learning for Large Margin Nearest Neighbor Classification," *Adv. Neural Inf. Process. Syst.*, pp. 1473–1480, 2005.
[21] R. Hanson, J. Stutz, and P. Cheeseman, "Bayesian Classification Theory," *Ames Res. Cent. - Artif. Intell. Res. Branch*, 1991.
[22] J. Grover, "A Literature Review of Bayes' Theorem and Bayesian Belief Networks (BBN)," in *Strategic Economic Decision-Making: Using Bayesian Belief Networks to Solve Complex Problems*, New York, NY: Springer New York, 2013, pp. 11–27.
[23] J. Ivansson, "Situation assessment in a stochastic environment using Bayesian Networks," Linköping University, 2002.
[24] F. Faltin and R. Kenett, "Bayesian Networks," *Encyclopedia of Statistics in Quality & Reliability*. Wiley & Sons, 2007.
[25] J. G. Birnberg, "Bayesian Statistics: A Review," *J. Account. Res.*, vol. 2, no. 1, pp. 108–116, 1964.
[26] J. M. Bernardo and A. F. M. Smith, *Bayesian Theory*. John Wiley & Sons, Inc., 2008.
[27] A. Shmilovici, "Support Vector Machines," in *Data Mining and Knowledge Discovery Handbook*, O. Maimon and L. Rokach, Eds. Boston, MA: Springer US, 2005, pp. 257–276.
[28] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intelligent Systems and their Applications*, vol. 13, no. 4. pp. 18–28, 1998.
[29] G. Mountrakis, J. Im, and C. Ogole, "Support vector machines in remote sensing: A review," *ISPRS J. Photogramm. Remote Sens.*, vol. 66, no. 3, pp. 247–259, May 2011.
[30] A. Ben-Hur and J. Weston, "Data mining techniques for the life sciences," in *A User's Guide to Support Vector Machines*, Humana Press, 2010, pp. 223–239.
[31] S. Ruping, "SVM Kernels forTime Series Analysis," *Dartmond's Univ.*, 2001.
[32] B. Scholkõpf, "The kernel trick for distances.," *Microsoft Res.*, 2000.
[33] R. Bajcsy and S. Kovačič, "Multiresolution Elastic Matching," *Comput. Vis. Graph. Image Process.*, vol. 46, no. 1, pp. 1–21, 1989.
[34] D. Graupe, *Principles of Artificial Neural Networks*, 2nd ed. River Edge, NJ, USA: World Scientific Publishing Co., Inc., 2007.
[35] J. S. Yadav, M. Yadav, and A. Jain, "Artificial Neural Network," *Int. J. Sci. Res. Educ.*, vol. 1, no. 6, pp. 108–118, 2014.
[36] S. Bajpai, K. Jain, and N. Jain, "Artificial Neural Networks," *Int. J. Soft Comput. Eng.*, vol. 1, pp. 13–14, 2011.
[37] E. P. Kumar and E. P. Sharma, "Artificial Neural Networks - A Study," *Int. J. Emerg. Eng. Res. Technol.*, vol. 2, no. 2, pp. 143–148, 2014.
[38] S. Agatonovic-Kustrin and R. Beresford, "Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research," *J. Pharm. Biomed. Anal.*, vol. 22, no. 5, pp. 717–727, Jun. 2000.
[39] I. A. Basheer and M. Hajmeer, "Artificial neural networks : fundamentals , computing , design , and application," *J. Microbiol. Methods*, vol. 43, pp. 3–31, 2000.
[40] A. Jain and I. Mao, "Artificial Neural Networks: A Tutorial," in *Artificial Inteligence: A New Synthesis*, ELSEVIERA, 1996, pp. 31–44.
[41] E. Alpaydin, *Introduction to Machine Learning Second Edition*, 2º ed. London: MIT Press, 2010.
[42] Z. Lipton, "A Critical Review of Recurrent Neural Networks for Sequence Learning." 2015.
[43] J. T. Connor, R. D. Martin, and L. E. Atlas, "Recurrent neural networks and robust time series prediction," *IEEE Transactions on Neural Networks*, vol. 5, no. 2. pp. 240–254, 1994.
[44] M. Sundermeyer, I. Oparin, J. L. Gauvain, B. Freiberg, R. Schlüter, and H. Ney, "Comparison of feedforward and recurrent neural network language models," *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. pp. 8430–8434, 2013.
[45] S. Garfield and S. Wermter, "Comparing Support Vector Machines, Recurrent Networks, and Finite State Transducers for Classifying Spoken Utterances," in *Proceedings of the 2003 Joint International Conference on Artificial Neural Networks and Neural Information Processing*, 2003, pp. 646-653.
[46] M. Cernansky and P. Tino, "Comparison of Echo State Networks with Simple Recurrent

