NOVA SCHOOL OF SCIENCE & TECHNOLOGY

DEPARTMENT OF PHYSICS

BEATRIZ AFONSO MENDONÇA

Bachelor of Science in Biomedical Engineering

CLASSIFICATION OF OPERATOR'S WORKLOAD BASED ON PHYSIOLOGICAL RESPONSE

DISSERTATION TO OBTAIN THE MASTER'S DEGREE ON BIOMEDICAL ENGINEERING

MASTER IN BIOMEDICAL ENGINEERING

NOVA University Lisbon
January, 2023

NOVA SCHOOL OF SCIENCE & TECHNOLOGY

DEPARTMENT OF PHYSICS

CLASSIFICATION OF OPERATOR'S WORKLOAD BASED ON PHYSIOLOGICAL RESPONSE

DISSERTATION TO OBTAIN THE MASTER'S DEGREE ON BIOMEDICAL ENGINEERING

BEATRIZ AFONSO MENDONÇA

Bachelor of Science in Biomedical Engineering

Adviser: Hugo Gamboa
Associated Professor, NOVA University Lisbon

Co-adviser: Luís Silva
Postdoctoral Researcher, NOVA University Lisbon in Biosignals Group at LIBPhys-UNL

Examination Committee

Chair: Ricardo Vigário
Associated Professor, FCT-NOVA

Rapporteur: João Vaz
Associate Professor, Egas Moniz School of Health and Science

Adviser: Luís Silva
Postdoctoral Researcher, NOVA University Lisbon in Biosignals Group at LIBPhys-UNL

MASTER IN BIOMEDICAL ENGINEERING

NOVA University Lisbon
January, 2023

Classification of operator's workload based on physiological response

Copyright © Beatriz Afonso Mendonça, NOVA School of Science and Technology, NOVA University Lisbon.
The NOVA School of Science and Technology and the NOVA University Lisbon have the right, perpetual and without geographical boundaries, to file and publish this dissertation through printed copies reproduced on paper or on digital form, or by any other means known or that may be invented, and to disseminate through scientific repositories and admit its copying and distribution for non-commercial, educational or research purposes, as long as credit is given to the author and editor.

This document was created with the (pdf/Xe/Lua)LaTeX processor and the NOVAthesis template (v6.9.15) [1].

This work is dedicated to my guardian angel, my star, my sister.

Acknowledgements

Firstly, I would like to thank my advisors. To Professor Hugo Gamboa for believing, encouraging, and integrating me in this OPERATOR project. I also thank him for demanding, not only from me, but from all his students, rigor in work, cooperation between all his team, and for worrying about the well-being of all and each of us. To Luís Silva, who was tireless in giving me support, in explaining over and over what was needed, in encouraging me to appreciate what I was doing and not be afraid to ask. Moreover, I thank him for teaching far beyond the strictly necessary. Last, but not least, to Mariana Dias, who also has always been available to help me and discuss ideas. Regardless of what have been said, it's impossible to forget the members of the LibPhysics laboratory, who, not only in working hours, but even at lunch time have always integrated me, contributing to increase my motivation for this course, which had been slightly lost with the COVID-pandemic 19, due to classes online, at home, without Knowing when things would return to normal.

Falando agora em português, quero agradecer à minha família de Sangue, do Coração e àquela que Deus me deu, por me apoiarem sempre em tudo e por me darem o amor e carinho incondicional, que jamais conseguirei retribuir. Aos meu pais por me mostrarem que basta querer aprender para se saber fazer. Ao meu pai, por me despertar, desde cedo, o interesse em observar, experimentar, testar e aprender coisas novas, desde mudar cabos de eletricidade, passando por me desenrascar com computadores, cozinhar ou até mesmo bordar. Mas, principalmente e em especial, à minha mãe, a pessoa que mais admiro: uma mulher lutadora, resiliente e visionária, que soube levantar os vários pedregulhos que encontrou no caminho, transportá-los, construir uma ponte com eles, atravessá-la e conseguir alcançar o que muitos diziam ser impossível. Aos meus padrastos por fazerem os meus pais felizes e por me tratarem também como uma filha.

Aos amigos que nesta vida fui criando e que permaneceram sempre no meu coração. Aos amigos dos Açores, do Repolariza-te e Bandidas. Destaque especial à grande família que criei na residência, Gente não Deescreta, porque a convivência diária com estas miúdas ajudou-me a criar competências que a faculdade não me deu. Às companhias no Caleidoscópio, Galveias, no Zoom ou no Gather que permitiram juntas, chegarmos todas a este fim.

"Fatigue, discomfort, discouragement are merely symptoms of effort." (Morgon Freeman)

Abstract

People spend most of their lives at work, during which time they are exposed to mechanical and environmental conditions that can harm their health. This risk can occur in an hour-long or over long periods, even when performed at a light to moderate intensity due to cumulative fatigue. Several measures have been proposed in order to prevent or reduce fatigue-inducing repetitive work. However, these measures are essentially subjective or only measure fatigue locally. Wearables are an attractive solution to measure work-related fatigue globally and at any time.

The purpose of this study is to quantify biosignals information for the determination of fatigue while performing repetitive work. Electrocardiogram (ECG), electromyography (EMG), respiratory inductance plethysmography (RIP) and Accelerometer (ACC) signals were collected from 25 healthy participants. The participants were instructed to perform a repetitive task after induced fatigue. Their biosignals were processed, and different families of features were extracted. These features were used to fit a classifier in order to evaluate fatigue. Self-Similarity Matrix (SSM) was used to select and segment the data in Baseline and Fatigue.

Autocorrelation of inertial measures, respiratory synchrony, and the root mean square of the cardiovascular load features achieved 88% of accuracy. It was possible to verify that the ACC's features lead to the best classification results, followed by the RIP, EMG and finally the ECG's features.

Multimodal data allows global classification of when a person is working after experiencing fatigue. Motor information contributes significantly to this classification due to compensations that occur while performing the repetitive task. More studies should be done to develop an index characterising the fatigue state.

Keywords: Workload, Fatigue, Machine Learning, Operator, Occupational Risks, Industry 4.0, ECG, EMG, RIP, ACC

Resumo

As pessoas passam a maior parte da sua vida a trabalhar. A exposição a condições mecânicas e ambientais no trabalho pode ser prejudicial à sua saúde. Este risco pode ocorrer devido à fadiga cumulativa. Lesões podem surgir tanto em curtos como em longos períodos, mesmo quando a tarefa tem uma intensidade leve a moderada. Várias medidas foram propostas para prevenir ou reduzir o trabalho repetitivo que induz fadiga, no entanto, estas medidas são essencialmente subjetivas ou apenas medem a fadiga localmente. Os *wearables* são uma solução interessante para medir a fadiga relacionada ao trabalho a nível global e em qualquer momento.

O objetivo deste estudo foi quantificar informações de biosinais para a determinação da fadiga durante a realização de trabalhos repetitivos. Os sinais de eletrocardiograma (ECG), eletromiografia (EMG), pletismografia de indutância respiratória (RIP) e acelerómetro (ACC) foram recolhidos de 25 participantes saudáveis. Os participantes realizaram uma tarefa repetitiva onde fadiga foi provocada. Os biosinais foram processados, e diferentes famílias de métricas foram extraídas. Estas métricas foram usadas para classificar a fadiga. Recorreu-se a Matrizes de Auto-Similaridade (SSM) para selecionar e segmentar os dados em fadiga e não fadiga.

A autocorrelação das medidas inerciais, a sincronia respiratória e o quadrado médio da raiz da carga cardiovascular alcançaram 88% de precisão. Foi possível verificar que as features do ACC tiveram os melhores resultados de classificação, seguindo-se do RIP, EMG e, por último, de ECG.

Os dados multimodais permitiram a classificação global de quando uma pessoa está a trabalhar, após sentir fadiga. A informação motora contribui, significativamente, para esta classificação devido às compensações que ocorrem durante a realização da tarefa repetitiva. Futuro trabalho deve ser feito com fim a determinar um índice que possa caracterizar o estado de fadiga.

Palavras-chave: Carga de Trabalho, Fadiga, *Machine Learning*, Operário, Riscos Ocupacionais, Indústria 4.0, ECG, EMG, RIP, ACC

Contents

List of Figures x

List of Tables xii

Acronyms xiv

1 Introduction 1
- 1.1 Context and Motivation . . . 1
- 1.2 Objectives . . . 2
- 1.3 Document Structure . . . 3

2 Theoretical Background 4
- 2.1 Occupational Health . . . 4
  - 2.1.1 Work-related Musculoskeletal Disorders (WRMDs) . . . 4
  - 2.1.2 Ergonomics . . . 7
  - 2.1.3 Fatigue Mechanisms . . . 9
- 2.2 Biosignals . . . 10
  - 2.2.1 Electrocardiogram . . . 10
  - 2.2.2 Respiratory Inductance Plethysmography . . . 11
  - 2.2.3 Electromyography . . . 12
  - 2.2.4 Accelerometer . . . 12
- 2.3 Machine Learning . . . 13
  - 2.3.1 Supervised Learning Models . . . 13

3 State of Art 15

4 Methods and Materials 18
- 4.1 Method Schema . . . 18
- 4.2 Data Acquisition . . . 19
  - 4.2.1 Participants . . . 19
  - 4.2.2 Task . . . 19

4.2.3 Materials 19
4.2.4 Protocol 21
4.3 Pre-processing 22
4.3.1 Transfer Function 22
4.3.2 ECG Signal Processing 23
4.3.3 RIP Signal Processing 32
4.3.4 EMG signals processing 33
4.3.5 ACC Signal Processing 34
4.3.6 Damaged data deletion 34
4.4 Annotation 35
4.5 Features Selection and Extraction 38
4.6 Classification Model Development 39

5 Results **41**
5.1 Features Selection 41
5.2 Model's Classification 45

6 Discussion **47**

7 Conclusion and Future Work **51**

Bibliography **53**

Annexes

I R Peaks Detetion **66**

II Segmentation **69**

III Features Selection **71**

IV Models Development **72**

List of Figures

2.1 Percentage of workers from different job sectors. 5
2.2 The main work-related musculoskeletal disorders. 5
2.3 Illustration of the shoulder rotation in task under study, by temporal step order. 8
2.4 The Figure shows how the Cardiac Conduction System is constituted. 11
2.5 Types of rotation were studied in the wrist through an accelerometer. 12

4.1 Thesis work' schema. 18
4.2 The layout of the manual handling of the box done on this study's acquisition date. 19
4.3 Devices used in data acquisition. 20
4.4 Data acquisition protocol's schema 21
4.5 HRV's Representation of the same ECG signal calculated by some tested algorithms for R peak detection. 24
4.6 The same signal was used in these graphs to detect the R Peaks with different models. 25
4.7 Schema of the proposed R peak detection algorithm. 25
4.8 ECG original signal comparison with bandpass filters of different order frequency ranges. 26
4.9 Shape of different Wavelets Transform tested 26
4.10 ECG Signal's comparison. 28
4.11 Analysis of the various wavelets. 29
4.12 Comparation of the filtered signals between wavelet db2 and bandpass. . . 29
4.13 Representation of the different steps' signals for detection of the QRS Complex. 30
4.14 In graph **A**, the detected R peaks' location can be seen in the QRS Complex Restriction Signal, and in the graph **B**, the same is compared with the original signal. 31
4.15 HRV signal calculated using the model used in this study. 31
4.16 RIP signals processing schema. 33
4.17 EMG signals processing's schema. 34

4.18 ACC signals processing's schema. 34
4.19 Analysis of different signals with the SSM matrix and the extracted function for the same individual. 36
4.20 Example of 4 signals, one of each signal type corresponding to the same subject sample used later to extract features. 37

5.1 The evolution of the Accuracy of the various algorithms is based on the number of features added to the signal. 42
5.2 The Accuracy of the various algorithms is based on the features' number added to the signal. 43
5.3 Scatter Plots between different features. 44

I.1 These graphs represent the HRV of the same ECG signal calculated by some tested algorithms for R peak detection - Part I 66
I.2 These graphs represent the HRV of the same ECG signal calculated by some tested algorithms for R peak detection - Part II 67
I.3 HRV signal of various subjects and tasks obtained here after processing. . . 68

II.1 Another example of SSM targeting a subject 69
II.2 All segmented signals from a sample. 70

IV.1 Evaluation metrics by all algorithms of classifiers trained with Leave-one-out Cross Validation and with ECG data set. 72
IV.2 Evaluation metrics by all algorithms of classifiers trained with Leave-one-out Cross Validation and with RIP data set. 73
IV.3 Evaluation metrics by all algorithms of classifiers trained with Leave-one-out Cross Validation and with EMG data set. 74
IV.4 Evaluation metrics by all algorithms of classifiers trained with Leave-one-out Cross Validation and with ACC data set. 75
IV.5 Evaluation metrics by all algorithms of classifiers trained with Leave-one-out Cross Validation and with All data sets. 76
IV.6 Receiver Operating Characteristic (ROC) curve for general classification . . 77
IV.7 Evaluation metrics by all algorithms of classifiers trained with Leave-one-out Cross Validation and with ECG and RIP data sets. 78
IV.8 Evaluation metrics by all algorithms of classifiers trained with Leave-one-out Cross Validation and with EMG and ECG data sets. 79
IV.9 Evaluation metrics by all algorithms of classifiers trained with Leave-one-out Cross Validation and with EMG and RIP data sets. 80
IV.10 Evaluation metrics by all algorithms of classifiers trained with Leave-one-out Cross Validation and with EMG, ECG and RIP data sets. 81
IV.11 Evaluation metrics by all algorithms of classifiers trained with Leave-one-out Cross Validation and with EMG and ACC data sets. 82

List of Tables

2.1 Main factors that potentially contribute to the development of MSDs. 6

4.1 Sensors used in Data Acquisition 20
4.2 Time Series where the features were extract per signal type 39
4.3 The number of subjects and samples used in each training model considering the signal(s) type of the features applied. 40

5.1 Best features selected considering each type of signal. 42
5.2 Best features selected considering all signals. 43
5.3 Evaluation metrics of workload classification model based on All time-series processed. 45
5.4 Evaluation metrics of workload classification model based on ACC's time-series processed. 45
5.5 Evaluation metrics of workload classification model based on RIP's time-series processed. 46
5.6 Evaluation metrics of workload classification model based on EMG's time-series processed. 46
5.7 Evaluation metrics of workload classification model based on ECG's time-series processed. 46

6.1 Results of studies that used classifiers to detect fatigue, also having as experimental Manual Material Handling task. 50

III.1 Features selected considering each type of signal 71

Acronyms

ABD Abdominal *(pp. 11, 12)*
ACC Accelerometer *(pp. xii, 12, 16, 19, 20, 22, 32, 34, 38, 41, 43, 45, 47, 48)*
AdaBoost Adaptive Boost *(pp. 13, 14, 41, 45)*
ANC Automatic Nervous System *(p. 10)*
AV Atrioventricular *(p. 10)*

CVL Cardiovascular Load *(pp. 11, 16, 32, 43)*
CVS Cardiovascular Strain *(pp. 11, 16, 32)*

ECG Electrocardiogram *(pp. xii, 2, 10, 11, 19–23, 27, 28, 30–32, 34, 38, 41, 43, 45–49)*
EEG Electroencephalogram *(p. 17)*
EMD Empirical Mode Decomposition *(p. 33)*
EMG Electromyography *(pp. xii, 2, 8, 12, 17, 19–22, 32–35, 38, 41, 43, 45–48)*
EOG Electrooculography *(p. 17)*
EU-OSHA European Agency for Safety and Health at Work *(pp. 1, 6)*

FPCEUP Faculty of Psychology and Educational Sciences of the University of Porto *(p. 2)*

HR Heart Rate *(pp. 10, 11, 16, 17, 43, 48, 49)*
HRV Heart Rate Variability *(pp. 10, 11, 16, 23, 31, 32, 48, 49)*

IEC International Electrotechnical Commission *(p. 15)*
ILO International Labor Organization *(pp. 1, 4)*
IMU Inertial Measurement Unit *(pp. 16, 17)*
INE Portugal's National Institute of Statistics *(p. 1)*
ISO International Organisation for Standardisation Directives *(p. 15)*

KNN K-Nearest Neighbours *(pp. 13, 48)*

ML Machine Learning *(pp. 13, 45)*
MMH Manual Material Handling *(pp. xii, 15, 17, 49, 50)*
MSD Musculoskeletal Disorder *(p. 7)*
MSDs Musculoskeletal Disorders *(pp. xii, 4–6)*
MVC Maximum Muscle Contraction *(pp. 12, 21, 33, 48)*

NOVA SST NOVA School of Sciences and Technology *(p. 2)*

RC Rib Cage *(pp. 11, 12)*
RHR Relative Heart Rate *(pp. 11, 16, 32)*
RIP Respiratory inductance plethysmography *(pp. xii, 2, 11, 17, 19–22, 32–34, 38, 41, 43, 45–48)*
RMS Root Mean Square *(p. 28)*
RNER Removed Noise Energy Ratio *(p. 27)*
ROC Receiver Operating Characteristic *(pp. 39, 45)*

SA Sinoatrial *(p. 10)*
SE Shannon Energy *(p. 28)*
SEE Shannon Energy Envelope *(pp. 28, 29)*
SSM Self-Similarity Matrix *(pp. 13, 35, 47, 48)*
SVM Support Vector Machine *(pp. 13, 41, 45)*

TSFEL Time Series Feature Extraction Library *(p. 38)*

WHO World Health Organisation *(p. 4)*
WRMD work-related musculoskeletal disorder *(p. 1)*
WRMDs work-related musculoskeletal disorders *(pp. 1, 4, 5, 9, 15)*

1

Introduction

1.1 Context and Motivation

In the manufacturing sector, the operators are often exposed to repetitive and demanding tasks, such as manual material handling over long periods. In addition to the workers' incorrect postures, these conditions result in the appearance of occupational diseases as work-related musculoskeletal disorders (WRMDs). Although with the increasing development of industry 4.0, the labour effort of the workers has decreased significantly, the occurrence of these diseases continues to exist.

The WRMD represent 60% of the total occupational diseases, according to the European Agency for Safety and Health at Work (EU-OSHA) and may contribute to the development of chronic diseases [2]. All this has repercussions, not only in the execution of the work but especially in everyday life of the worker and thus suppresses the quality of life [3].

The industry sector is one of those with the most significant economic impact. According to Portugal's National Institute of Statistics (INE) and International Labor Organization (ILO), about 23% of the active population, in the last few years, works in the industrial field in Portugal and also worldwide [4, 5]. The number of people working in the industry is significant, so there is a high need to prevent these diseases in this sector.

Legislation and recommendations have been applied at the labour level to reduce as much as possible the number of individuals with these health problems [6]. Limitations in load weight [7], repetition frequency [8], and the definition of pauses or task rotation have been applied [9].

Despite the weight reduction, repetitive manual material handling over a specific time can lead to fatigue and discomfort for an operator [10]. Depending on age, gender, physical condition or even health, people have different resistance to work and with this, do not enter at the same time into fatigue [11]. Applying the same schedule to all workers has limitations in combating injuries. An attempt to customise these intervals has been made. However, the use of self-reports or observational levels is very subjective and may not be effective [12].