Networks and Variable-Length Markov Models on Symbolic Sequences ⋆," in *Artificial Neural Networks*, 2007, pp. 618–627.
[47] W. De Mulder, S. Bethard, and M.-F. Moens, "A survey on the application of recurrent neural networks to statistical language modeling," *Comput. Speech Lang.*, vol. 30, no. 1, pp. 61–98, Mar. 2015.
[48] S. Fernández, A. Graves, and J. Schmidhuber, "An Application of Recurrent Neural Networks to Discriminative Keyword Spotting," in *Proceedings of the 17th International Conference on Artificial Neural Networks*, 2007, pp. 220–229.
[49] A. Malek, "Applications of Recurrent Neural Networks to Optimization Problems," in *Recurrent Neural Networks*, Balasubramaniam, Ed. InTech, 2008.
[50] T. G. Dietterich, "Ensemble Methods in Machine Learning," in *Proceedings of the First International Workshop on Multiple Classifier Systems*, 2000, pp. 1–15.
[51] D. R. Faria, C. Premebida, and U. Nunes, "A probabilistic approach for human everyday activities recognition using body motion from RGB-D images," *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*. pp. 732–737, 2014.
[52] D. R. Faria, M. Vieira, C. Premebida, and U. Nunes, "Probabilistic Human Daily Activity Recognition towards Robot-assisted Living," in *IEEE RO-MAN'15: IEEE International Symposium on Robot and Human Interactive Communication*, 2015.
[53] A. Loverro, "Fractional Calculus : History , Definitions and Applications for the Engineer," 2004.
[54] M. S. Couceiro, F. M. L. Martins, R. P. Rocha, and N. M. F. Ferreira, "Introducing the Fractional Order Robotic Darwinian PSO," in *9th International Conference on Mathematical Problems in Engineering, Aerospace and Sciences: ICNPAA 2012*, 2012.
[55] M. S. Couceiro, F. Clemente, F. Martins, and J. Machado, "Dynamical Stability and Predictability of Football Players: The Study of One Match," *Entropy*, vol. 16, no. 2, 2014.
[56] P. W. Ostalczyk, "A note on the Grünwald–Letnikov fractional-order backward-difference," *Phys. Scr.*, no. T136, pp. 1-5, 2009.
[57] O. Heino and V. Sillanpää, "Forecasting football match results - A study on modeling principles and efficiency of fixed-odds betting markets in football," Aalto University, 2013.
[58] S. Declan, "Modelling football match results and testing the efficiency of the betting market," National University of Ireland, Maynooth, 2012.
[59] K. S. Courneya and A. V. Carron, "The home advantage in sport competitions: a literature review," *J. Sport Exerc. Psychol.*, vol. 14, pp. 18–27, 1992.
[60] J. P. Jamieson, "The home field advantage in athletics: a meta-analysis," *J. Appl. Soc. Psychol.*, vol. 40, pp. 1819–1848, 2010.
[61] T. Moskowitz and J. Wertheim, "The 12th man," *The Economist*, London, 2014.
[62] A. M. Nevill and R. L. Holder, "Home advantage in sport: an overview of studies on the advantage of playing at home.," *Sports Med.*, vol. 28, no. 4, pp. 221–236, Oct. 1999.
[63] B. Schwartz and S. F. Barsky, "The Home Advantage," *Soc. Forces* , vol. 55 , no. 3 , pp. 641–661, Mar. 1977.
[64] D. L. Greer, "Spectator Booing and the Home Advantage: A Study of Social Influence in the Basketball Arena," *Soc. Psychol. Q.*, vol. 46, no. 3, pp. 252–261, 1983.
[65] A. M. Nevill, N. J. Balmer, and M. A. Williams, "The influence of crowd noise and experience upon refereeing decisions in football," *Psychol. Sport Exerc.*, vol. 3, no. 4, pp. 261–272, Oct. 2002.
[66] J. Goddard and I. Asimakopoulos, "Forecasting football results and the efficiency of fixed-odds betting," *J. Forecast.*, vol. 23, no. 1, pp. 51–66, 2004.
[67] R. Audas, S. Dobson, and J. Goddard, "The impact of managerial change on team performance in professional sports," *J. Econ. Bus.*, vol. 54, no. 6, pp. 633-650, Nov. 2002.
[68] C. Rusling, "THE FUNCTIONAL MOVEMENT SCREENING TOOL DOES NOT PREDICT INJURY IN FOOTBALL," *Prog. Orthop. Sci.*, vol. 1, no. 2, pp. 41–46, 2015.
[69] E. Giza, K. Mithofer, L. Farrell, B. Zarins, T. Gill, and S. Drawer, "Injuries in women's professional soccer," *Br. J. Sports Med.*, vol. 39, no. 4, pp. 212–216, Apr. 2005.
[70] S. Kampakis, "Comparison of machine learning methods for predicting the recovery time

of professional football players after an undiagnosed injury." Springer-Verlag Berlin Heidelberg, 2011.

[71] S. W. Shaffer and T. L. Uhl, "Preventing and Treating Lower Extremity Stress Reactions and Fractures in Adults," *J. Athl. Train.*, vol. 41, no. 4, pp. 466–469, 2006.

[72] C. M. Ferrara and E. Hollingsworth, "Physical characteristics and incidence of injuries in adult figure skaters.," *Int. J. Sports Physiol. Perform.*, vol. 2, no. 3, pp. 282–291, Sep. 2007.

[73] S. D. Mair, A. V Seaber, R. R. Glisson, and W. E. Garrett, "The Role of Fatigue in Susceptibility to Acute Muscle Strain Injury," *Am. J. Sport. Med.* , vol. 24 , no. 2 , pp. 137–143, Mar. 1996.

[74] S. Drawer and C. Fuller, "Evaluating the level of injury in English professional football using a risk based assessment process," *Br. J. Sports Med.*, vol. 36, no. 6, pp. 446–451, Dec. 2002.

[75] M. R. Al-mulla, F. Sepulveda, and M. Colley, "A Review of Non-Invasive Techniques to Detect and Predict Localised Muscle Fatigue," *Sensors*, vol. 11, pp. 3545–3594, 2011.

[76] O. Faude, T. Koch, and T. Meyer, "Straight sprinting is the most frequent action in goal situations in professional football.," *J. Sports Sci.*, vol. 30, no. 7, pp. 625–631, 2012.

[77] E. Rampinini, F. M. Impellizzeri, C. Castagna, A. J. Coutts, and U. Wisloff, "Technical performance during soccer matches of the Italian Serie A league: effect of fatigue and competitive level.," *J. Sci. Med. Sport*, vol. 12, no. 1, pp. 227–233, Jan. 2009.

[78] S. A. Dugan and W. R. Frontera, "Muscle fatigue and muscle injury.," *Phys. Med. Rehabil. Clin. N. Am.*, vol. 11, no. 2, pp. 385–403, May 2000.

[79] S. J. Cormack, M. G. Mooney, W. Morgan, and M. R. McGuigan, "Influence of Neuromuscular Fatigue on Accelerometer Load in Elite Australian Football Players," *Int. J. Sports Physiol. Perform.*, vol. 8, no. 4, pp. 373–378, Jul. 2013.

[80] G. Jones, "More than Just a Game: Research Developments and Issues in Competitive Anxiety in Sport," *Br. J. Psychol.*, vol. 4, pp. 449–478, 1995.

[81] S. Hanton, O. Thomas, and I. Maynard, "Competitive Anxiety Responses in the Week Leading up to Competition: The Role of Intensity, Direction and Frequency Dimensions," *Psychol. Sport Exerc.*, vol. 2, pp. 169–181, 2004.

[82] Z. Zhang, W. Zhou, Z. Cai, and S. Cui, "The Characteristics of Brain Function in Chinese Elite Table Tennis Players," *Chinese J. Sport. Med.*, vol. 5, pp. 452–457, 2002.

[83] Y. Lv and Q. Ma, "Research on the Characteristics of Information Processing of Athletes in the Brain with the Events Relative Potentials (ERP)," *J. Tianjin Inst. Phys. Educ.*, vol. 4, pp. 1–2, 1995.

[84] I. Hale, "Engineering Football. Why Spain will win...," *Eng. Technol.*, vol. 5, no. 18, pp. 24–27, 2010.