Physical fatigue occurs in different body systems, and thus different physiological

responses may occur in the practice of prolonged physical activity [13]. Therefore, detecting different physiological signals can help understand when a person should stop performing a specific task.

The project "OPERATOR - Digital Transformation in Industry with a Focus on the Operator 4.0" [14], among witch this thesis is conducted, intends to develop technology to ensure that the industry has high productivity and to protect the welfare of workers in this area.

The OPERATOR project has its origins in a consortium by Zenithwings, the Fraunhofer Portugal Research Association, the Faculty of Psychology and Educational Sciences of the University of Porto (FPCEUP), the NOVA School of Sciences and Technology (NOVA SST), Volkswagen Autoeuropa, NST Apparel (Europe) Lda., Controlconsul and the University of Minho and MIT - Institute for Medical Engineering and Science [14].

The creation of a device, capable of simultaneously acquiring signals from different body sources and through artificial intelligence classifiers, can be an asset to the objectives of this project. By implementing this technology, it will be possible to understand how the body adapts to the excess workload and warn of the danger in advance.

At the level of muscle fatigue, for classification of more local fatigue, the median of frequency and also the amplitude of the EMG signal have been referred to as the metrics that most evidence fatigue during a repetitive task [15, 16].

1.2 Objectives

In order to avoid accumulated fatigue and possible injury sequences, this thesis focused on trying to understand the behaviour of several individuals' physiological responses to determine fatigue.

The acquisition of various physiological signals during a given task can allow an adaptation of the duration of a work task in real-time in a personalized, objective way. With this, a repetitive task of manual material handling was performed interspersed with exercises with weights to cause maximum muscle contraction and to ensure that the person entered into fatigue.

This classification was based on the processing of various physiological signals acquired from several individuals, such as Electrocardiogram (ECG), Electromyography (EMG), Respiratory inductance plethysmography (RIP) signals and other signals related to movement, during the tasks. Through these acquired and processed signals, it was possible:

- Understand how different biosignals change over time when the individual is physical fatigued;
- Analyse and select the best metrics able to target fatigue state and reference;
- Study and select the workload classification models.

By understanding which features contribute most to the identification of fatigue, future studies may converge on these metrics, better perceiving their relationship with the physiological response to overwork. It then allows identifying more accurately when a worker goes into fatigue, thus preventing musculoskeletal problems that cause pain and abstinence from work.

1.3 Document Structure

This thesis is divided into seven chapters. The contextualisation, motivation and objectives are referred to in this chapter. In chapter 2, theoretical concepts are addressed at the level of Occupational Health, Biosignals and also at the level of Machine Learning. State of the art presented in chapter 3 shows the solutions and attempts regarding the prevention of muscular occupational diseases, focusing on the tasks of Handling material manual. Chapter 4 describes how the data acquisition, processing, segmentation and development of fatigue classification models were made. Chapter 5 presents the best-selected Features and the results of the various classification models. Chapter 6 presents the discussion, and finally, the last chapter represents the conclusions and future work. In the annexes are figures and tables with additional study information. Annexes I, II, III and IV correspond to complementary data from R peaks detection, segmentation, features selection and model development respectively.

2

Theoretical Background

2.1 Occupational Health

The World Health Organisation (WHO) characterise *occupational health* as a matter of public health. This matter aims to maintain and promote not only the physical but also mental and social well-being of workers in any occupation by improving working conditions and environment and promoting worker's safety, health, and their ability to work [17].

According to the ILO, about 2.78 million deaths per year are due to work-related accidents or diseases of which 2.4 million are disease-related [18]. These numbers reflect the need to strengthen and increase research and technology in occupational health, more specifically in the prevention of occupational disorders.

In the world of industry, workers are exposed to both physical and mental health risks, including injuries, accidents and illnesses [19]. In this area, depending on their type of assignment, workers tend to suffer, mostly from accidents related to fire or hot substances, falls or even electrocution. They may also suffer hearing loss or Musculoskeletal Disorders (MSDs) due to, for example, repetitive movement with microtasks [19].

According to the Protocol of 2002 to the Occupational Safety and Health Convention at Geneva in 1981, the concept of *occupational disease* "covers any disease contracted as a result of an exposure to risk factors arising from work activity" [20].

Among the various work-related diseases, as can be seen in Figure 2.1, MSDs are the most common problem, representing 60% of reports, followed by major mental disorders such as stress or depression and other disorders such as headache, eyestrain and also cardiovascular and pulmonary disorders [3]. Here the MSDs will emphasized to try to understand how to avoid them.

2.1.1 Work-related Musculoskeletal Disorders (WRMDs)

The *work-related musculoskeletal disorders* are diseases related to the soft tissues and their surrounding structures, such as muscles, bones, tendons, joints, blood vessels, and peripheral nerves, leading to trouble in movement [21]. These conditions are caused by occupational overstrain, exclude damage from acute or instant accidental events such as

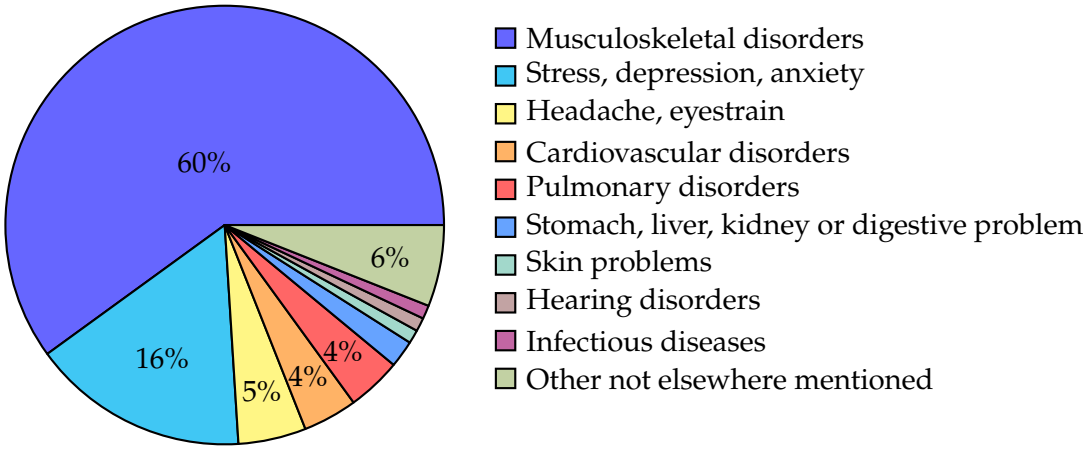

Figure 2.1: Percentage of workers from different job sectors, reporting an occupational health problem, by type of problem, according by Eurostat, Labour Force Survey in EU-27, 2013 [3].

falls or splits, and are made worse by the work and the way it is done. The WRMDs are characterised as painful, which causes discomfort not only in the workplace but also in the person's quality of life and even decreases the person's ability to make an active contribution to society [22–25].

Most of these diseases occur mainly in the back (representing 43% of MSDs), in the upper limbs (representing 41% of MSDs) and with less impact in the lower limbs (with 29%). In Figure 2.2 the primary musculoskeletal disorders related to the workplace can be seen.

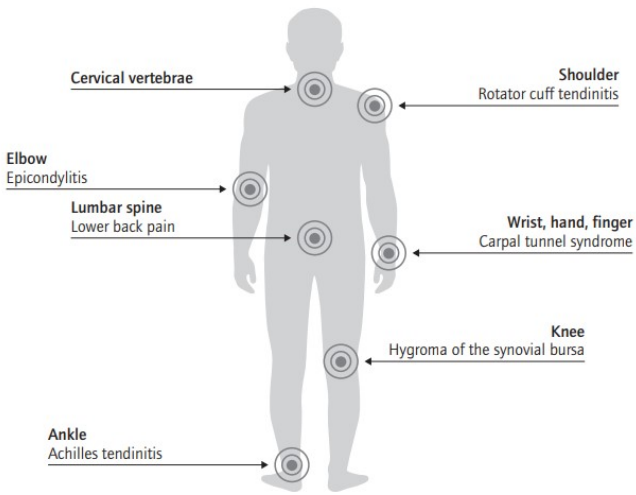

Figure 2.2: The main work-related musculoskeletal disorders [23]

2.1.1.1 Risks factors for musculoskeletal disorders

The cause of a musculoskeletal disease is often associated with work because the worker is exposed to various risk factors such as tiring and/or painful positions, repetitive work,

carrying or moving heavy loads and also, in some specific jobs are exposed to vibration from machinery [25].

Although these working conditions have an enormous weight in the development of these occupational resources, the physical condition, age, and clinical and mental state of the individual also contribute to the development of MSDs [26].

The risk exposure can be related to the amount of work in a short time, such as a day or a week, but can relate to cumulative exposure to risk for several years [27]. Table 2.1 describes the various factors contributing to developing these diseases. In this table, the risks described may be associated with various work activities in the manufacturing sector.

Table 2.1: Main factors that potentially contribute to the development of MSDs more precisely related to work, table based on European Agency for Safety and Health at Work. [26, 28]

Occupational Factors	
Biomechanical and environmental	Lifting, carrying, pushing or pulling loads or using tools; Repetitive or forceful movements; Prolonged physical effort; Awkward and static postures; Hand-arm or whole-body vibration; Cold or excessive heat.
Organisational	Long working hours; Lack of breaks or opportunities to change working postures; Work at a fast pace/working to tight deadlines; Repetitive and monotonous work.
Psychosocial	Lack of control over tasks and workload analysed; Unclear/conflicting roles; Heavy mental / cognitive workload; Lack of support from colleagues and/or supervisor.
Extra-Occupational Factors	
Individual Factors	Age and Gender; Genetic predispositions; Obesity/overweight; Pregnancy; Medical and surgical history.
Extra-Occupational Activities	Heavy or prolonged activities such as exercise, DIY activities or gardening.

This research work was intended to understand the workload, analysing the various risk factors described at the biomechanical level through electrophysiological signals. In the experiential task of the search, through repeating lifting and loading of a box for a long time without a break, one can then analyse physiologically, taking into account age and gender and physical activity of the person, how the body reacts to exposure to occupational risk factors such as [29]:

- Lifting, carrying, pushing or pulling loads
- Repetitive and monotonous movements
- Prolonged physical effort
- Lack of breaks

Epidemiological studies have associated these risk factors mainly with the risk of shoulder tendinitis and low back disorders [22]. For this cause, it is necessary to analyse the ergonomics of the movements related to the protocol analysed here to understand how the body reacts to the long period to exercise the movement of lifting and load of weights.

2.1.2 Ergonomics

2.1.2.1 Ergonomics Definition

Ergonomics is defined as "the science of work" [30] that studies people in their occupational work and aims to solve and modify the way the work is performed in order to provide security to the individual eliminating possible discomfort and risk of accidents or injuries due to the practice of particular labour activity, thus ensuring good occupational health. This science is divided into three essential themes to analyze in work: the physical, cognitive levels of the individual and the organisational level. In this study, we will emphasise physical Ergonomics; more precisely, in manual material handling, an MSD risk factor which increases its severity with repetitive activity over a long period without breaks [31].

Physical ergonomics aims to analyse anatomical, anthropometric, physiological and human biomechanical characteristics related to physical activity [30, 31].

2.1.2.2 Manual Material Handling Ergonomics in Industry

Manual material handling requires operators to lift, lower, push, pull, load, throw, move, and hold objects or equipment that are often heavy and carry the risk of injury and pain [32].

This section focus on ergonomics at the level of lifting weights with low back inclination and movement of the upper limbs without bending the back or flexion of the lower limbs to understand the workload on the upper body without counting on awkward positions.

The Figure 2.3 represent the manual material handling by steps, in order to represent the varied movements of the shoulder. The task begins in the resting position (A0); the individual tilts slightly to low back and slightly flexes the shoulder (A1) to grab the box (A2) to lift the box to the top shelf (UP) is made an abduction of the shoulder (A3) followed by horizontal adduction (A4) returning to the original position (A5, B0). To lower the box, the shoulder is flexed to catch (B1) the box following an extension (B2), placing the box on

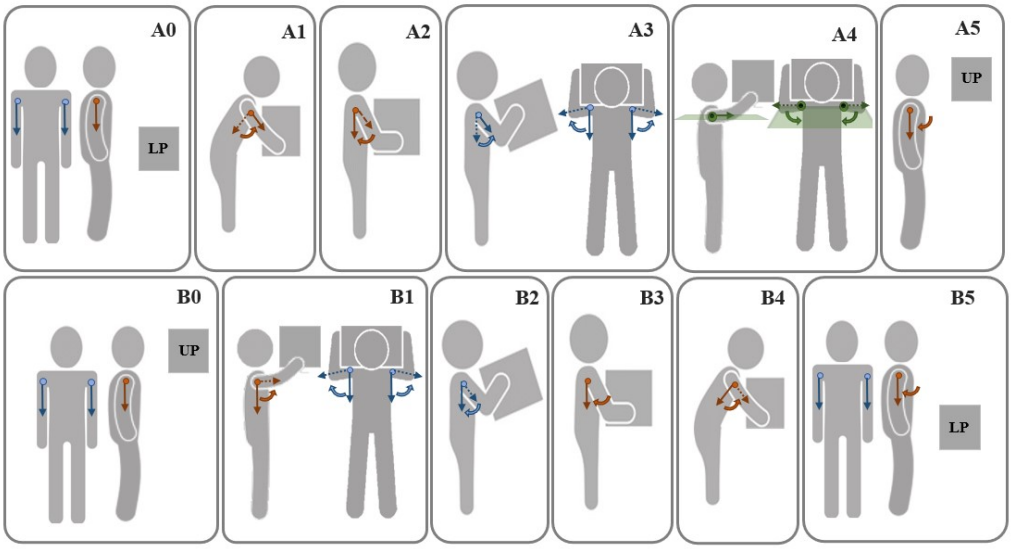

Figure 2.3: Illustration of the shoulder rotation in task under study, by temporal step order. The rotation's description is done at the level of the sagittal plane (arrows in orange), the frontal plane (blue arrows) and the transversal plane (green arrows). The blunt arrows reflect the final position at rest, and the dashed represent the initial position. Figures with **A** represent the lifting of the box and with a **B** represent the box's lowering.

the bottom shelf (B4) in the same way as in step A1 returning to the initial step (B5=A0). Concerning the movements of the elbows, there is a predominance of bending in the lifting and extension when lowering the box. To understand how these movements are made, it is necessary to realize the ergonomics of the shoulder and elbow joints. Although this task also has repercussions on the low back, it is on the shoulders and elbows that this thesis was focused on since the muscles of this joint make more effort and, therefore, more quickly go into fatigue

During the performance of the manual material handling, several combined muscles allow various movements with the arms, making it possible to pick up the box. In addition to the agonist, the primary muscle responsible for a particular movement, the antagonists maintain the arm's stability, thus playing an essential role in carrying a weight with the hands. In addition to these, synergist muscles can either assist the agonist or help as the matter fixes its origin, the end of the muscle where they are stuck [33].

For the study, two muscles were selected: *Bicep Brachii* and *Upper Trapezius*, because they work most in the task of picking up the box, in different specific movements (as agonist, antagonist or synergist) and also because they are the most superficial, that is, easier to detect non-invasive signal of EMG [34].

2.1.2.3 Bicep Brachii Muscle

The Bicep Brachii lies at the front of the arm; it is composed of two origins, the short head, located in the scapula's coracoid process, and the long head, located in the scapula's supraglenoid tubercle. It is inserted into the radial tuberosity. There are responsible for the

movements of the glenohumeral joint and the elbow joint. In the shoulder joint, this muscle promotes arm abduction (by both heads) and arm flexion (by the short head). In the elbow joint, this bicep accomplishes forearm supination and flexion. The musculocutaneous nerves also innervate this muscle, from the vertebrae C5 and C6 to the muscular bellies. [35].

2.1.2.4 Upper Trapezius Muscle

The upper trapezius muscle is located in the posterior neck and back and is the trapezius' smallest area. It originates at the base of the neck and is inserted into the clavicle (lateral third) and scapula (acromion and spine). This muscle is responsible for various neck and scapula movements such as shrugging the shoulders, shoulder abduction, and flexion [34, 36, 37].

2.1.3 Fatigue Mechanisms

As a worker exerts a particular workload, the body will physiologically adapt to meet the required task. When physical activity is done for a particular time without pause, the body can no longer respond, reducing the work output; that is, the body enters into fatigue. Although fatigue can be seen at the subjective level, with the mental decrease concentration and motivation or at the objective level, through the reduction of work output, it is the physiological fatigue that interests here, it means physically exhaustion [38, 39].

Physical fatigue is associated with external manifestations such as the inability to maintain the desired strength result, muscle tremors and localised pain [38, 40], which is why it is essential to detect when work comes to this physical condition to prevent WRMDs.

Neuromuscular fatigue may be due to peripheral or central fatigue. Central fatigue occurs at the central nervous system level and is a physiological occurrence detected over days, months. In peripheral fatigue occurs between the medulla (for the muscles of the upper limbs, in the cervical vertebrae) and the skeletal muscle it causes and is due to metabolic or ionic disturbances, occurring more quickly [41].

During a particular task, hemodynamic and cardiovascular changes occur to maintain the amount of oxygen adequate to the metabolic needs of skeletal muscles performing functions [42]. Physical activity causes changes in the cardiovascular system's *Efferent Sympathetic* and *Parasympathetic Outputs*. These changes in the autonomic system will cause variations in the *heart rate*, *arterial blood pressure* and *blood flow distribution* [**TheExercisePressoReflex**]. In addition to the pulmonary level, insufficient oxygen and excess CO_2 will cause increased ventilation, decreased pulmonary resistance and dead physical spaces. All these adaptations aim to improve the *Ventilation to Perfusion Ratio (V/Q)* to ensure an adequate gas exchange [44].

2.2 Biosignals

2.2.1 Electrocardiogram

Blood circulation from the veins to the auricula, to the ventricle and then to the arteries is coordinated through a conductive electrical system. This system begins in the Sinoatrial (SA) node that conducts impulses across the Atrioventricular (AV) node and then to the rest of the pacemaker cells of the ventricle in order to contract it [45]. In normal conditions, the SA node generates an electrical stimulus regularly, 60 to 100 times per minute [46].

In the typical ECG signal, as shown at the bottom of Figure 2.4, it is possible to observe five different types of waves during a cardiac cycle. The P wave corresponds to the depolarization of the muscles of the upper cavities (atrial myocardium) that causes the displacement of blood from the auricles to the ventricles. The set of QRS waves (QRS-Complex) reflects the depolarization of the myocardium of the lower cavities (Ventricles), where ventricular contraction is initiated, allowing blood to flow from the blood to the entire body or lungs. The T wave corresponds to the repolarization of the ventricular muscles. The auricles' repolarization coincides with the QRS complex, causing it where it is not observed. The shape of the wave of an impulse and time intervals between various peaks give information about the possible disease and other information [45].