[85] R. H. Koning, "Balance in Competition in Dutch Soccer," *Stat.*, vol. 49, no. 3, pp. 419–431, 2000.

[86] M. Crowder, M. Dixon, A. Ledford, and M. Robinson, "Dynamic modeling and prediction of English football league matches for betting.," *Stat.*, vol. 51, pp. 157–168, 2002.

[87] P. Andersson, M. Ekman, and J. Edman, "Forecasting the fast and frugal way: a study of performance and information-processing strategies of experts and non-experts when predicting the world cup 2002 in soccer.," Stockholm, 2003.

[88] J. Goddard and I. Asimakopoulos, "Forecasting football results and the efficiency of fixed-odds betting.," *J. Forecast.*, vol. 23, pp. 51–66, 2004.

[89] F. Halicioglu, "Can we predict the outcome of the international football tournaments: The case of EURO 2000," *Doğuş Üniversitesi Derg.*, vol. 6, no. 1, pp. 112–122, 2005.

[90] A. P. Rotshtein, M. Posner, and A. B. Rakityanskaya, "Football predictions based on a fuzzy model with genetic and neural tuning.," *Cybern. Syst. Anal.*, vol. 41, no. 4, pp. 619–630, 2005.

[91] D. Buursma, "Predicting sports event from past result: Towards effective betting on football matches.," in *14th Twente Student Conference on IT*, 2011.

[92] W. Nivard and R. D. Mei, "Soccer analytics: Predicting the of soccer matches," University of Amsterdam, 2012.

[93] C. G. Kramer, T. Hagg, and B. Kemp, "Real-time measurement of muscle fatigue related changes in surface EMG.," *Med. Biol. Eng. Comput.*, vol. 25, no. 6, pp. 627–630, Nov. 1987.

[94] D. Moshou, I. Hostens, G. Papaioannou, and H. Ramon, “Dynamic muscle fatigue detection using self-organizing maps,” *Appl. Soft Comput.*, vol. 5, no. 4, pp. 391–398, Jul. 2005.

[95] J. H. Serfntein, “Aprediction model for the prevention of soccer injuries amongst youth players,” 2009.

[96] C. Carling, J. Bloomfield, L. Nelsen, and T. Reilly, “The Role of Motion Analysis in Elite Soccer,” *Sport. Med.*, vol. 38, no. 10, pp. 839–862, 2008.

[97] B. Eskofier, M. Oleson, C. DiBenedetto, and J. Hornegger, “Embedded Surface Classification in Digital Sports,” *Pattern Recognit. Lett.*, vol. 30, no. 16, pp. 1448–1456, 2009.

[98] H. Zhou and H. Hu, “Human motion tracking for rehabilitation - A survey,” *Biomed. Signal Process. Control*, vol. 3, no. 1, pp. 1–18, 2008.

[99] X. Chen, “Human Motion Analysis with Wearable Inertial,” 2013.

[100] M. S. Couceiro, F. M. Clemente, and F. M. L. Martins, “Analysis of football player’s motion in view of fractional calculus,” *Cent. Eur. J. Phys.*, vol. 11, no. 6, pp. 714–723, 2013.

[101] M. S. Couceiro, G. Dias, F. M. L. Martins, and J. M. A. Luz, “A fractional calculus approach for the evaluation of the golf lip-out,” *Signal, Image Video Process.*, vol. 6, no. 3, pp. 437–443, 2012.

[102] V. Correia, D. Araújo, K. Davids, O. Fernandes, and S. Fonseca, “Territorial gain dynamics regulates success in attacking sub-phases of team sports,” *Psychol. Sport Exerc.*, vol. 12, no. 6, pp. 662–669, Nov. 2011.

[103] J. Headrick, K. Davids, I. Renshaw, D. Araujo, P. Passos, and O. Fernandes, “Proximity-to-goal as a constraint on patterns of behaviour in attacker-defender dyads in team games.,” *J. Sports Sci.*, vol. 30, no. 3, pp. 247–253, 2012.

[104] J. L. Copete, J. Suzuki, Q. Wei, R. Iwaki, N. Endo, H. Mori, Y. Nagai, and M. Asada, “Estimation of Players’ Actions in Soccer Matches Based On Deep Autoencoder,” Osaka, Japan, 2015.

[105] N. Lee and K. M. Kitani, “Predicting wide receiver trajectories in American football,” in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2016.