The superficial ECG signal detector comprises negative (-) and positive (+) electrodes placed on the fur and an amplifier. Through these electrodes, it is possible to detect when depolarisation of pacemaker cells occurs. If the direction of propagation of depolarisation occurs in the electrodes from negative to positive, the detected current will be positive; otherwise, it will be negative [45]. In this study, we used a thoracic strap to measure the ECG, where a measurement is made just in the horizontal direction, also known as *Lead I*, according to the *Einthoven Lead System* [47], that typically electrodes are used in the arms.

In Figure 2.4, it can be seen how the system of electrical conduction of the heart is constituted and how this conduction spreads over time. In this image, one can also see a typical ECG signal acquired by electrodes (in the direction of *Lead I*) that is characterized by being a set of currents caused by the depolarization of different parts of the heart that occur in the direction of Lead [49].

2.2.1.1 Heart Rate Variability

The Heart Rate (HR) corresponds to the number of heartbeats for a particular time, usually for minutes, and how the time intervals oscillate between adjacent heartbeats corresponds to the Heart Rate Variability (HRV). The HR and HRV can be calculated by electrocardiogram by time between peaks R [48].

The HRV reflects the regulation of the functioning of the heart through the Automatic Nervous System (ANC) that adapts to the body's environmental and psychological challenges, such as activities of high physical intensity. The HRV is a consequence of body

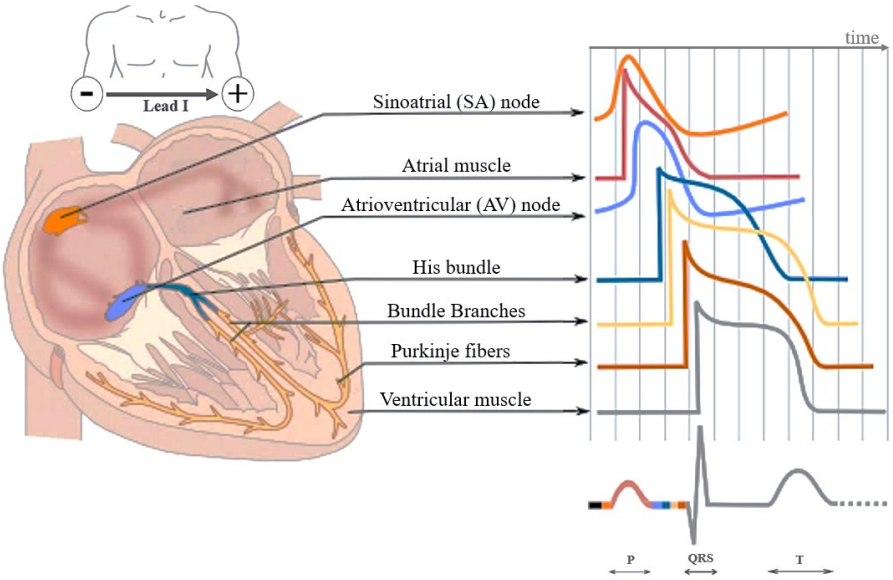

Figure 2.4: The Figure shows how the Cardiac Conduction System is constituted and how the acquired electrical signal affects each part of the system. Below is the set of all these signals that form the ECG signal corresponding to what is observed by the device. The image is based on [48].

regulation through the change of blood pressure, gas exchange, gut, heart, and vascular tone (blood vessels' diameter) [48].

Healthy biological systems exhibit spatial and temporal complexity and non-linearly (mathematical chaos); however, diseases can involve either a loss or increase of complexity [48]. People's age, gender and physical condition influence the form of HRV [50].

It is known that when physical exercise increases, in addition to the Heart Rate increase also, the stroke and load volume increase. With this, in addition to HRV, we also study other metrics related to the variation of heart frequency. These metrics can be analyzed to understand how the cardiovascular system evolves, taking into account the age of individuals and the HR at rest. Cardiovascular Load (CVL), Cardiovascular Strain (CVS) and Relative Heart Rate (RHR) can be highlighted [51–54].

2.2.2 Respiratory Inductance Plethysmography

Respiratory inductance plethysmography (RIP) is a noninvasive method that intends to analyse the evaluation of the respiratory system. It consists of two elastic bands in the abdominal and chest areas that measure the respiratory movements Rib Cage (RC) and Abdominal (ABD), respectively [55].

With the data collected by the two bands, having the sinusoidal signal of respiratory movement, it is possible to analyze several relationships between these signals, as is the case of *Phase Angle* (ϕ), the *Percent ribcage (RC%)*, and also the correlation between the two signals. The *Phase Angle* (ϕ) corresponds to the difference in Phase between the two signals, and the *RC%* is intended to measure the contribution of the Rib Cage in relation

to the *Tidial Volume (Vt)*, here seen as the sum of the two signals RC and ABD [56]. Here this sensor was used to understand how respiration develop when a person fatigues.

2.2.3 Electromyography

Electromyography (EMG) measures electrical activity caused by neuromuscular cells that promote muscle contraction. The surface method of EMG was used here, where sensors can measure the muscles' impulses through electrodes [57].

Through this type of signal, it is possible to observe the reduction of maximum strength exercised by the muscle when it enters into fatigue. In order to detect fatigue, a relationship between the EMG measured during a task and the maximal force of contraction produced before a task is called the Maximum Muscle Contraction (MVC) and must be greater than or equal to the power produced in the exercise. The person exercises his maximum strength in a position and angle strategically where it is known that the muscle contracts as much as possible [58].

2.2.4 Accelerometer

The Accelerometer (ACC) is a sensor that detects linear accelerations (along x, y and z) and is sensitive to the gravitic field. Through these sensors, it is possible to study and identify gestures [59].

Through the Cartesian coordinates, it is possible to remove information from different rotations of the pulse to perceive possible changes in the rotation of the pulse for the same type of task. On the wrist, the movements that can be analyzed are the *Pitch*, *Roll* and *Yaw*. The *Yaw* corresponds to the same axis of motion as the turf, so it only changes in *Pitch* and *Roll* interested in this studying. In the Figure 2.5 are represented the movements under study [60].

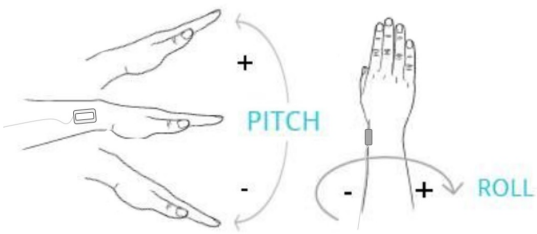

Figure 2.5: Types of rotation were studied in the wrist through an accelerometer [60].

2.3 Machine Learning

Machine Learning (ML) is a subclass of artificial intelligence and aims to make predictions through statistical models; it means to guess something that will happen. Three major categories of predictive models within ML are *Supervised*, *Unsupervised* and *Reinforcement* models [61]. The *Supervised* models are trained with a labelled dataset, unlike *Unsupervised* models, in which the model finds unidentifiable patterns by eye. *Reinforcement* models train the model by trial and error [62].

2.3.1 Supervised Learning Models

Although the exact moment the person goes into fatigue is not visible, using the Self-Similarity Matrix (SSM) method, the information used to train the models is already categorised, so only the *Supervised Machine Learning* models are used. For fatigue prediction, we used seven types of models:

- Decision Tree
- Random Forest,
- Support Vector Machine (SVM),
- Logistic Regression,
- Gaussian Naive Bayes,
- K-Nearest Neighbours (KNN) and,
- Adaptive Boost (AdaBoost)

The *Decision tree* is a popular supervised tool that intends to predict. Have a tree structure where the nodes represent the attributes, the branch represents the outcome, and the terminal nodes represent the class label [63].

The *Random Forest* is the meta-estimator that uses several sub-samples. For each sub-sample, obtain the Prediction of the Decision Tree, and in the end averages, the predictions to improve the classification [64].

Cortes and Vapnik proposed the *Support Vector Machine*. This algorithm defines boundaries in hyper-plan between the class types to predict. The space has the exact dimensions as the number of Features [65].

The *Logistic Regression* predicts binary classes by statistical method. This method uses the probability of occurrence of a binary event utilising a logistic function [66].

The *Gaussian Naive Bayes* is a *Naive Bayes* algorithm based on *Gaussian Distribution*. In this case, the prediction considers probabilities founded on the Bayes theorem. [67].

By calculating the distances, the *K-Nearest Neighbours* predicts the class through the k-nearest labelled values already known [68].

Freund and Schapire proposed the *Adaptive Boost*. It is based on the number of interactions where each result is compared with the true value. Only the data that was miscalculated is used in the next iteration. With this, it is possible to calculate an error function [69].

3

State of Art

The every day workload leads to occupational stress on a psychological, social and physical level. There are different stimuli depending on the type of work, but even low to moderate intensities can lead to occupational illness, namely when repetitive work is present. This fact is worrisome because the worker performance hisactivity under fatigue. Accumulated fatigue causes a decrease in effectiveness and increases the risks to the worker's health.

Since the beginning of the last century, there has been a need to find solutions for labour models focused on workers to increase the productivity and effectiveness of manufacturing workers. Lillian and Frank Gilbreth [70] pioneered this topic when they conducted a study in which they designed the movement of essential tasks to reduce repetitive movements that could affect workers' health.

Even before the industrial revolution, from the beginning of the 18th century, musculoskeletal diseases have been recognised as a consequence of occupational etiologic factors. However, only from the 1970s began to examine this fact, scientific literature on this subject [27].

To reduce WRMDs, international standards have been developed, such as the *International Organisation for Standardisation Directives (ISO)* and the *International Electrotechnical Commission (IEC)*, where, for example, at the level of lifting and carrying of certain loads (ISO: 11228-1:2003) are given limits of lifting and manual transport, depending on the weight, frequency and duration of the task [71]. In addition to this, the ISO has given advice on how the boxes should be loaded.

The *Health and Safety Executive* [8] has analysed tasks that require repetitive upper extremity movements and defined the task as low, medium or high level of risk. For this, risk classification was considered the duration of the task, the number of breaks and repetitions and the actual time to which a person is in a bad position.

To the repetitive Manual Material Handling (MMH), although there are load limits of intensity and frequency, is not sufficient to protect individuals from injury since the person over time becomes tired and even has musculoskeletal pains at the end of a working day

[7]. In this regard, several studies have proposed job rotation and breaks during a work shift [9]. This solution is an option indicated for the industry to try to prevent injuries and, with it possible work absences or even allowances; however, different people have physiological reactions and support different limits of weights and work intensities. These differences make different people get tired more or less quickly. For this reason, it is necessary to adapt the working time to each person.

Studies such as Jørgensen et al. [72] have demonstrated the importance of detailing fatigue, showing the reduction of endurance can lead people to load weights incorrectly, causing bad postures and consequent injuries. Detecting fatigue can thus be a good criterion for understanding when the person should rest.

Self-report questionnaires or scores classified as *Borg Scale* have been used to detect fatigue. Åhsberg et al. [12], Winwood et al. [73] and Michielsen et al. [74] have used fatigue scales that have taken into account lack of energy, physical exertion and discomfort, lack of motivation and also sleepiness [12]. Although this information can provide information on the level of the state of fatigue, they are subjective, they require time to fill out and their daily use is not practical.

Indices that characterise risk exposure have been used but are based on observational aspects. The *OCRA* proposed by Occhipnti [75] is a "concise index for assessing exposure to repetitive upper limb movements". The *NIOSH* equation [76] calculates exposure at the level of material handling tasks. The *Rapid Entire Body Assessment (REBA)* [77] was used for detection of postures and *RULA (Rapid Upper Limb Assessment)* was used to avoid work-related upper limb disorders [78].

Motion-detection wearables have been used for fatigue detection at work. *Motion Analysis System (MAS)*[79] is a human body digitalisation and concern productive and ergonomic analysis by evolving the joint angles and the bone postures. There are also studies based on ACC and Inertial Measurement Unit (IMU) data that aim to classify fatigue by taking into account the posture, the angle between joints [80] or even possible tremors caused by the strain.

As previously mentioned, long durations of repetitive movements cause the person to change their physiological responses as they get tired, forcing them to make too much effort for a simple task. Through the acquisition of physiological signals and movements, it is possible to perceive how the person changes these physiologic responses during work. Based on this, many studies analyse various physiologic signs to perceive the alterations at the local level, in this case, in the upper limb and at the general level.

Several studies have tried to perceive cardiovascular stress during a repetitive task for different classes of workers who require cyclical physical work. These studies used the metrics of RHR, HRV, HR, CVL and CVS in order to identify the accumulated fatigue of these workers [81–87]. The results of these studies demonstrated the impact these metrics have on fatigue determination.

At the level of respiratory stress, studies have used oxygen consumption to verify changes in aerobic capacity [88, 89]. Silva et al. [90] showed RIP correlation and Phase Synchronisation changes between baseline and fatigue. Aubier [91] proposed the detection of respiratory muscle fatigue.

At the level of muscle fatigue, for classification of more local fatigue, the median of frequency and also the amplitude of the EMG signal have been referred to as the metrics that most evidence fatigue during a repetitive task [15, 16].

Other physiological signs have also been used to detect fatigue. Mehta et al. [92] use skin temperature to understand if it relates to tiredness. At the neurological level was detect fatigue with the recurrence of Electroencephalogram (EEG) [93] and/or Electrooculography (EOG) [94].

Focus on repetitive MMH task, studies such as Darbandy et al. [95] and Nasirzadeh et al. [96] used HR to classify fatigue. Other studies, in addition to HR, still have other signs. Lambay et al. [97], and Maman et al. [29] added to the IMU classifiers located in various parts of the body in order to perceive the changes in movement, posture and cardiac activity. Umer et al. [98] and Aryal et al. [99] used a mixture of physiologic signals with the *Borg Scale.* Both studies were used in addition to HR, Temperature and the pain scale. The difference was that Lambay et al. [97] used signs of respiratory frequency, and Aryal et al. [99] used signs of EEG. Baghdadi et al. [100] got by just using the motion sensors.

In this study, it was intended to obtain a classifier based on more physiologic responses to obtain a more precise result and compare which physiologic response had the best results.

4

Methods and Materials

4.1 Method Schema

In order to try to understand what is the best way to obtain a workload classifier, this research work is divided into three main phases: *Data Acquisition*, *Pre-processing* and *Model Development and Validation*.

To train the classifier in Data Acquisition, it was necessary first to acquire physiological signals with known workload domain. The study was specifically study the load by lifting and transporting loads repeatedly. After the data had already been acquired, it was necessary to process them so that existing noise was removed and that it was possible to observe the signals of interest to be deeply analyzed later. Several features were extracted and selected from the various acquired signals, generating different design matrices to test in different types of Machine Learning Models that were validated with several evaluation metrics to determine the best model to classify the workload. The scheme presented in the Figure 4.1 shows how the thesis investigation proceeded.

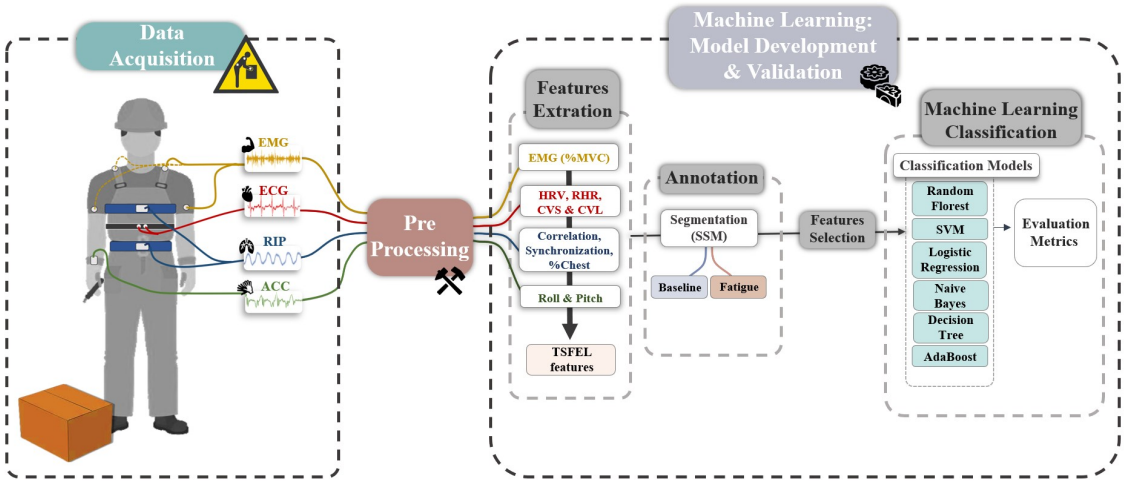

Figure 4.1: Thesis work' schema.

4.2 Data Acquisition

4.2.1 Participants

The experimental procedure was performed on 25 healthy participants (11 men and 14 women, age: 26.3 ± 8.0 yrs; height: 1.70 ± 0.09 m; mass: 61.89 ± 12.90 Kg) with different physical activity levels. The experimental activity focused on a repeated movement of taking a box from shelfs.

4.2.2 Task

The activity represented in Figure 4.2 was done cyclically over a long time in the same way as was explained in 2.1.2.2. Here, the top shelf (**UP**) has a height appropriate to the box's alignment with the individual's eyes. The bottom (**LP**) one is aligned with the tips of the individual's hands at rest without bending over.

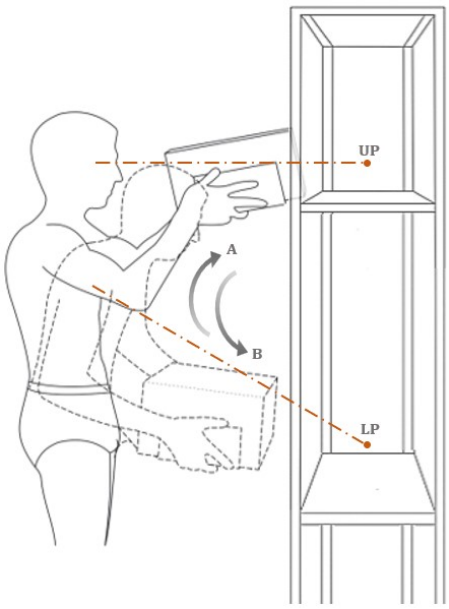

Figure 4.2: The layout of the manual handling of the box done on this study's acquisition date. Figure based on Oliveira et al.[101]. Movement **A** represents the box's lifting from the bottom shelf (**LP**-Lower position) to the top rack (**UP**-Upper position). With the counter movement, the **B** represents the box's lowering.

4.2.3 Materials

During the activity several physiological signals of ECG, EMG, RIP and ACC were acquired. To this end, the various sensors were placed before the activity began.

For ECG acquisition was used an elastic band connected to a bipolar sensor (Negative and Positive). Two elastic straps connected to an inductive sensor were used in the acquisition of the RIP signals. In the EMG Signal for brachial bicep and triceps, self-adhesive disposable Ag/AgCl electrodes were placed in the muscle with bipolar sensors, with an electrode

connected at the back of the elbow to serve as a ground to the electrical system. To ACC signals were used a tri-axial accelerometer.