[106] M. Joyner, “Physiological limiting factors and distance running: influence of gender and age on record performances.,” *Exerc. Sport Sci. Rev.*, vol. 21, pp. 103–133, 1993.

[107] M. Joyner and E. Coyle, “Endurance exercise performance: the physiology of champions,” *J. Physiol.*, vol. 586, no. Pt 1, pp. 35–44, Jan. 2008.

[108] S. Jerritta, M. Murugappan, R. Nagarajan, and K. Wan, “Physiological signals based human emotion Recognition: a review,” *Signal Processing and its Applications (CSPA), 2011 IEEE 7th International Colloquium on*. pp. 410–415, 2011.

[109] R. Merletti and P. A. Parker, *Electromyography: Physiology, Engineering, and Non-Invasive Applications*. John Wiley & Sons, 2004.

[110] F. Billaut, “Electromyography Assessment of Muscle Recruitment Strategies During High-Intensity Exercise,” in *Advances in Applied Electromyography*, J. B. T.-A. in A. E. Mizrahi, Ed. Rijeka: InTech, 2011.

[111] P. Konrad, “The ABC of EMG: A Practical Introduction to Kinesiological Electromyography,” *Noraxon, Inc.*, 2005.

[112] C. J. De Luca, “urface Electromyography: Detection and Recording,” 2002. [Online]. Available: https://www.delsys.com/Attachments_pdf/WP_SEMGintro.pdf.

[113] L. Bao and S. S. Intille, “Activity Recognition from User-Annotated Acceleration Data,” *Pervasive Comput.*, pp. 1–17, 2004.

[114] T. Huynh, “Human Activity Recognition with Wearable Sensors,” Technische Universität Darmstadt, 2008.

[115] J. Suutala, S. Pirttikangas, and J. Roning, “Discriminative temporal smoothing for activity recognition from wearable sensors,” *Ubiquitous Comput. Syst.*, pp. 182–195, 2007.

[116] M. Ermes and J. Parkk, “Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions,” *IEEE Trans. Inf. Technol. Biomed.*, pp. 20-26, 2008.

[117] P. J. Beek, C. E. Peper, and D. F. Stegeman, “Dynamical models of movement coordination,” *Hum. Mov. Sci.*, vol. 14, no. 4, pp. 573–608, 1995.

[118] F. H. Chan, Y. S. Yang, F. K. Lam, Y. T. Zhang, and P. A. Parker, “Fuzzy EMG classification for prosthesis control.,” *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 3, pp. 305–311, Sep. 2000.

[119] J. Jeong, W. Cho, Y. Kim, and H. Choi, “Recognition of Lower Limb Muscle EMG Patterns by using Neural Networks during the Postural Balance Control BT - 3rd Kuala Lumpur International Conference on Biomedical Engineering 2006: Biomed 2006, 11 – 14 December 2006 Kuala Lumpur, Malaysia,” F. Ibrahim, N. A. A. Osman, J. Usman, and N. A. Kadri, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 82–85.

[120] C. D. Joshi, U. Lahiri, and N. V Thakor, “Classification of gait phases from lower limb EMG: Application to exoskeleton orthosis,” *2013 IEEE Point-of-Care Healthcare Technologies (PHT)*. pp. 228–231, 2013.

[121] S. Barris and C. Button, “A review of vision-based motion analysis in sport.,” *Sports Med.*, vol. 38, no. 12, pp. 1025–1043, 2008.

[122] M. B. Raez, M. S. Hussain, and F. Mohd-Yasin, “Techniques of EMG signal analysis: Detection, processing, classification and applications,” *Biol. Proc. Online*, vol. 8, pp. 11–35, 2006.

[123] A. Phinyomark, S. Thongpanja, H. Hu, P. Phukpattaranont, and C. Limsakul, “The Usefulness of Mean and Median Frequencies in Electromyography Analysis,” G. R. B. T.-C. I. in E. A.-A. P. on C. A. and F. C. Naik, Ed. Rijeka: InTech, 2012, p. Ch. 08.

[124] A. Avci, S. Bosch, M. Marin-Perianu, R. Marin-Perianu, and P. Havinga, “Activity Recognition Using Inertial Sensing for Healthcare, Wellbeing and Sports Applications: A Survey,” *23th International Conference on Architecture of Computing Systems 2010*. pp. 1–10, 2010.

[125] Ó. D. Lara and M. A. Labrador, “A survey on human activity recognition using wearable sensors,” *Commun. Surv. Tutorials, IEEE*, vol. 15, no. 3, pp. 1192–1209, 2013.

[126] M. Zhang and A. A. Sawchuk, “Human Daily Activity Recognition With Sparse Representation Using Wearable Sensors,” *IEEE J. Biomed. Heal. Informatics*, vol. 17, no. 3, pp. 553–560, 2013.

[127] U. Jensen, M. Ring, and B. Eskofier, “Generic Features for Biosignal Classification,” in *Proceedings Sportinformatik 2012 9. Symposium der dvs-Sektion Sportinformatik, Univ. Konstanz*, 2012, pp. 162–168.

[128] W. Mohd, B. Wan, A. B. Yahya, C. S. Horng, and M. F. Sulaima, “Features Extraction of Electromyography Signals in Time Domain on Biceps Brachii Muscle,” *Int. J. Model. Optim.*, vol. 3, no. 6, pp. 3–7, 2013.

[129] S. Sharma, G. Kumar, S. Kumar, and D. Mohapatra, “Techniques for Feature Extraction from EMG Signal,” *Int. J. Adv. Res. Tech. Featur. Extr. from EMG Signal*, vol. 2, no. 1, 2012.

[130] A. Del Boca and D. C. Park, “Myoelectric signal recognition using fuzzy clustering and artificial neural networks in real time,” in *IEEE International Conference on Neural Networks, IEEE World Congress on Computational Intelligence*, 1994, vol. 5, pp. 3098–3103.

[131] M. Al-Mulla and F. Sepulveda, “Predicting the time to localized muscle fatigue using ANN and evolved sEMG feature,” in *Proceedings of 2010 International Conference on Autonomous and Intelligent Systems (AIS)*, 2010, pp. 1–6.

[132] M. Cifrek, V. Medved, S. Tonkovic, and S. Ostojic, “Surface EMG based muscle fatigue evaluation in biomechanics.,” *Clin. Biomech. (Bristol, Avon)*, vol. 24, no. 4, pp. 327–340, May 2009.

[133] R. Merletti and P. A. Parker., *Surface Electromyography: Physiology, Engineering and Applications*. Wiley, 2016.

[134] U. Maurer, A. Smailagic, D. P. Siewiorek, and M. Deisher, “Activity recognition and monitoring using multiple sensors on different body positions,” *International Workshop on Wearable and Implantable Body Sensor Networks (BSN’06)*. p. 4 pp.–116, 2006.

[135] K. Van Laerhoven and O. Cakmakci, “What shall we teach our pants?,” *Digest of Papers. Fourth International Symposium on Wearable Computers*. pp. 77–83, 2000.

[136] J.-Y. Yang, J.-S. Wang, and Y.-P. Chen, “Using acceleration measurements for activity recognition: An effective learning algorithm for constructing neural classifiers,” *Pattern Recognit. Lett.*, vol. 29, no. 16, pp. 2213–2220, Dec. 2008.

[137] T. Huynh and B. Schiele, “Analyzing Features for Activity Recognition,” in *Joint*

Conference on Smart objects and ambient intelligence: innovative context-aware services: usages and technologies, 2005.

[138] M. Oladazimi, F. Molaei-Vaneghi, and M. J. Safari, “A Review For Feature Extraction of EMG Signal Processing,” in *4th International Conference on Computer and Automation Engineering (ICCAE 2012)*, 2012.

[139] J. Mantyjarvi, J. Himberg, and T. Seppanen, “Recognizing human motion with multiple acceleration sensors,” *2001 IEEE International Conference on Systems, Man and Cybernetics. e-Systems and e-Man for Cybernetics in Cyberspace (Cat.No.01CH37236)*, vol. 2. pp. 747–752 vol.2, 2001.

[140] N. Bidargaddi, A. Sarela, L. Klingbeil, and M. Karunanithi, “Detecting walking activity in cardiac rehabilitation by using accelerometer,” *2007 3rd International Conference on Intelligent Sensors, Sensor Networks and Information*. pp. 555–560, 2007.

[141] E. Jang, B. Park, S. Kim, and J. Sohn, “Emotion Classification by Machine Learning Algorithm using Physiological Signals,” in *Proceedings of 2012 4th International Conference on Machine Learning and Computing*, 2012, vol. 25, pp. 1–5.