All these sensors were connected to the same HUB, a device that collects the data and transfer all signals to his software at the same time, where the data can be saved. Table 4.1 describes the various sensors used and where they were placed, and in Figure 4.3, several sensors used for data acquisition can be seen. All these equipment are part of device *biosignalsplux* Kit [102].

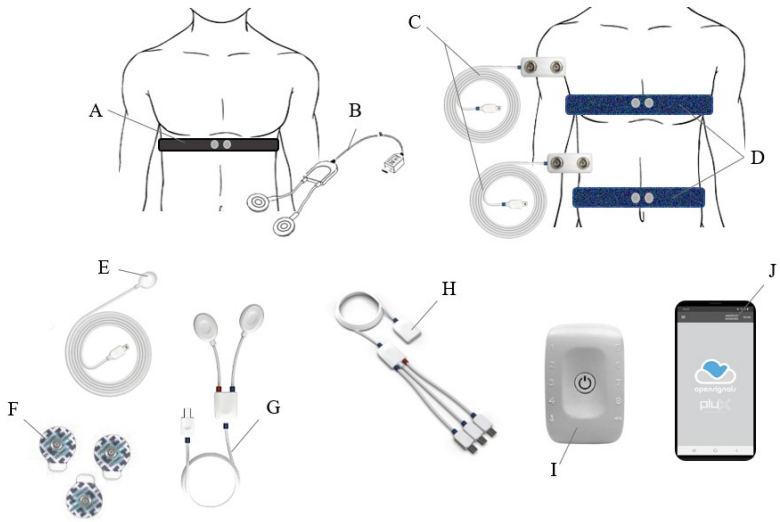

Figure 4.3: Devices used in data acquisition. **A** - ECG's chest strap, **B** - ECG's Bipolar Sensor, **C** - RIP's Inductive Sensors, **D** - RIP's chest strap, **E** - Ground Cable, **F**-self-adhesive disposable Ag/AgCl electrodes. **G** - tri-axial Accelerometer, **I**-Hub, **J**-App *opensignals* used to get data.

Type of sensors	Location on the body
ECG's Chest Strap 1 signal source	Chest Belt - below the outer bone
EMG's Electrodes 2 signal sources	Biceps - Dominant Arm Trapezius - Dominant Arm
RIP's Chest Strap 2 signal sources	Chest Belt - in hight bust line, below the armpit Abdominal Belt - in waist line, on the navel
ACC's sensors 1 signal source	Medial side of the right wrist

Table 4.1: Sensors used in Data Acquisition

4.2.4 Protocol

In a data acquisition protocol, after the individual read and signed the informed concention, some personal details began to be extracted as age, gender, height, weight, and exercise frequency. The ECG and RIP bands were placed, the skin was prepared to place the EMG electrodes, and the sensors and the hub were linked. Then, so that the box had 5.25% of the person's body, an estimate made of the usal weight carried by the workers, the weight of the box and the height of the shelf was adjusted to the height of the person.

After the sensor placement, a reference acquisition protocol was made; that is, signs were taken from the person standing at rest for 2 min to normalize later the ECG and MVC signal, where the person had to do four times the maximum contraction of each muscle, to normalize the EMG.

In protocol mainly, the person had to take the box from a bottom shelf (the height of the hands in anatomical position), put it on the shelf at eye height, and then put it back in the initial place, as can be seen in Figure 4.2, during 10 minutes, ending the first task. Two more tasks were done in the same way.

A fatigue protocol was applied to ensure the person entered more quickly in fatigue between tasks. That consisted of the person doing the flexion and extension of the arms with weights (used in the Gym) until the person could no more execute the exercise. In the Figure 4.4 is a schematic of the data acquisition procedure.

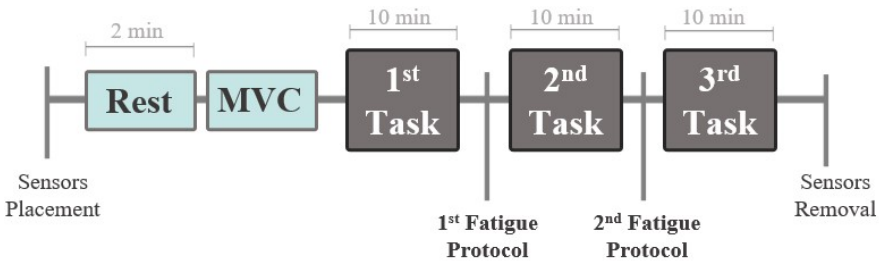

Figure 4.4: Data acquisition protocol's schema

4.3 Pre-processing

In order to be able to understand what data and variables to use in the fatigue classification model, it was necessary to process them, it means, to convert to original units, filter, analyse them and extract variables associated with them.

4.3.1 Transfer Function

When signals are acquired, all information transferred from the HUB to the software is in the same units, in Volts. It is necessary, first of all, to use a transfer function that varies from sensor to sensor to convert the system output data into the original input signal. In this acquisition, the data were obtained at 16 bits per sensor ($n = 16$) and with an operation voltage (VCC) value of $3V$ [103–106].

In the case of ECG and EMG, the *Transfer Function* equation is the same and is represented in the Equation 4.1 and 4.2 respectively, where the G_{ECG} is 1019 and G_{EMG} is 1000. The ADC represent the value sampled [104, 105].

$$ECG(mV) = \frac{(\frac{ADC}{2^n} - \frac{1}{2}).VCC}{G_{ECG}}.10^3 \quad (4.1) \qquad EMG(mV) = \frac{(\frac{ADC}{2^n} - \frac{1}{2}).VCC}{G_{EMG}}.10^3 \quad (4.2)$$

In the case of RIP, the input signal of each sensor is given in percentage (%), for it the Transfer function just takes into account the number of bits, as can be seen in the Equation 4.3 [106].

$$RIP(\%) = (\frac{ADC}{2^n} - \frac{1}{2}).100\% \quad (4.3)$$

For the ACC the sensor needs to be calibrated with a calibration protocol previously done (slow rotation in 360º) to know the min and the max of calibration (C_{min} and C_{max}), values needed to use in the transfer function represented in Equation 4.4 where the g represent the g-force [103].

$$ACC(g) = (\frac{ADC - C_{min}}{C_{max} - C_{min}}).2 - 1 \quad (4.4)$$

It should be added that the values that signals can acquire, in ECG vary from $[-1.47mV, +1.47mV]$, in the EMG is $[-1.5mV, +1.5mV]$, in the RIP is $[-50\%, +50\%]$ and finally in the ACC is $[-3.60g, +3.60g]$ [103–106].

4.3.2 ECG Signal Processing

In this work, the ECG signal acquisition was used mainly to analyse the HRV and related metrics, so the signal processing of this signal was focused on R peaks detection to calculate the interval of heart cycles - RR Intervals. Several studies, as was previously said, have shown that HRV varies significantly when the person is or is not fatigued; because of this, it was necessary to ensure that the processing of this metric was correctly done not to generate errors at the end of the classification, having thus occupied a large part of this investigation's time.

4.3.2.1 R Peaks Detection

Several existing models were tested to detect the R peaks, such as those found in the *NeuroKit* and *biosignalsnotebooks Python* packages [107, 108]. However, when visualising HRV graphs of different acquired signals, it was realised that the peaks were being poorly detected because many abrupt variations between RR Intervals were not natural in humans' heart rate. Heart rate varied from one beat to another at about 100 bpm and this was caused by the failure of one or more R peaks or false peak detection.

When testing different detection methods, different forms were observed with different heart rate means, as can be seen in Figure 4.5. In Figure 4.6, the problems in R peaks detection before mentioned can be seen. In the annexes, also the results of this detection in different signals and in different tested algorithms can be observed and confirm the poor detection of R peaks.

Because of these flaws in the detection, it was necessary to develop a new algorithm to ensure that as many R peaks as possible were detected. This algorithm was based on previously proposed methods for detecting R peaks, such as Manikandan et Soman [109] or Suboh et al. [110] models , and the QRS complex, such as Xu et Du model [111], which are based on the use of *Shannon Entropy* [109–111] and, in some models, followed by *Hilbert-Transformation* [109, 111] to detect these fiducial points. Here after several tests it was realized that the *Hilbert-Transformation* did not ensure a good R peak detection in the acquired signals and not resorting to this was obtained, since it also had a considerably unstable processing time. The method proposed is schematized in Figure 4.7. This method is divided into three parts, the first corresponds to the signal filtering, the second part corresponds to the QRS complex's detection and the last part is intended to detect the R peaks.

ECG Filtering

In order to observe the ECG signal converted to highlight the QRS complex, several band-pass filters with different frequency ranges were tested. In Figure 4.8, you can see the filtering of an ECG signal with reference band-pass filters. The first filter shown (Fig. 4.8-A) is used in the *Pantompkins* method, which uses a frequency range between 5-15Hz

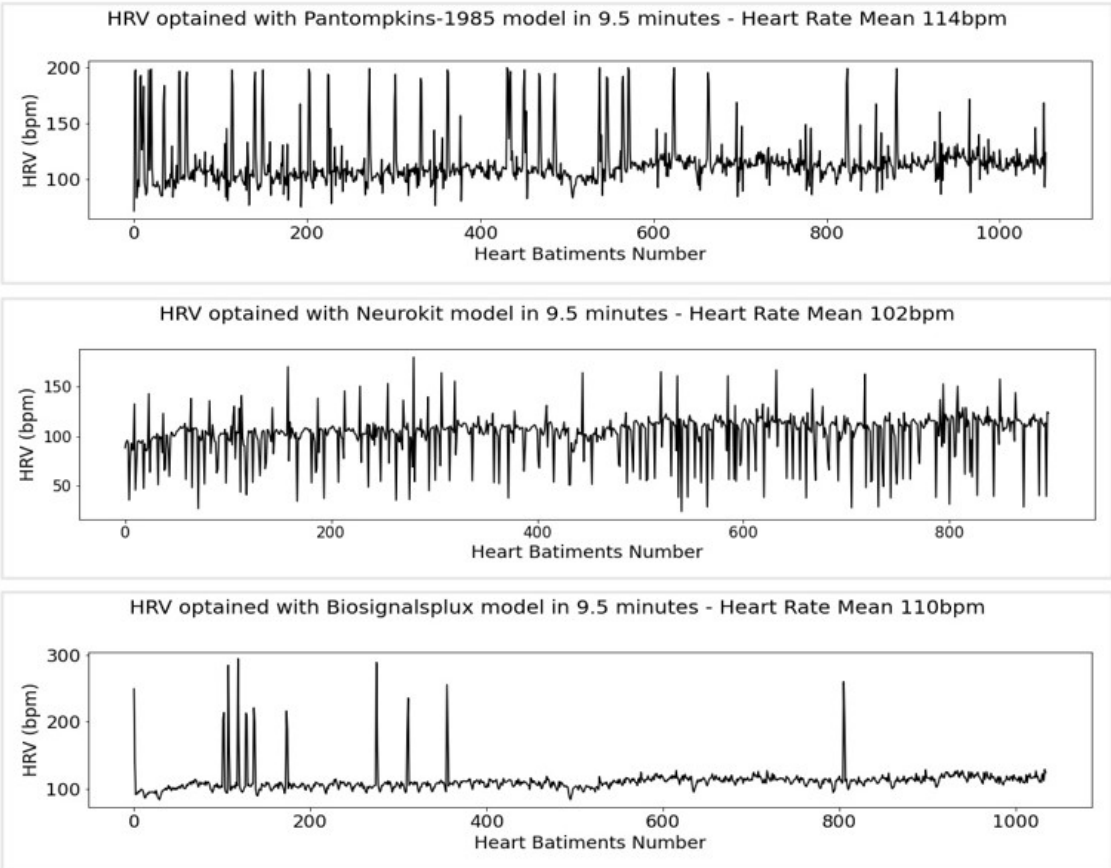

Figure 4.5: These three graphs represent the HRV of the same ECG signal calculated by some tested algorithms for R peak detection. The first signal corresponds to the signal acquired by the most classic model of peak detection R, the *Pantompkins-1985* model, the second corresponds to the proper model of the *NeuroKit*, and the third signal corresponds to the signal acquired through the model suggested by *Plux Wireless Biosignals* [107, 108]. The HRV has different shapes and sharp peaks that show flaws in RR intervals.

in third order [112] and the second filter (Fig. 4.8-B) is used in the methods where this model was based on [109, 111] and which use a fourth-order band-pass filter with a range of 6-18 Hz.

In addition to the bandpass filters also tested, different *Discrete Wavelet Transforms* were used as a form of filtering. Here the wavelets of the family *Daubechies* and *Symlet* were used since these wavelets families are more used in the filtering of ECG Signal [112–114]. From these families, we chose to use db2, db3, db4, and db10 of the *Daubechies* family and sym2, sym3, sym4, and sym5 of the *Symlet* family. In Figure 4.9, you can see the shape of these wavelets.

An existing function for python equivalent to the *modwt* function of Matlab [115, 116] was used for presenting a more significant signal cleanup than that existing in the *PyWavelets* software [117], selecting only the coefficients with a frequency range between 5.625-22.5

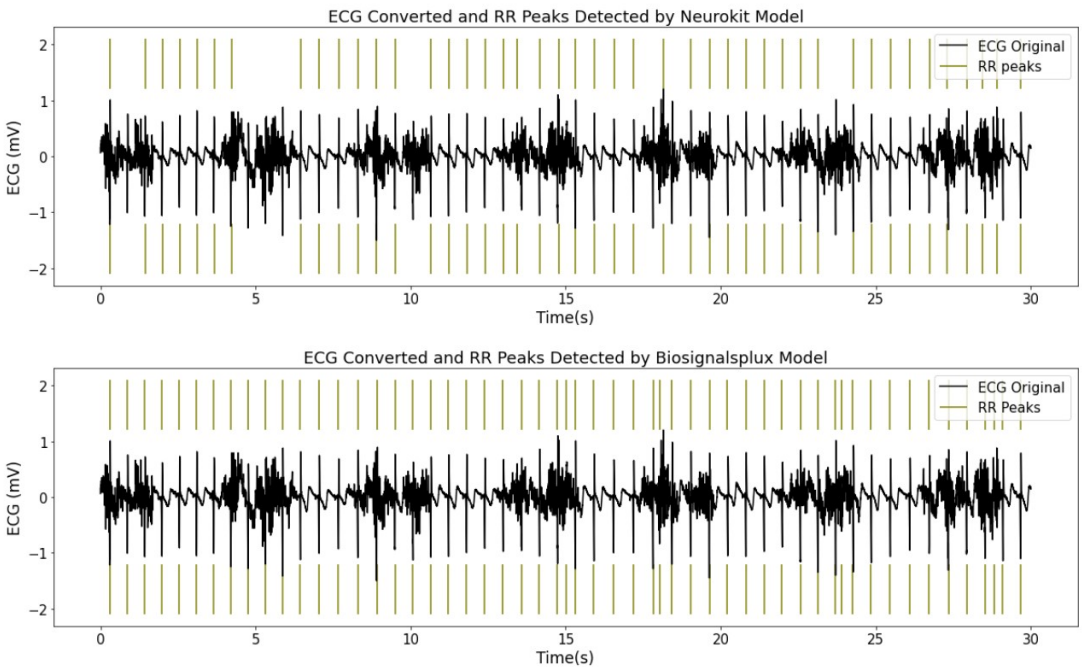

Figure 4.6: The same signal was used in these graphs to detect the R Peaks with different models. The graph above represents the detection with the *NeuroKit* model, and the bottom is the detection with the *Biosignalsnotebook* model. Here there are R peaks not detected and wrong peaks detected.

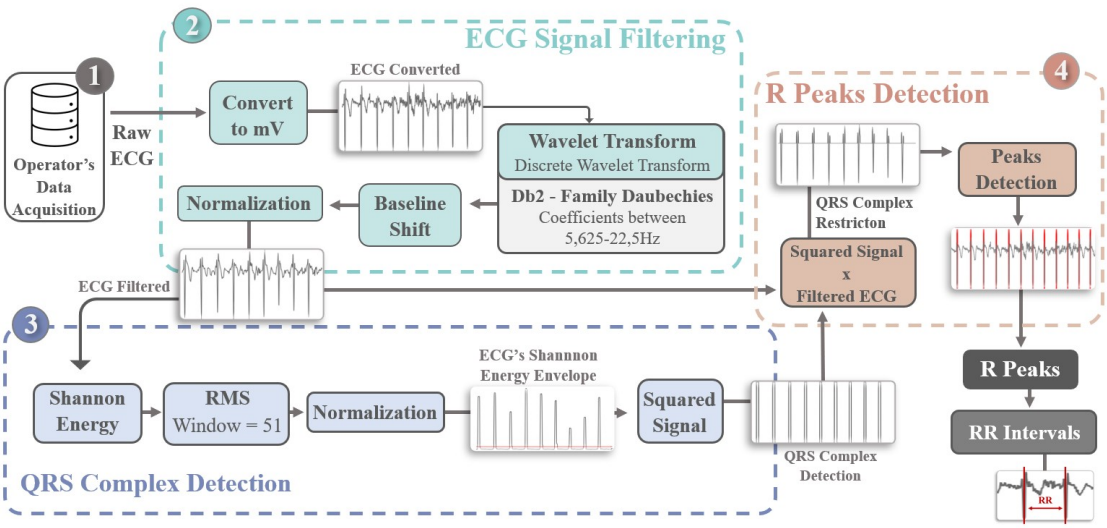

Figure 4.7: Schema of the proposed R peak detection algorithm.

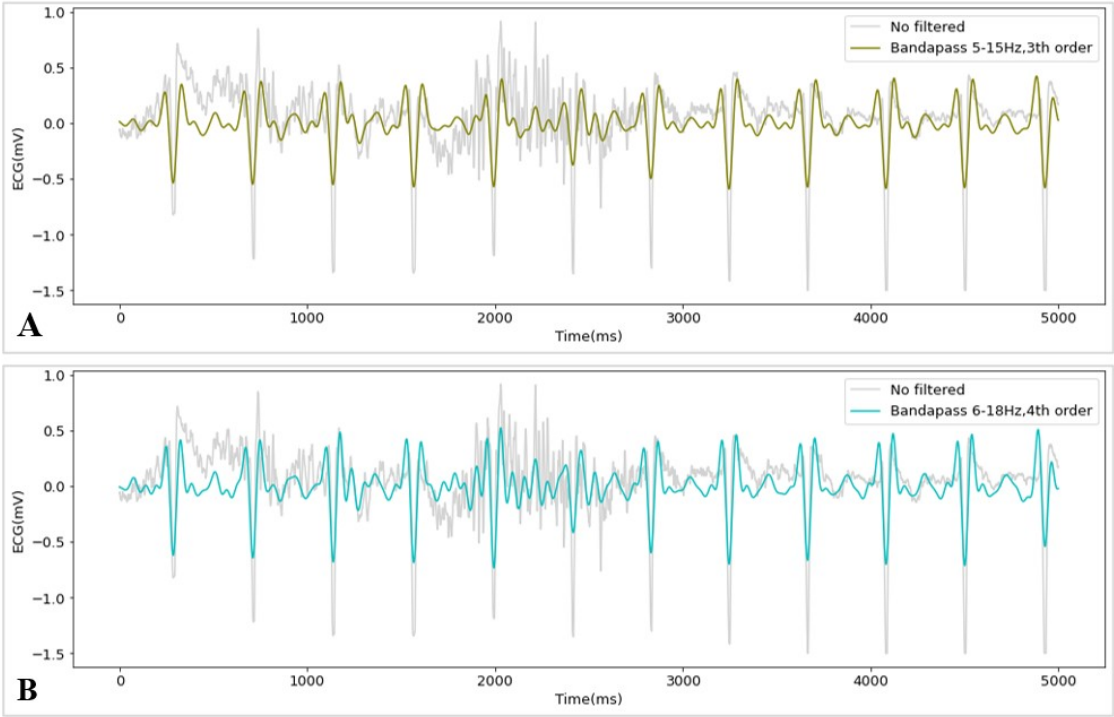

Figure 4.8: ECG original signal comparison with bandpass filters of different order frequency ranges. In image **A**, the filter is third order and has a frequency range between 5-15Hz, and in image **B**, the filter is fourth order between 6-18Hz. Both filters shown here are Butterworth digital Filter, and although they have managed to remove much of the noise, there is also much attenuation of the R peaks, making it difficult to distinguish between the QRS complex and the rest.