[142] E. Jang, B. Park, S. Kim, M. Park, and J. Sohn, “Classification of Human Emotions from Physiological signals using Machine Learning Algorithms,” in *ACHI 2013 : The Sixth International Conference on Advances in Computer-Human Interactions*, 2013, no. c, pp. 395–400.

[143] I. Mohino-Herranz, R. Gil-Pita, J. Ferreira, M. Rosa-Zurera, and F. Seoane, “Assessment of Mental, Emotional and Physical Stress through Analysis of Physiological Signals Using Smartphones,” *Sensors (Basel).*, vol. 15, no. 10, pp. 25607–25627, Oct. 2015.

[144] A. Granero, F. Fuentes-Hurtado, V. Naranjo Ornedo, J. Guixeres Provinciale, J. M. Ausín, and M. Alcañiz Raya, “A Comparison of Physiological Signal Analysis Techniques and Classifiers for Automatic Emotional Evaluation of Audiovisual Contents,” *Front. Comput. Neurosci.*, vol. 10, p. 74, Jul. 2016.

[145] J. Wang, P. Lei, K. Wang, L. Mao, and X. Chai, “Mood States Recognition of Rowing Athletes Based on Multi-Physiological Signals Using PSO-SVM,” *E-Health Telecommun. Syst. Networks*, vol. 3, pp. 9-17, 2014.

[146] N. Ravi, N. Dandekar, P. Mysore, and M. L. Littman, “Activity Recognition from Accelerometer Data,” in *Proceedings of the 17th Conference on Innovative Applications of Artificial Intelligence - Volume 3*, 2005, pp. 1541–1546.

[147] E. A. Heinz, K. S. Kunze, M. Gruber, D. Bannach, and P. Lukowicz, “Using Wearable Sensors for Real-Time Recognition Tasks in Games of Martial Arts - An Initial Experiment,” *2006 IEEE Symposium on Computational Intelligence and Games*. pp. 98-102, 2006.

[148] D. M. Karantonis, M. R. Narayanan, M. Mathie, N. H. Lovell, and B. G. Celler, “Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 10, no. 1. pp. 156–167, 2006.

[149] M. J. Mathie, B. G. Celler, N. H. Lovell, and A. C. F. Coster, “Classification of basic daily movements using a triaxial accelerometer,” *Med. Biol. Eng. Comput.*, vol. 42, no. 5, pp. 679–687, 2004.

[150] B. Park, E. Jang, S. Kim, C. Huh, and J. Sohn, “Three Differential Emotions and Classification Using Physiological Signals,” in *Proceedings of 2012 4th International Conference on Machine Learning and Computing*, 2012, vol. 25, pp. 55–59.

[151] P. Zappi, C. Lombriser, T. Stiefmeier, E. Farella, D. Roggen, L. Benini, and G. Tröster, “Activity Recognition from On-Body Sensors: Accuracy-Power Trade-Off by Dynamic Sensor Selection,” in *Wireless Sensor Networks: 5th European Conference, EWSN 2008, Bologna, Italy, January 30-February 1, 2008. Proceedings*, R. Verdone, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 17–33.

[152] J. Vital, D. Faria, G. Dias, M. S. Couceiro, F. Coutinho, and N. Ferreira, “Combining discriminative spatiotemporal features for daily life activity recognition using wearable motion sensing suit,” *Pattern Anal. Appl.*, pp. 1–16, 2016.

[153] C. I. Christodoulou and C. S. Pattichis, “A new technique for the classification and decomposition of EMG signals,” in *Proceedings of IEEE international conference on neural networks (ICNN’95)*, Perth, Western Australia, 1995, pp. 2303–2308.