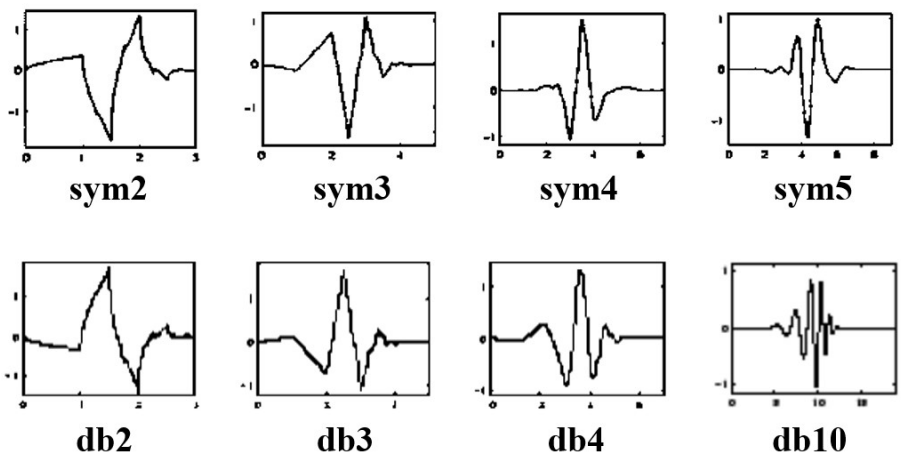

Figure 4.9: Shape of different Wavelets Transform tested

Hz was considered. The coefficients were chosen considering the range of frequencies closest to the band-pass referenced before.

When analysing the various wavelets in each signal to verify which one is better, it was found that, for different ECG signals, the wavelet to be applied would be different. Therefore, for the choice of the best wavelet, we tried to apply a more quantitative criterion instead of an analytic one.

For a good detection of QRS Complex, there must be a good discrepancy between the R peaks of the rest of the signal. Although R peaks usually have a much intenser signal strength than the other peaks, they have little impact on signal power because they represent only a tiny part of the signal samples. Knowing that, we can have better detection when the signal amplitude is as high as possible and there is less noise at the same time.

$$P_x = \frac{1}{N}\sum_{n=0}^{N-1} x[n]^2 = \frac{1}{N}.E_x \tag{4.5}$$

In the Equation 4.5, P_x and E_x represent respectively the power and energy of a signal x, N the signal's length and n the sample.

Thus, to choose which wavelet to apply, the amount of noise removed was used as a criteria, assuming that most of the information taken from the wavelet is noisy, but taking care to remove the minimum possible information from the QRS complex. The graph in Figure 4.10 shows different types of wavelets compared with the original ECG and the corresponding removed noise.

An energy ratio was used to determine the amount of noise removed, based on existing methods of ECG de-noising through Wavelet Energy, referenced by Zhang et al. [118]. This ratio was called Removed Noise Energy Ratio (RNER), a relation between the energy of the information removed and the original signal energy represented in Equation 4.6, where x represents the original signal, and s represents the filtered signal.

$$RNER(\%) = \left(\frac{\sum_{n=0}^{N-1}(s[n]-x[n])^2}{\sum_{n=0}^{N-1} x[n]^2}\right).100\% \tag{4.6}$$

Here, when analysing this ratio and the relation of the amplitude of the filtered signal with the original, it was possible, as shown in the graph of Figure 4.11, to choose the wavelet db2 as a filter, since it was the one that could get more noise and the one in which amplitude even increased compared with the original signal.

As we can see in the figure 4.12 the chosen wavelet gives greater projection to the QRS Complex than the band-pass filter 6-18Hz, referenced before. After choosing the wavelet db2 to filter the signal, baseline shift and normalisation were performed to ensure that the baseline of ECG signal was at $0mV$ and to avoid variations in signal amplitudes for

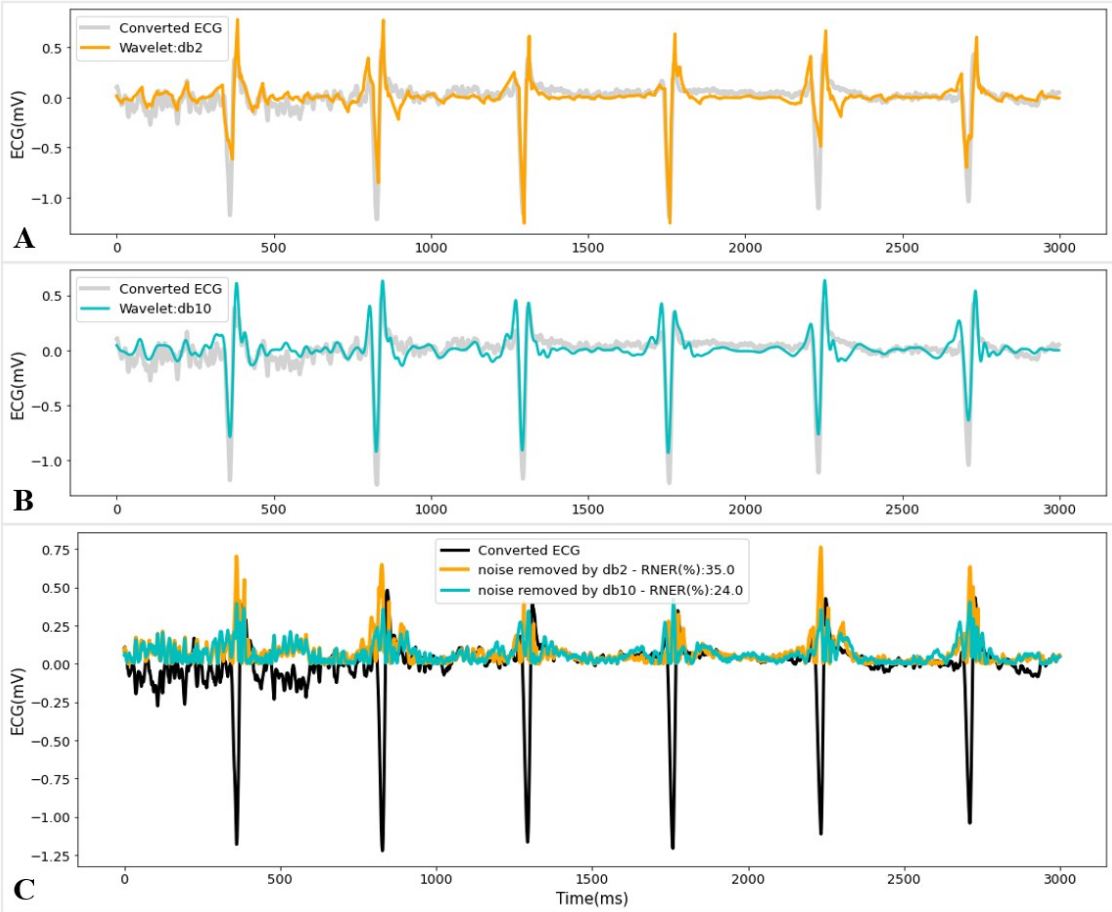

Figure 4.10: The same original ECG Signal compared with Wavelet db10 in image **A**, with Wavelet db2 in image **B** and with the absolute value of the removed noise signal of both filters in image **C**

better detection of R peaks. The normalisation made in this study is based on the Z-score of the integer signal, which is represented in the equation 4.7, where here x is the filtered signal, t is the filtered signal and $\bar{x}$ and σ represent the mean and the standard deviation respectively of the signal x.

$$t[n] = \frac{x[n] - \bar{x}}{\sigma} \tag{4.7}$$

QRS Complex Detection

The Shannon Energy Envelope (SEE) was used to detect the part of the QRS Complex. For this, the Shannon Energy (SE), s, of the filtered ECG signal, x, was calculated (Equation 4.8) [110], followed by a Moving Root Mean Square (RMS) filter, with a window of 61 samples (0.61 seconds) (Equation 4.9)[119].

$$SE[n] = -x[n]^2.log(x[n]^2) \tag{4.8}$$

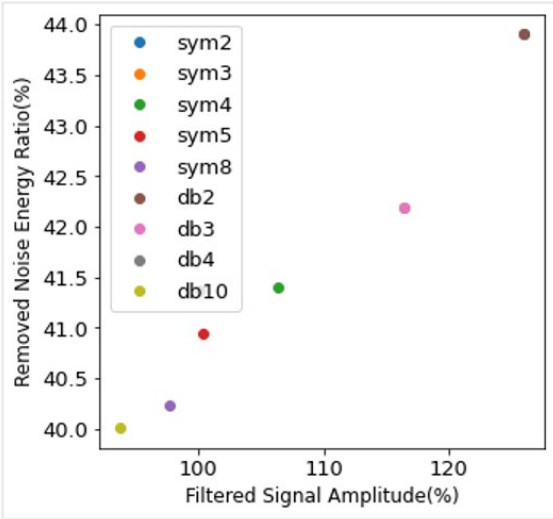

Figure 4.11: This figure shows the analysis of the various wavelets taking into account the average of removed noise energy ratio of the various ECG signals acquired and the average of the amplitude ratio of the filtered signals. In all wavelets, only coefficients with a frequency range between 5.625-22.5Hz were used.

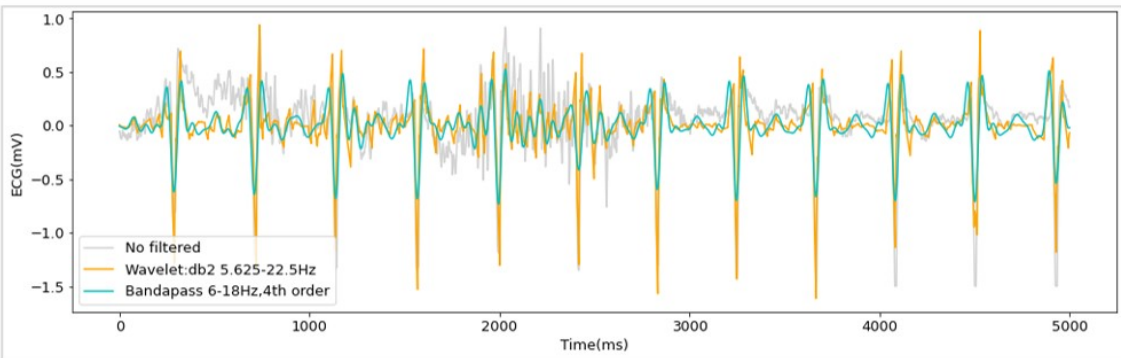

Figure 4.12: The exact signal used in Figure 4.8, but now to compare the filtered signal with the Wavelet db2 and the band-pass filter of 6-18Hz.

$$RMS_{moving}[n] = \sqrt{\frac{1}{N}\sum_{i=0}^{N-1} s^2[n-i]} \tag{4.9}$$

After the SEE was calculated and normalised (Equation 4.7), a threshold, th, was chosen, representing 35% of the absolute minimum value of the SEE signal, r, where a *square signal*, SS, was generated, and values higher than the threshold were placed at one and the rest at zero, as seen in the Equation 4.10.

$$SS[n] = \begin{cases} 1 \text{ if } r[n] > th \\ 0 \text{ if } r[n] \leq th \end{cases} \tag{4.10}$$

In Figure 4.13 we can observe how the the *Shannon Energy* (fig.4.13-B) and *Shannon Energy Envelope* (fig.4.13-C) accentuates the QRS Signal and you can also see the result of the detection of the QRS complexes by the chosen threshold.

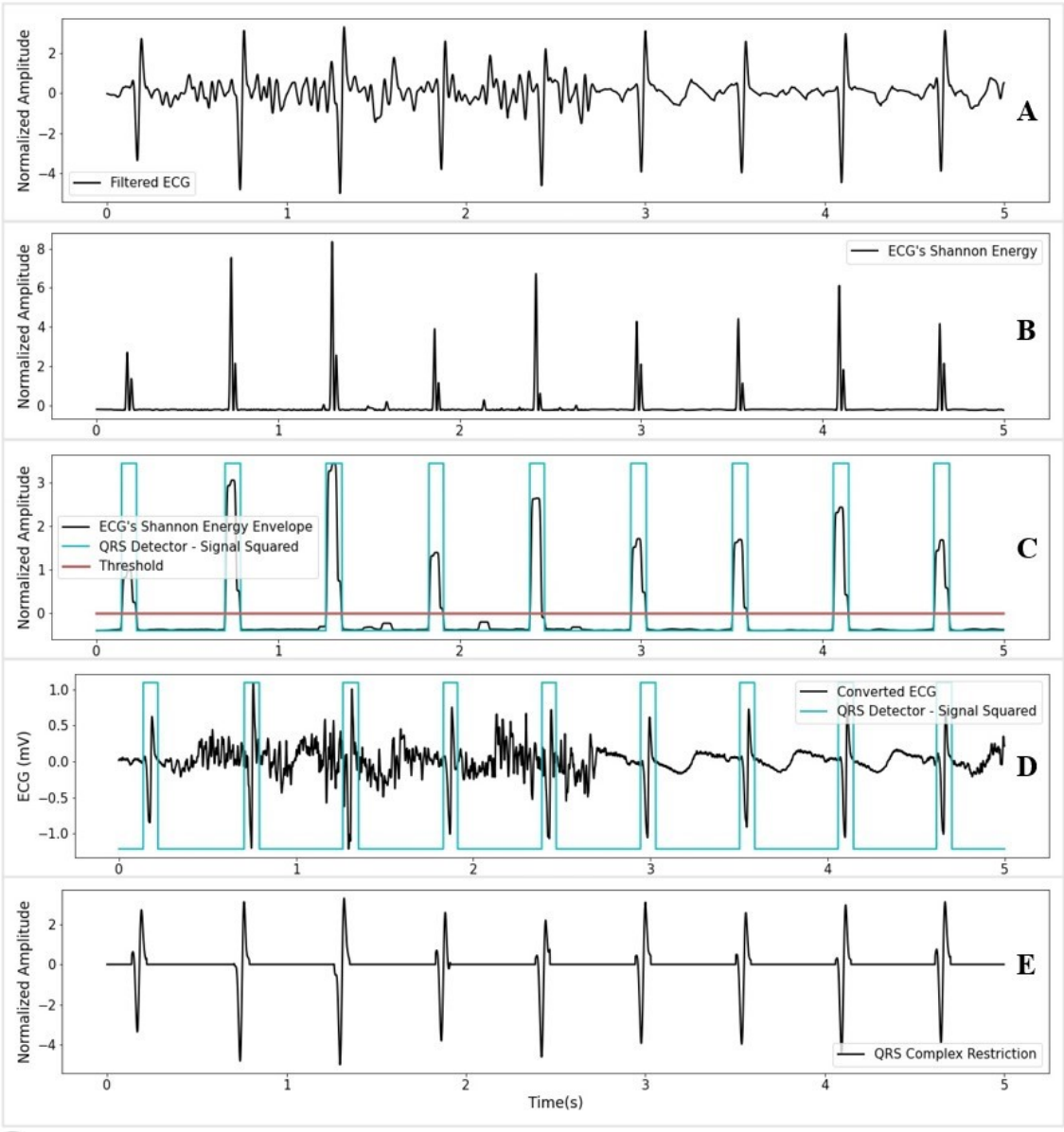

Figure 4.13: Representation of the different steps' signals for detection of the QRS Complex. In the graph **A** the filtered ECG after baseline shift and normalisation is represented. In the graph **B**, the Filtered ECG's Shannon Energy can be seen. The graph **C** shows how the squared signal was obtained by a defined threshold in ECG's Shannon Energy Envelope. In the graph **D** the QRS complex detection is represented on the top of the original ECG signal. Finally, in the graph **E**, the QRS complex restriction can be seen.

R Peaks Detection - Final Phase

After detecting the QRS complexes, it was restricted in the ECG signal so that noise

between QRS complexes could not interfere with the R peaks detection. After this, a peak detection function was used with a minimum interval between peaks. Here, as the intention is to detect the maximum correct R peaks to use the parameters of HRV a minimum interval of 0.25 ms was chosen, considering that a healthy person's heart rate does not exceed 240 bpm [120].

Although this model showed better detection of R peaks, there were still certain peaks that were not well detected, thus having the need, after segmentation of the signals for classification, to correct certain outliers with an interval of 20 seconds. Here peaks that were outside the range $[\bar{x} - 1.45\sigma, \bar{x} + 1.45\sigma]$ were eliminated. The annexes show the various signals' results of HRV, where the signals before and after the correction of outliers are compared. For example, the Figure 4.14 shows part of an ECG signal and its peak detection and the figure graph 4.13 shows the HRV signal obtained with the same example of the graphs shown in the Figure 4.5.

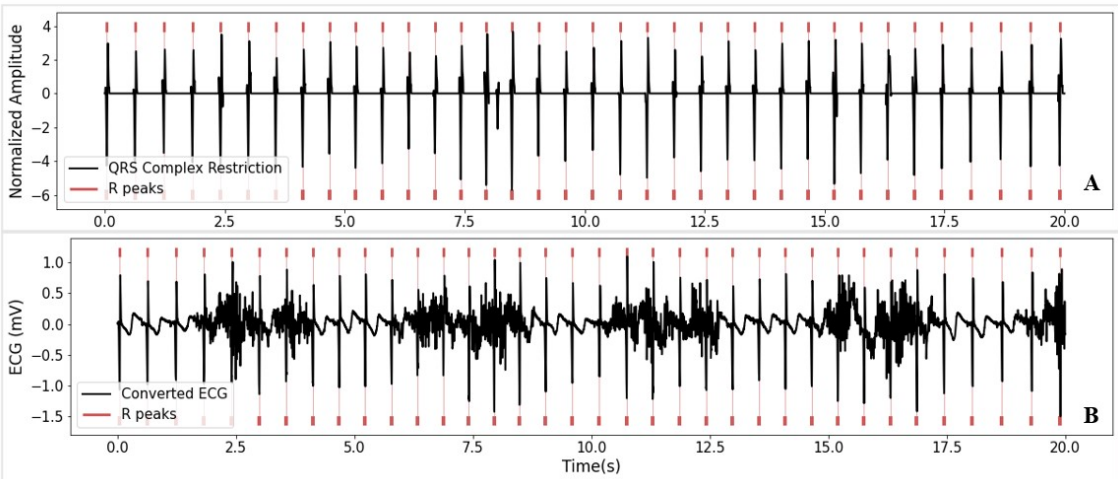

Figure 4.14: In graph **A**, the detected R peaks' location can be seen in the QRS Complex Restriction Signal, and in the graph **B**, the same is compared with the original signal.

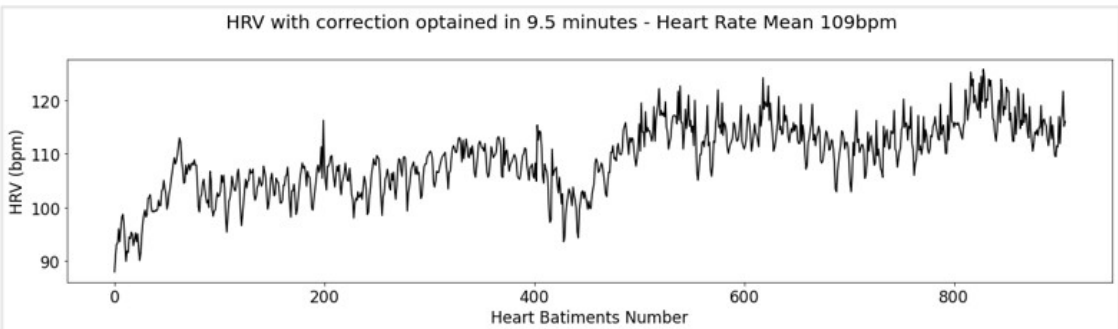

Figure 4.15: HRV signal calculated using the model used in this study, referring to the same signal used in Figure 4.6

4.3.2.2 ECG-derived metrics

In addition to HRV, other metrics were extracted from the ECG signal, which were calculated from the RR intervals signals. From the several existing metrics, we chose to calculate the Relative Heart Rate, represented in Equation 4.11 [53], the Cardiovascular Strain, in the Equation 4.12 [51], and the Cardiovascular Load, in Equation 4.13 [52]. It should be noted that, for classification, these four *time-series* were only computed after segmentation to maintain synchronisation between signals.

$$RHR[n](\%) = \left(\frac{HR_{Work}[n] - HR_{Rest}}{HR_{Max} - HR_{Rest}}\right).100\% = \left(\frac{HR_{Work}[n] - HR_{Rest}}{(208 - 0,7 \times age) - HR_{Rest}}\right).100\% \tag{4.11}$$

$$CVS[n](\%) = \left(\frac{HR_{Work}[n] - HR_{Rest}}{HR_{Rest}}\right).100\% \tag{4.12}$$

$$CVL[n](\%) = \left(\frac{HR_{Work}[n] - HR_{Rest}}{HR_{Max(8h)}}\right).100\% = \left(\frac{HR_{Work} - HR_{Rest}}{\frac{1}{3}.(220 - age) + HR_{Rest}}\right).100\% \tag{4.13}$$

All these metrics make it possible to insert the data into the more generalised model, less of the person in question, by considering factors such as age, resting heart rate (HR_{Rest}) and the minimum heart rate (HR_{Min}) detected in rest. The maximum heart rate are estimated using formulas such as *Tanaka* (HR_{Max}) and *Karvonen* ($HR_{Max(8h)}$) .

4.3.3 RIP Signal Processing

For the RIP, EMG and ACC signals, the processing steps were the same as the ones used in the article published by the same research group to which this thesis belongs, Silva et al. [90], to study this signals about the analysis of fatigability during repetitive work.

Based on this, the RIP signal processing was done for each respiratory band (Chest and Abdominal) independently. After the signal conversion was done, the signal was filtered through a *Finitude Impulse Response (FIR)* band-pass with frequency values ranging between 0.15 Hz to 0.45 Hz, considering that a breathing cycle of a healthy adult varies between 2.5 - 5 seconds [121]. Then, due to the fact that the acquisition was made at an acquisition rate of 1000Hz, a very high frequency compared with breathing frequency downsampling to the signal was made to 40 Hz,decreasing the processing time consume.

After these steps, we proceeded to extract the *percent rib cage contribution (%RC)*, represented in equation 4.14 [55, 122].

$$\%RC[n] = \frac{|RC[n]|}{|RC[n]| + |ABD[n]|}.100\% \tag{4.14}$$

Next, the Masked Sift Empirical Mode Decomposition (EMD) was implemented, a method that decomposes the signal non-linearly to extract the first intrinsic mode function [123].

In order to calculate the correlation between abdominal breathing and chest, a slicing average was applied to determine the coupling between the two signals [124].

The phase synchrony was determined by the Hilbert transform, expressed in the equation 4.15, and from its imaginary part (equation 4.16) was calculated the difference between the phase of the respiration of chest and abdominal, described in the equation 4.17.

$$g[n] = \begin{cases} \frac{2}{\pi} \sum_{t\ odd} \frac{f[t]}{n-t} & n\ even \\ \frac{2}{\pi} \sum_{t\ even} \frac{f[t]}{n-t} & n\ odd \end{cases} \tag{4.15}$$

$$\varnothing[n] = imag(g[n]) \tag{4.16}$$

$$\varphi[n] = 1 - sin\left(\left|\varnothing_{chest}[n] - \frac{\varnothing_{abdominal}[n]}{2}\right|\right) \tag{4.17}$$

In the diagram of the Figure 4.16, the applied RIP signal processing is synthesized.

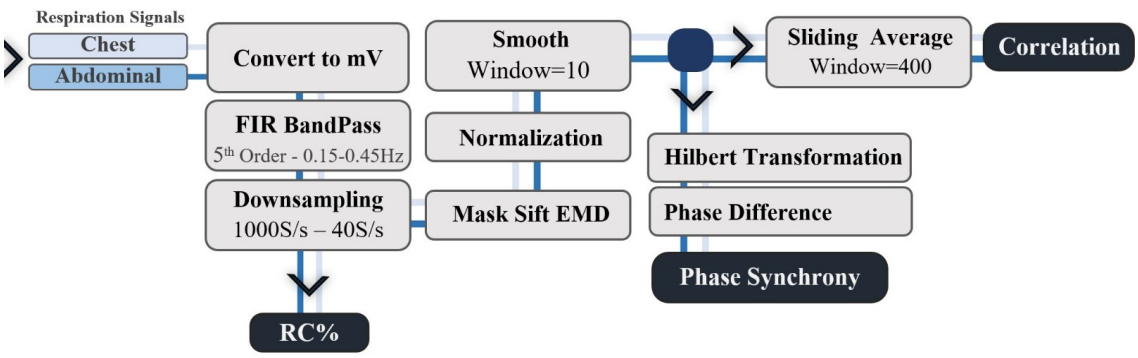

Figure 4.16: RIP signals processing schema.

4.3.4 EMG signals processing

For the EMG signal, the processing was less complex: after the conversion to mV, a baseline shift was performed, as well as a restriction of the frequency range of 10-499 Hz, through a *Finitude Impulse Response band-pass*. In order to normalize the data, for each subject, the EMG signals were divided by the maximum value acquired in the MVC protocol, as represented in the Equation 4.18. A schema of the applied acrshortEMG signal processing is represented in Figure 4.17.

$$EMG[n]_{mV} = \frac{EMG[n]_{\%MVC}}{max(MVC[n])} \tag{4.18}$$

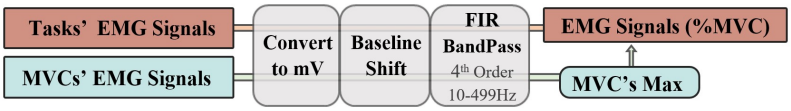

Figure 4.17: EMG signals processing's schema.

4.3.5 ACC Signal Processing

For the accelerometer signals, after conversion, a baseline shift and Butterworth band-pass filtering between 0.1-10Hz of fourth order were once again done, followed by a downsampling of 1000 Hz to 100 Hz. This process is represented in the Figure 4.18.

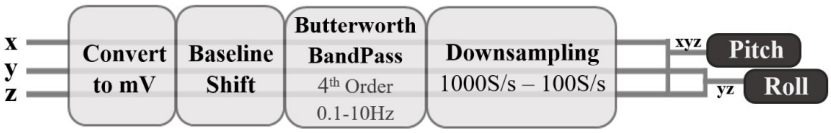

Figure 4.18: ACC signals processing's schema.

After the signals were filtered the *Roll* and *Pitch* were calculated, using the equations 4.19 and 4.20, respectively.

$$Roll[n] = arctan\left(\frac{Acc_y[n]}{Acc_z[n]}\right) \tag{4.19}$$

$$Pitch[n] = arctan\left(-\frac{Acc_x[n]}{\sqrt{Acc_y[n]^2 + Acc_z[n]^2}}\right) \tag{4.20}$$

4.3.6 Damaged data deletion

Before moving to the annotation part, there was a need to exclude some parts of acquired data that could cause possible noise to the classification models. This elimination was due to acquisition problems (namely disconnection of the sensor's cables or human errors) during the protocol or the existence of great levels of noise that could not be removed with filters.

In the ECG signals, there was the need to eliminate samples of 5 subjects, 4 for failure in the acquisition and 1 for losses in R peaks detection. In the RIP signals, there was only one failure in the individual's acquisition and two in the EMG. In the ACC, there was no problem.

4.4 Annotation

To segment the data, we could have divided it into three trials by labelling them *Baseline, Fatigue 1*, and *Fatigue 2* or *Baseline* and *Fatigue*. However, immediately after performing the fatigue protocol, there is a period in which some muscle energy resources are recovering.

Fatigue is usually defined locally by EMG; the best-known indicators are the median frequency and the signal amplitude [16, 125]. At submaximal efforts, the median amplitude is not as sensitive as desirable [90], so we have looked for changes in amplitude. Physiological changes were considered because of compensations that may be present due to the different motor strategies of the participants.

To avoid this situation, we segmented the signal with Self-Similarity Matrix (SSM) [126] as can be seen in Figure 4.19. Here we use this method to identify when the person is or is not fatigued. This segmentation was based on the SSM model, a method used in audio signals and which is currently being applied to physiological signals. This work is still under development by J. Rodrigues.

Self-Similarity Matrix is a method used mainly in music to study similarity between different time series by extracting specific characteristics from a particular window and comparing it with others to understand whether they are similar [127]. SSM is a technique that converts a time series into a suitable feature sequence. After conversion, each feature sequence's element is compared with all the other aspects of that sequence [126].

The dot product of the features matrix by itself represents the SSM. The SSM plotted a recurrence map showing when there are transitions in the signal that could correspond to events. If homogeneity properties are present block-like structures will appear. In the same way, if repetitive properties are present, a path-like structure will be drawn, that is, diagonal lines will show up when two subsequences are similar to each other. Abrupt changes will target the identification of any alteration in the signal.

The Figure 4.19 represented different types of signals of the same individual, with the corresponding SSM and *Extracted Function*. The *Extracted Function* represents the correlation over time. The peaks of this function indicate that there have been sharp changes in signal patterns [128].

After identifying the *Baseline* and *Fatigue* parts of the signal, each of these 2 parts was further divided into 8 samples of signal, each labelled as *Baseline* or *Fatigue*, accordingly.Figure 4.20 shows different signals to the exact part divided.

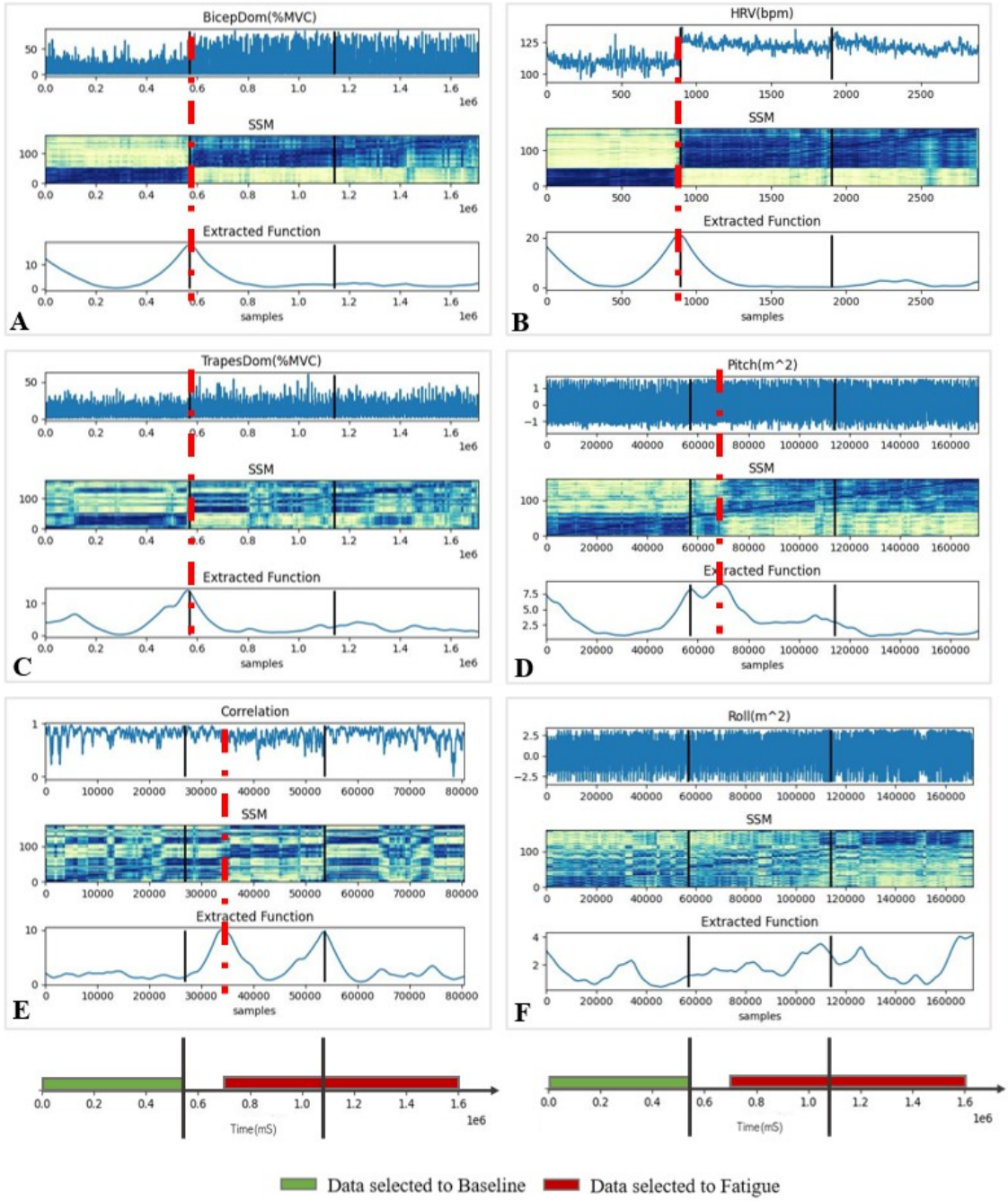

Figure 4.19: Analysis of different signals with the SSM matrix and the extracted function for the same individual, where the tasks are concatenated (time break shown with vertical black bars). The vertical red bars show the point where the pattern change is verified. Considering these changes, a particular samples' time was selected for *Baseline* (green bar) and another part for *Fatigue* (horizontal red bar). In the time intervals where there was uncertainty about whether or not the person was in fatigue, it was decided not to use the corresponding data.

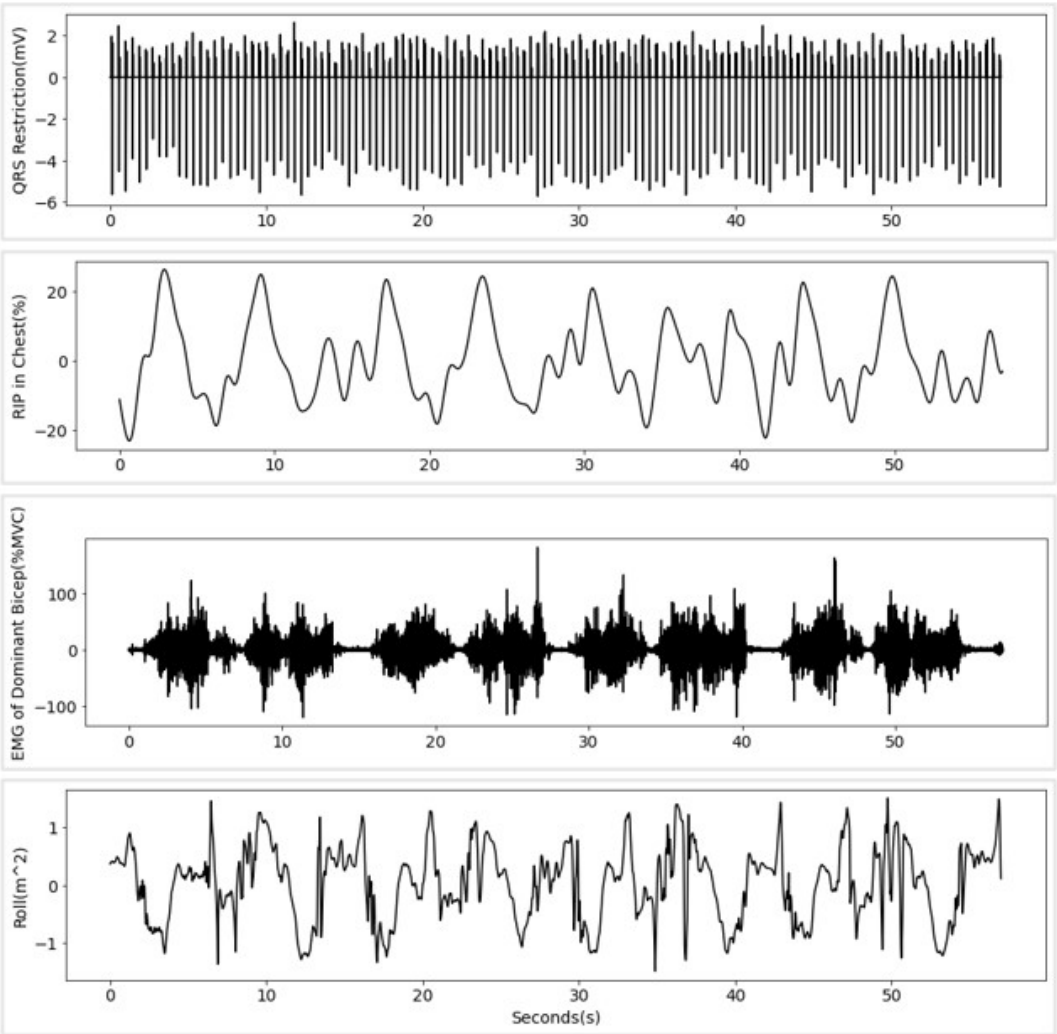

Figure 4.20: Example of 4 signals, one of each signal type corresponding to the same subject sample used later to extract features.The signal from above corresponds to an already processed signal coming from the ECG, followed by a signal from RIP, EMG and ACC.