[154] K. Veer and T. Sharma, “A novel feature extraction for robust EMG pattern

recognition.," *J. Med. Eng. Technol.*, vol. 40, no. 4, pp. 149–154, 2016.
[155] FIRSTBEAT, "Professional Sports for Teams," 2016. [Online]. Available: https://www.firstbeat.com/en/professional-sports/team-solutions/. [Accessed: 01-Nov-2016].
[156] FIRSTBEAT, "Technical," 2016. [Online]. Available: https://www.firstbeat.com/en/professional-sports/team-solutions/technical/. [Accessed: 01-Nov-2016].
[157] FIRSTBEAT, "Clients," 2016. [Online]. Available: https://www.firstbeat.com/en/professional-sports/team-solutions/client-list/. [Accessed: 01-Nov-2016].
[158] Catapult, "Catapult," 2016. [Online]. Available: http://www.catapultsports.com/. [Accessed: 01-Nov-2016].
[159] Catapult, "Clients," 2016. [Online]. Available: http://www.catapultsports.com/clients/. [Accessed: 01-Nov-2016].
[160] GPSports, "Football," 2016. [Online]. Available: http://gpsports.com/football/. [Accessed: 01-Nov-2016].
[161] GPSports, "Hardware," 2016. [Online]. Available: http://gpsports.com/spi-hpu/. [Accessed: 01-Nov-2016].
[162] Myontec, "Elite Sports," 2016. [Online]. Available: http://www.myontec.com/benefits/elite-sport/. [Accessed: 01-Nov-2016].
[163] Myontec, "Muscle Monitor," 2016. [Online]. Available: http://www.myontec.com/products/muscle_monitor/. [Accessed: 01-Nov-2016].
[164] STATS, "Team Performance - Football Soccer," 2017. [Online]. Available: https://www.stats.com/football/#dynamix. [Accessed: 27-Jan-2017].
[165] L. Ingeniarius, "TraXports," 2016. [Online]. Available: http://ingeniarius.pt/?page=traxports#. [Accessed: 21-Oct-2016].
[166] E. J. Ciaccio, S. M. Dunn, and M. Akay, "Biosignal pattern recognition and interpretation systems," *IEEE Engineering in Medicine and Biology Magazine*, vol. 12, no. 3. pp. 89–95, 1993.
[167] K. Kiesel and W. G. Hall, "CAN SERIOUS INJURY IN PROFESSIONAL FOOTBALL BE PREDICTED BY A PRESEASON FUNCTIONAL MOVEMENT screen ? CORRESPONDENCE :," vol. 2, no. 3, pp. 147–158, 2007.
[168] U. Johnson and A. Ivarsson, "Psychological predictors of injuries among junior soccer players," *Scand. J. Med. Sci. Sport*, vol. 21, no. 1, pp. 129–136, 2010.
[169] M. S. Couceiro, A. Araújo, S. Pereira, G. Dias, and D. Araújo, "A Pervasive Wearable Tracking System for Team Sports Analysis," 2017.
[170] "Particle," 2016. [Online]. Available: https://www.particle.io/.
[171] Mathworks, "Key Features," 2017. [Online]. Available: https://www.mathworks.com/products/matlab/features.html. [Accessed: 30-Jan-2017].
[172] D. Kleppner and R. Kolenkow, *An introduction to mechanics*. Cambridge University Press, 2013.
[173] A. Perea, J. Castellano, L. Alday, and A. Hernández-Mendo, "Analysis of behaviour in sports through Polar Coordinate Analysis with MATLAB®," *Qual. Quant.*, vol. 46, no. 4, pp. 1249–1260, 2012.
[174] C. Penas and M. Anguera, "Use of the polar coordinates technique to study interactions among professional Soccer players," .*Rev Port Cien Desp*, vol. 2, no. 4, pp. 21–40, 2002.
[175] W. Rose, "Electromyogram analysis," *Math. Signal Process. Biomech.*, 2014.
[176] R. Altimari, J. L. Dantas, M. Bigliassi, T. Kanthack, A. Moraes, and T. Abrão, "Influence of Different Strategies of Treatment Muscle Contraction and Relaxation Phases on EMG Signal Processing and Analysis During Cyclic Exercise," in *Computational Intelligence in Electromyography Analysis - A Perspective on Current Applications and Future Challenges*, Rijeka: InTech, 2012.
[177] R. G. T. Mello, L. F. Oliveira, and J. Nadal, "Digital Butterworth filter for subtracting noise from low magnitude surface electromyogram.," *Comput. Methods Programs Biomed.*, vol. 87, no. 1, pp. 28–35, Jul. 2007.
[178] M. Dash and H. Liu, "Feature selection for classification," *Intell. Data Anal.*, vol. 1, no. 1, pp. 131–156, 1997.

[179] J. Davis and M. Goadrich, “The Relationship Between Precision-Recall and ROC Curves,” in *Proceedings of the 23rd International Conference on Machine Learning*, 2006, pp. 233–240.

[180] K. Morton and P. Torrione, “MATLAB Pattern Recognition Open Free and Easy.” [Online]. Available: http://covartech.github.io/. [Accessed: 08-May-2017].