4.5 Features Selection and Extraction

After the sample segmentation, features related to each time-series present in Table 4.2 were extracted. To this end, part of the temporal and statistical domain features and the median frequency for the EMG case were used [129]. This extraction was made through the Time Series Feature Extraction Library (TSFEL) package [130]. For each time series, specified in table 4.2, the following variables were selected [131]:

- **Statistical Domain:**
  - Interquartile range;
  - Kurtosis;
  - Mean;
  - Mean absolute deviation;
  - Median;
  - Median absolute deviation;
  - Root mean square;
  - Skewness;
  - Standard deviation and
  - Variance.
- **Temporal Domain:**
  - Autocorrelation and
  - Peak to peak distance or Amplitude.
- **Spectral Domain:**
  - Median frequency - just for EMG.

After the features were extracted, features with *Missing values* were eliminated (because of divisions by 0). Due to the existence of large amounts of features, there was a need to seal the best features, eliminating features that were highly correlated with others (more than 95% of correlation) [132] and then those features that had a variance below 80% (chosen value), since these do not contribute to a good model. This variance considers the Bernoulli Distribution where $Var[X] = p(1-p)$, for $p = 80\%$ threshold was 0.16 [133]. Here the selection of features was made individually by signal type. That means features of EMG, ECG, RIP and ACC were studied independently.

To understand which features could best contribute to the model, features were ordered in order of best score, in this case we used the *chi-Squared* distribution test (X^2) [134]

Table 4.2: Time Series where the features were extract per signal type

ECG	RIP	EMG	ACC
CVS	% Rib cage contribution	Trapezius' EMG (%MVC)	Roll
CVL	Phase Sycronization	Bicep Brachii's EMG (%MVC)	Pitch
HRV	Correlation		
RHR			

4.6 Classification Model Development

Seven different types of classification algorithms were experimented to find the best model for fatigue classification. These models were: *Random Forest*, *Support Vector Machine*, *Logistic Regression*, *Gaussian Naive Bayes*, *K-Nearest Neighbours*, *Decision Tree* and *AdaBoost*. In addition, different models were tested, considering each signal individually or in groups.

Four evaluation metrics were used to analyse the classification models. *Accuracy, Precision, Recall* and *F1-Score* are represented in Equations 4.21,4.22, 4.23 and 4.24 [135].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.21}$$

$$Precision = \frac{TP}{TP + FP} \tag{4.22}$$

$$Sensitivity\ (Recall) = \frac{TP}{TP + FN} \tag{4.23}$$

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4.24}$$

Due to a considerably low number of samples and individuals, it was observed that different trains and test sets would lead to markedly disparate results. Therefore, it was decided to use "Leave-one subject-out cross-validation". This cross-validation was done interactively by excluding one individual for the Test and the rest for training in order to evaluate the classifiers.

This *leave-one-out cross-validation* was considered the average metrics of a training cycle. In each cycle, a subject was separated for testing, and the rest was used as training to avoid overfitting [136].

The table 4.3 shows the number of subjects and samples used for training each model, considering the features corresponding to one or more types of acquired signals. After deciding which were the best features and classification algorithms, the models were tested in order to understand the importance of the various features for the classifier and the sensitivity and specificity were calculated to compute the Receiver Operating Characteristic (ROC) curve [137] of each classifier. In this step, 20% of the subjects were used for testing and the rest for validation and training.

To understand the model's features importance a *Permutation feature importance* was used [138]. The dataset used to train the models (80%) was divided between tests and validation(20% of 80%).

Table 4.3: The number of subjects and samples used in each training model considering the signal(s) type of the features applied.

Total number	**ECG**	**EMG**	**RIP**	**ACC**	**All Signals**
Subjects	20	23	24	25	17
Samples	320	366	382	398	272
Total number	**ECG + EMG**	**ECG + RIP**	**EMG + RIP**	**ACC + EMG**	**ECG+EMG+RIP**
Subjects	18	19	22	23	17
Samples	288	304	350	366	272

5

Results

This section presents the more revealing features of the classification models and the evaluation metrics of different algorithms. In this study, we chose first to analyze the data of the different types of signals independently. That is, it was started by analyzing datasets of ECG, EMG, ACC and RIP separately and then analyzed the datasets in a combined way.

5.1 Features Selection

After the exclusion of features and their ordering by *chi2*, the behaviour of the various algorithms with the number of features used to feed the classifiers was analyzed. The features were added to the models cyclically in an orderly manner. The models were started with the feature with the best score, and then the model was evaluated. For this, the average accuracy was calculated through Leave-one-out cross-validation.

The Figure 5.1 shows four charts corresponding to each dataset. Here, the evolution of the accuracy of the various algorithms as ordered features were added. Graph **A** shows the results of the EMG Dataset. It is observed that from the six best features, the models' results did not improve or even worsened. The same phenomenon happens in graph **B** with the ECG Dataset. In the graph **C**, with the RIP Dataset, although there are algorithms that decrease its accuracy, it was chosen ten features based on Random Forest and AdaBoost algorithms. In graph **D**, with the ACC Dataset, although SVM and Logistic Regression algorithms present a more considerable accuracy with only one feature, they decrease their accuracy by adding features. Thus, we chose 8 features based on Random Forest.

Based on this analysis, 6 features were chosen for EMG and ECG, as well as 10 and 8 features for RIP and ACC, respectively. In the Table 5.1, the features that were considered for the classification models are declared, separated by signal type.

Regarding the study of classification models that combine all features, an analysis similar to the previous process was made. Figure 5.2**-A** shows the evolution of accuracy over the 56 ordered features that were not excluded by correlation or variance. It was found that the highest accuracies occurred with a low number of features. Thus, Figure 5.2**-B** shows

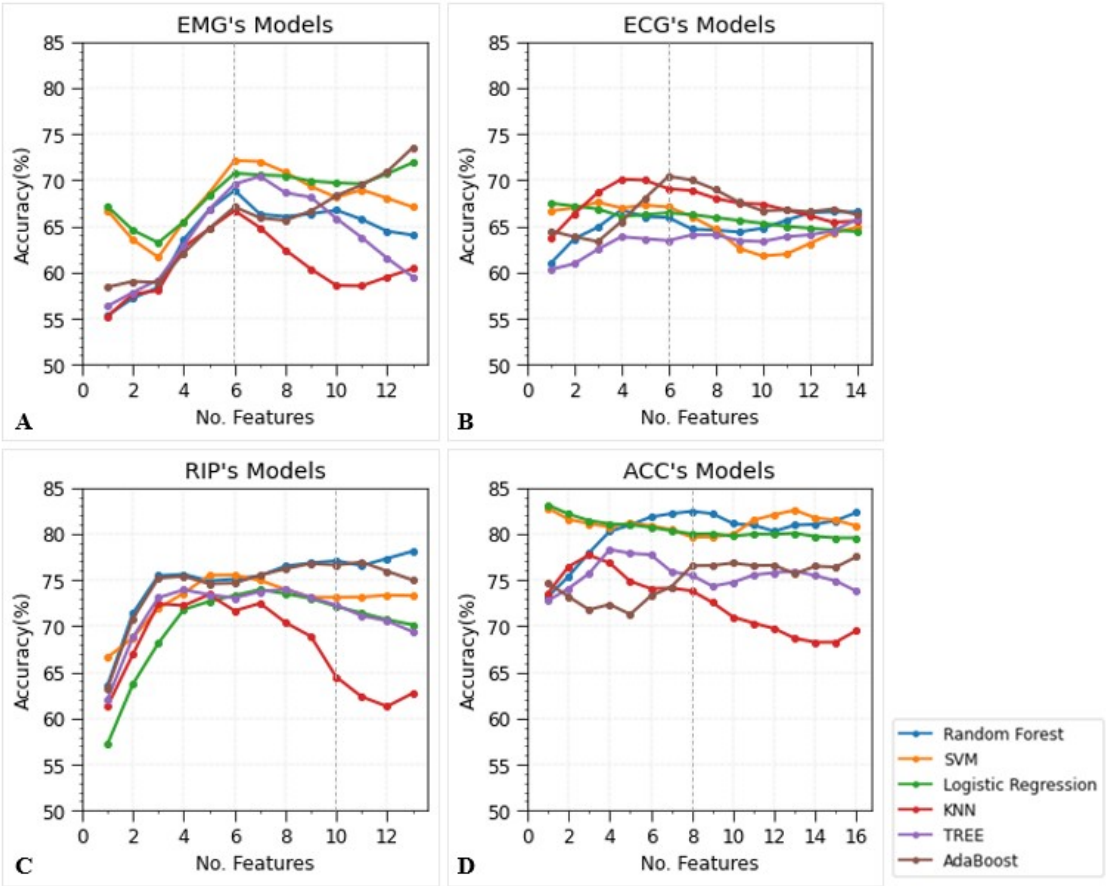

Figure 5.1: The evolution of the Accuracy of the various algorithms is based on the number of features added to the signal. The features are added decreasingly from the *Chi2* score. Graphs A, B, C and D represent the models that use the dataset of the EMG, ECG, RIP and ACC signals, respectively.

Table 5.1: Best features selected considering each type of signal.

Chi2	**ECG Signals**	**EMG Signals**
1º	CVL-Root mean square	Bicep-Autocorrelation
2º	CVL-Mean	Trapezius-Autocorrelation
3º	HRV-Mean	Trapezius-Median
4º	RHR-Mean	Bicep-Kurtosis
5º	CVS-Mean	Trapezius-Mean
6º	CVS-Variance	Trapezius-Amplitude
Chi2	**RIP Signals**	**ACC Signals**
1º	Phase Syncr.-Autocorrelation	Pitch-Autocorrelation
2º	Correlation-Autocorrelation	Roll-Autocorrelation
3º	Phase Syncr.-Kurtosis	Roll-Mean absolute deviation
4º	Correlation-Kurtosis	Pitch-Root mean square
5º	%RC-Skewness	Roll-Interquartile range
6º	PhaseSyn-Skewness	Pitch-Mean absolute deviation
7º	Correlation-Skewness	Roll-Kurtosis
8º	%RC-Interquartile range	Roll-Median
9º	%RC-Mean	
10º	%RC-Kurtosis	

that the number of features indicated for classification will be 7. Table 5.2 shows in an orderly fashion the features that were used to train the general classifiers.

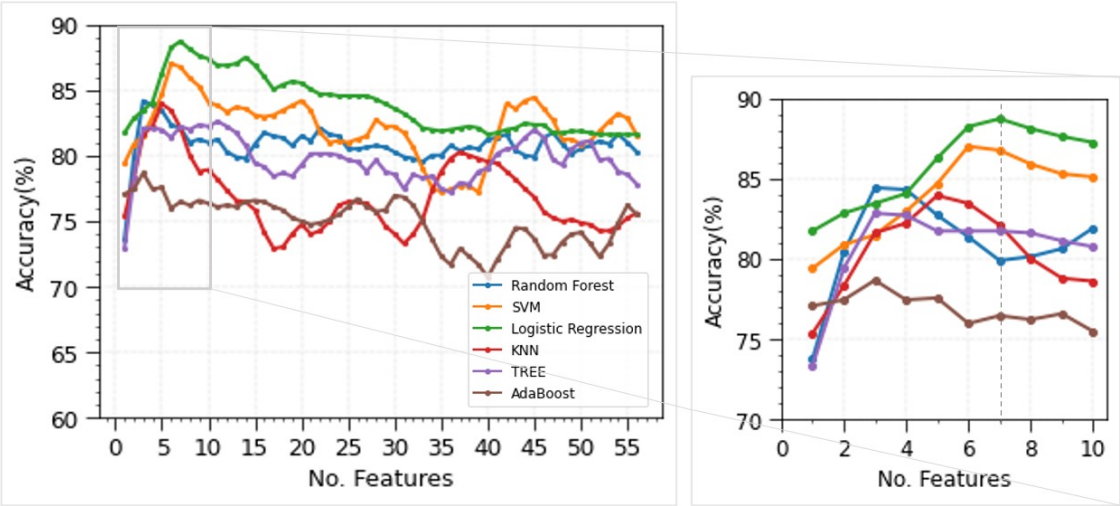

Figure 5.2: The Accuracy of the various algorithms is based on the features' number added to the signal. These models take to account features of all signal types.

Although the selection of the number of features was based on Accuracy, the F1-Score, Precision and Recall were also considered. Its evolution can be observed in the annexes.

Table 5.2: Best features selected considering all signals.

Chi2	**Best Features**
1º	Pitch-Autocorrelation
2º	CVL-Root mean square
3º	Roll-Autocorrelation
4º	Phase Syncr.-Autocorrelation
5º	Bicep-Autocorrelation
6º	CVL-Mean
7º	HR-Mean

The autocorrelation feature shows more relevant information for the models, highlighting the accelerometer, phase synchrony and Bicep autocorrelations. In addition, when we look at all datasets, the average value of HR and CVL and also CVL's root mean square showed to be informative in the classification.

After this, the relationships between the most relevant features were observed, considering the class of the samples. In the figure, one can observe four scatter plots showing the behaviour of eight features, two best features of each signal type. For example, graphs **A** and **B** show the relationship between ECG and RIP features. Graphs **C** and **D** show the relationship of EMG with ACC. This relationship was chosen because ECG and RIP are associated with more general fatigue and EMG and ACC with fatigue at the upper limb level. Through this analysis, it is possible to observe a distinction between fatigue and not fatigue evidence in these features, for example.

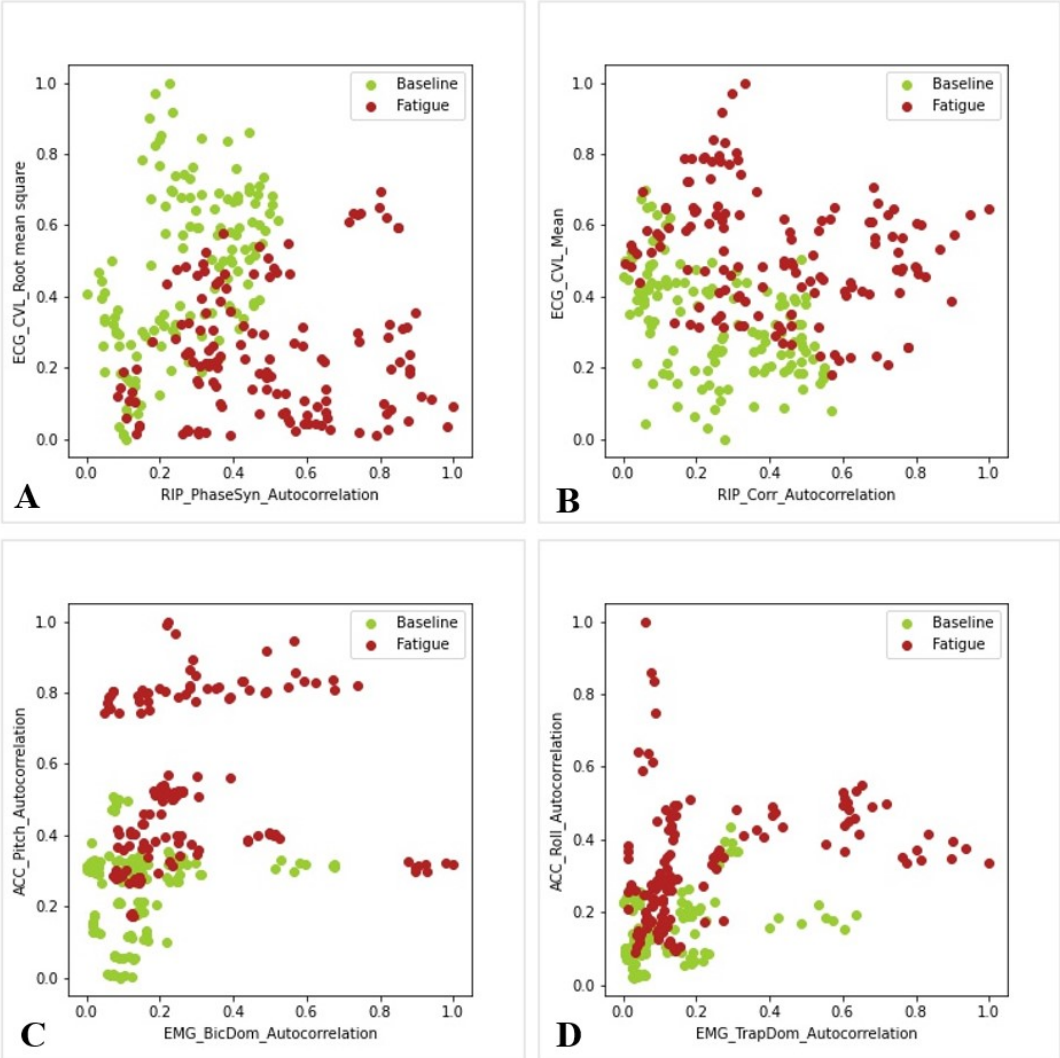

Figure 5.3: Scatter Plots between different features. Plot **A** represents the relationship between Phase Autocorrelation Synchronization with the root Mean square of the CVL. The relationship between the RIP Correlation Autocorrelation and CVL mean can be seen in plot **B**. Plot **C** shows the relationship between the EMG autocorrelation of the Bicep and the pulse pitch. In Plot **D**, the relationship between the autocorrelation of the Trapezium with the Wrist Roll.

5.2 Model's Classification

Having defined the features to use, several models were trained with the different ML algorithms through *Leave-one-out cross-validation* (between subjects).

The best classifier was the general model, which used all kinds of signals. In the general model, the algorithm with the highest results was the Logistic Regression with an Accuracy mean of 87.87% ± 15.08%. For both models trained with ACC and RIP data, the algorithm that showed the best results was Random Forest, with an average Accuracy of 82.11% ± 19.32% for ACC and 76.12% ± 17.06%. As for EMG, SVM stood out with an average Accuracy of 73.29% ± 20.98%. Finally, the ECG was the signal that showed the worst results in the fatigue classification, the best classifier being the AdaBoost, with only an average accuracy of 70.31% ± 19.55%. The tables 5.3 , 5.4, 5.5, 5.6 and 5.7 show the mean and standard deviation (std) of Accuracy, Precision, Recall, and F1-Score for each different algorithms and data types used. These tables are sorted by classification's quality. Having defined the features' number, several models were trained with the different ML algorithms.

Table 5.3: Evaluation metrics of workload classification model based on All time-series processed, using the seven higher *Chi-Square* features.

	All Times Series							
	Accuracy (%)		**Precision (%)**		**F1 Score (%)**		**Recall (%)**	
Algorithm	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Random Forest	80.15	20.31	80.15	20.31	77.00	29.52	75.50	26.74
SVM	86.03	17.75	86.03	17.75	88.82	18.23	83.51	21.94
Logistic Regression	**87.87**	**15.08**	**87.87**	**15.08**	**88.88**	**18.22**	**86.39**	**18.16**
Gaussian Naive Bayes	85.29	19.41	85.29	19.41	85.32	23.96	82.19	24.55
KNN	79.04	17.01	79.04	17.01	81.08	22.03	75.61	21.62
Decision Tree	80.88	20.10	80.88	20.10	82.33	23.94	77.11	25.22
AdaBoost	77.94	21.77	77.94	21.77	79.43	25.02	75.01	25.38

Table 5.4: Evaluation metrics of workload classification model based on ACC's time-series processed, using the eight higher *Chi-Square* features.

	ACC's Times Series							
	Accuracy (%)		**Precision (%)**		**F1 Score (%)**		**Recall (%)**	
Algorithm	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Random Forest	**82.11**	**19.32**	**82.11**	**19.32**	**82.98**	**24.82**	**79.54**	**23.51**
SVM	78.75	21.72	78.75	21.72	73.19	31.81	73.77	28.27
Logistic Regression	79.82	20.20	79.82	20.20	77.49	28.31	75.75	25.85
Gaussian Naive Bayes	81.18	20.75	81.18	20.75	77.14	30.24	76.58	27.34
KNN	72.14	19.65	72.14	19.65	73.39	24.21	68.46	23.53
Decision Tree	72.25	19.78	72.25	19.78	70.90	25.91	66.97	24.65
AdaBoost	76.07	20.39	76.07	20.39	74.15	28.12	71.85	25.68

Despite presenting the results, the *standard deviation* is relatively high, suggesting that the models depend a lot on the test subject, which indicates a low number of samples and subjects for the model. When testing the best algorithms for each feature's data set, with 20% of a random subjects, it was found that the results improved with particular sets of individuals, with the Area Under Curve (AUC) of the ROC charts above 90%. However,

Table 5.5: Evaluation metrics of workload classification model based on RIP's time-series processed, using the ten higher *Chi-Square* features.

	RIP's Times Series							
	Accuracy (%)		**Precision (%)**		**F1 Score (%)**		**Recall (%)**	
Algorithm	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Random Forest	**76.12**	**17.06**	**76.12**	**17.06**	**80.06**	**14.57**	**73.57**	**18.73**
SVM	73.55	21.73	73.55	21.73	73.80	27.29	69.80	25.77
Logistic Regression	71.17	19.39	71.17	19.39	71.51	25.89	67.16	23.45
Gaussian Naive Bayes	67.86	20.13	67.86	20.13	67.42	25.87	64.00	23.94
KNN	59.82	16.83	59.82	16.83	61.11	18.25	58.51	17.38
Decision Tree	72.21	17.73	72.21	17.73	76.37	21.03	69.65	21.04
AdaBoost	76.08	19.70	76.08	19.70	76.84	23.02	73.84	22.57

Table 5.6: Evaluation metrics of workload classification model based on EMG's time-series processed, using six higher *Chi-Square* features.

	EMG's Times Series							
	Accuracy (%)		**Precision (%)**		**F1 Score (%)**		**Recall (%)**	
Algorithm	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Random Forest	67.00	24.80	67.00	24.80	66.34	32.21	62.35	28.87
SVM	**73.29**	**20.98**	**73.29**	**20.98**	**68.85**	**30.98**	**66.99**	**27.63**
Logistic Regression	70.65	18.29	70.65	18.29	68.96	28.11	64.30	24.56
Gaussian Naive Bayes	59.47	15.43	59.47	15.43	50.43	29.53	48.03	21.67
KNN	63.47	21.57	63.47	21.57	64.35	27.90	58.65	24.79
Decision Tree	67.62	21.08	67.62	21.08	60.88	30.91	61.73	26.73
AdaBoost	65.95	20.95	65.95	20.95	64.00	28.12	61.14	24.89

Table 5.7: Evaluation metrics of workload classification model based on ECG's time-series processed, using six higher *Chi-Square* features.

	ECG's Times Series							
	Accuracy (%)		**Precision (%)**		**F1 Score (%)**		**Recall (%)**	
Algorithm	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Random Forest	63.44	15.57	63.44	15.57	61.32	26.90	55.46	21.25
SVM	66.88	19.07	66.88	19.07	61.75	30.79	58.03	26.15
Logistic Regression	66.25	18.69	66.25	18.69	61.47	30.85	57.59	25.21
Gaussian Naive Bayes	67.50	18.18	67.50	18.18	67.20	28.59	59.46	24.62
KNN	68.75	15.18	68.75	15.18	69.92	24.79	63.47	20.43
Decision Tree	65.62	16.09	65.62	16.09	64.04	25.79	59.85	21.00
AdaBoost	**70.31**	**19.55**	**70.31**	**19.55**	**68.76**	**28.39**	**64.24**	**25.39**

with other sets, the classification worsened (several results can be seen in annexes IV). It was realised that depending on the sets, features that had greater importance for the model also varied according to the sample.

In addition, models were also tested using combinations of two and three types of signals, but the results were not better.

6

Discussion

The aim of this study yielded to characterised the presence of fatigue while performing repetitive tasks through biosignals. To achieve this purpose, both variables related to local physiological changes (at the muscle level) and general (large muscle masses) were included. These adaptations will occur depending on the relationship between the stimulus that the task represents and the morpho-physiological characteristics of the worker. In addition, different organic systems respond independently but at different times to the energy demand that is required for a particular task.

In this study, features were evaluated separately (e.g., EMG, ACC, ECG, and RIP) to realise how is the contribution by themselves, as well as all the signals together. The results showed, for the proposed task, that there is a change in physiological response after the presence of fatigue. The segmentation algorithm (SSM) used to identify changes in the signals was sensitive to them, differentiating the signals during *Baseline* compared to what occurs in *Fatigue*.

The classification with all variables achieved the greatest accuracy (88%) and showed a high contribution of motion-associated features (ACC) compared to physiological ones. In contrast, the information from the signal that is usually used to measure local fatigue, i.e.,the EMG showed less relevance when compared to the other signals. The accelerometer showed a high accuracy (82.11%) when used alone, as it was also the one that provided the most relevant features. The EMG only achieved 73.29% of accuracy.

Autocorrelation of inertial measures, respiratory synchrony, and the root mean square of the cardiovascular load were the features with the highest contrast between the two classes and the highest weight for the fatigue classifiers.

When analyzing the various signals individually, it was possible to see physiologic changes throughout the tasks. It can be distinguished when the person was or was not in fatigue. Increased heart rate, changes in breathing patterns, movement and muscle contractility, showed fatigue status in the tasks performed during the experiment.

It was observed that the wrist movement (ACC) is the signal that presents the best results,

with 83% of accuracy and a sensitivity of 80%, which is not bad when compared to studies using various accelerometers located in different parts of the body with gyroscope as Baghdadi et al. [100], showing 90% of accuracy. Some studies tried to combine ACC with HR. Lambay et al. [97] and Maman et al. [29] showed an accuracy of 65% and a sensitivity between 79% and 100%, respectively. Testing in the same way as the articles referred, the combination of ACC and ECG metrics (with the KNN algorithm), the accuracy and sensitivity increased 3% and 2%, respectively. The autocorrelation of both Roll and Pitch are the ones that show more importance for the ACC models. The changes found in the accelerometer and its discriminative capacity seem to be associated with different motor strategies that the participants performed to compensate fatigue.

Regarding the respiration signals, the study by Silva et al. [90], which uses the same data as this study, showed differences between baseline and fatigue. In this study, both phase synchrony and correlation decreased. This decrease is due to the fatigue of the respiratory muscles, such as the diaphragm, as the skeletal muscles no longer have enough energy for the task. The classifier's results of this study (76% of Accuracy) were better compared to data from the study by Umer et al. [98], which presents an Accuracy of 39.3%. This significant difference between studies may be due to the fact that the article uses only the respiratory rate while here used more complex metrics that consider that take into account the relationship between the chest and abdominal breathing.It should be noted that the RIP model showed an F1-Score of 80%, which reinforces the model's good quality.

Although several studies [15] indicate that muscle fatigue results in decreased muscle contractility at the EMG level, the models did not show significant results. Looze et al. [16] suggest that the "combination of frequency decrease and amplitude increase over time"; however, in this study, these features did not show relevance but, yes, the autocorrelation.

During the protocol for lifting the box, the individual usually does shoulder abduction, but what was observed was that there were people who began the task with abduction and when they began to get tired, they tended to change to flexion.

In addition, there were still people who began with flexion of the shoulder and ended with abduction. These two movements cause muscle contraction of different structures. Abduction tends to cause a more significant contraction of the upper trapezius. Then, in the bicep and flexion, the opposite happens. Given this, it is possible to explain the results in question. Due to a variety of movements between tasks and individuals, it may have happened that there were individuals. For example, the bicep shows a higher %MVC and then descends from a particular moment, but there were others where the opposite happens, eventually 'confusing' the model.

Finally, the ECG classification was the data set that presented less precision. Although the SSM matrices, at the individual level, show more evident disbelief in HRV than in the other signs. This result may be due to intersubject variability. Individual characteristics such as age, frequency of physical exercise, or even consumption or not of caffeine can

lead to different HR and HRV.

The processing level of R peaks, due to the high amount of noise, was not possible to fully identify them. These minor flaws may also have had some weight in the poor fatigue rating.

Despite this, when we compare the results obtained with other articles that use heart signals, it is clear that these results are not isolated. In the studies by Darbandy et al. [95], Nasirzadeh et al. [96], Aryal et al.[99] and Umer et al. [98], the accuracy was 78.18%, 77.81%, 59% and 56.6%, while in this study, it was 69% to Manual Material Handling. Darbandy et al. [95], despite only resorting to the Heart rate, has low sensitivity, maximum of 60%, which indicates not being able to identify the classes correctly. The Nasirzadeh et al. [96] shows better results despite using only HRV, and reference features Entropy and Shannon Energy as the best features. Here, it was not possible to calculate these features due to the exact time difference between samples. Aryal et al. [99] and Umer et al. [98] have inferior results to this study; however, it is necessary to take into account that they only resort to the heart rate.

After all, the intention was to create a classifier capable of detecting fatigue in any person. The variability of inter-individual signals can be pointed out as one of the leading causes of the presented results. This statement can be confirmed, taking into account the high value of the standard deviation, about 20%, which shows much dependence of the classifier on the individual under study.

It was also noticed, in the segmentation phase, that for the same individual, there were signals that showed better the change of pattern than the others. Despite this evidence, in different subjects, the signal highlighted was not the same. There were people whose pattern change was recorded in the ECG and Biceps signal, but there were others whose change was more noticeable, for example, in Tricep or the?? correlation. This evidence may also be behind the classification results. With different features representing different, weights and individuals, they can justify poor results in models and justify that multimodal classifiers have better results than models with only one signal. Table 6.1 presents the results of the different studies found, which also uses supervised classifiers to detect fatigue at work, also having as an experimental Manual Material Handling task.

Table 6.1: Results of studies that used classifiers to detect fatigue, also having as experimental Manual Material Handling task. TR stands for Thermoregulation and Resp. R, the respiration rate

Year	Study	Signals Used	Accuracy	Participants
2021	Lambay et al. [97]	IMU, HR	65%	total of 100 epochs
2020	Darbandy et al. [95]	HR	78.18%	8
2020	Umer et al. [98]	HR, HRV, Resp. R, TR, Borg's Scale	All features: 42,1%-95.3% HR and HRV data: 56,6% Resp. R data: 39,3% TR data: 87,5%	10
2020	Nasirzadeh et al.[96]	HR	90.36%	8
2018	Baghdadi et al. [100]	IMU (ACC and Gyroscope)	90%	20
2017	Maman et al. [29]	ACC ankle, wrist, hip and torso; HR	imbalanced data sensitivity: 65-100% specificity: 70-79%	6
2017	Aryal et al. [99]	HR, TR, EEG, Borg's Scale	All data: 82% HR data: 59% TR data: 79%	12

7

Conclusion and Future Work

Detecting when the person is exerting too much effort for a simple task is necessary to understand when the person should rest. Fatigue classification models were proposed in order to be able to develop a wearable capable of analyzing the physical state of the person in real-time. The physiological response to fatigue is diverse and occurs at both the neuromuscular and cardiovascular levels.

The task of loading boxes or other objects, when it is done repetitively for long periods, leads to the appearance of fatigue. Even with a box not excessively heavy and without the possibility of awkward positions, tiredness and during the task. Fatigue is a physical condition characterized by a combination of bodily changes. Developing a classifier capable of identifying fatigue must consider various physiological signals. By collecting data from ECG, EMG, RIP and ACC, it was possible to record the different responses of the body.

Multimodal data showed better classification than data of only one type; still, motor features showed greater relevance than physiological features. Motor strategies happen in this type of task, so the detection of muscle activity only of one muscle individually should not be used.

The results of the models trained by *Leave-one-out cross-validation* can demonstrate good fatigue indicators when we observe the various metrics evaluation. Due to the low amount of individuals and high intersubject variability appear to have lower classifications. Increasing the number of individuals can increase the robustness of the models.

As for the R peaks detection algorithm, although it was not the objective of this study, it demonstrated a better ability to locate these points than the other existing algorithms.

During the processing of these signals, it was observed that the physiological responses of the various individuals, despite different beings, had repeated physiological responses. This evidence can justify the results of the models. It may be interesting, in the future, to analyze the possibility of the existence of these patterns.

Future work should focus on determining fatigue indice through objective data such as

biosignals. The amount of multimodal data extracted allows us to draw several exciting conclusions and various types of analysis using EMD. In this way, it could be possible to understand how the most relevant features evolve for each different subjects. Different physiological adaptations may arise depending on the individual, physical condition or age. Surveys or recordings of mood swings should be considered to understand the physiological changes in the person's mental state.

Better detection of R peaks or reducing ECG noise as much as possible can help improve Accuracy. Resorting to a combination of EMG signals, rather than isolated signals, can be an asset for identifying muscle fatigue. The segmentation of signals in samples with the same time interval should be used in order to be able to use Features such as energy, the area under the curve or entropy. Here, only divided the tasks by parts, it was impossible to use these features for the models, but they can give relevant information.

Bibliography

[1] J. M. Lourenço. *The NOVAthesis LaTeX Template User's Manual*. NOVA University Lisbon. 2021. URL: https://github.com/joaomlourenco/novathesis/raw/master/template.pdf (cit. on p. ii).

[2] A. Williams et al. "Musculoskeletal conditions may increase the risk of chronic disease: a systematic review and meta-analysis of cohort studies". en. In: *BMC Med.* 16.1 (2018), p. 167 (cit. on p. 1).

[3] E. R. Observatory. "Eurostat (Labour Force Survey ad hoc module". In: European Health, 2013. ISBN: 978-92-9479-145-0. DOI: 10.2802/66947. URL: http://europa.eu (cit. on pp. 1, 4, 5).

[4] INE. "Estatísticas da Produção Industrial - 2019". In: *Instituto Nacional de Estatística, I. P.* (2020), p. 41. ISSN: 0872-9298 (cit. on p. 1).

[5] *Statistics on employment - ILOSTAT*. URL: https://ilostat.ilo.org/topics/employment/ (cit. on p. 1).

[6] *Manual handling (Labour administration and inspection)*. URL: https://www.ilo.org/global/topics/labour-administration-inspection/resources-library/publications/guide-for-labour-inspectors/manual-handling/lang--en/index.htm (cit. on p. 1).

[7] "Health and Safety Executive Health and Safety Executive Manual handling assessment charts (the MAC tool)". In: (). URL: www.hse.gov.uk/msd/mac/scoresheet.htm. (cit. on pp. 1, 16).

[8] "Health and Safety Executive Assessment of repetitive tasks of the upper limbs (the ART tool)". In: (). URL: www.hse.gov.uk/msd/uld/art. (cit. on pp. 1, 15).

[9] J. Santos et al. "The influence of task design on upper limb muscles fatigue during low-load repetitive work: A systematic review". In: *International Journal of Industrial Ergonomics* 52 (2016-03), pp. 78–91. ISSN: 0169-8141. DOI: 10.1016/J.ERGON.2015.09.010 (cit. on pp. 1, 16).

[10] M. F. Antwi-Afari et al. "Biomechanical analysis of risk factors for work-related musculoskeletal disorders during repetitive lifting task in construction workers". en. In: *Autom. Constr.* 83 (2017), pp. 41–47 (cit. on p. 1).

[11] *Chapter 3 | Surgeon General Report | CDC*. URL: https://www.cdc.gov/nccdphp/sgr/chap3.htm (cit. on p. 1).

[12] E. Åhsberg, F. Garnberale, and A. Kjellberg. "Perceived quality of fatigue during different occupational tasks Development of a questionnaire". In: *International Journal of Industrial Ergonomics* 20.2 (1997-08), pp. 121–135. DOI: 10.1016/s0169-8141(96)00044-3. URL: https://doi.org/10.1016/s0169-8141(96)00044-3 (cit. on pp. 1, 16).

[13] S. M. Marcora, A. Bosio, and H. M. de Morree. "Locomotor muscle fatigue increases cardiorespiratory responses and reduces performance during intense cycling exercise independently from metabolic stress". en. In: *Am. J. Physiol. Regul. Integr. Comp. Physiol.* 294.3 (2008), R874–83 (cit. on p. 2).

[14] "OPERATOR | DIGITAL TRANSFORMATION IN INDUSTRY WITH A FOCUS ON THE OPERATOR 4.0 E I N F O R M AT I O N A N D C O M M U N I C AT I O N S O L U T I O N S-A I C O S Project Overview". In: (). URL: http://operator-i40.com (cit. on p. 2).

[15] D. C. Toledo-Pérez et al. "Support Vector Machine-Based EMG Signal Classification Techniques: A Review". In: *Applied Sciences* 9.20 (2019-10), p. 4402. DOI: 10.3390/app9204402. URL: https://doi.org/10.3390/app9204402 (cit. on pp. 2, 17, 48).

[16] M. de Looze, T. Bosch, and J. van Dieën. "Manifestations of shoulder fatigue in prolonged activities involving low-force contractions". In: *Ergonomics* 52.4 (2009). PMID: 19401894, pp. 428–437. DOI: 10.1080/00140130802707709. eprint: https://doi.org/10.1080/00140130802707709. URL: https://doi.org/10.1080/00140130802707709 (cit. on pp. 2, 17, 35, 48).

[17] *Occupational health*. URL: https://www.who.int/health-topics/occupational-health (cit. on p. 4).

[18] "SAFETY AND HEALTH AT THE HEART OF THE FUTURE OF WORK Building on 100 years of experience". In: () (cit. on p. 4).

[19] V. Patel et al. "Trends in Workplace Wearable Technologies and Connected-Worker Solutions for Next-Generation Occupational Safety, Health, and Productivity". In: *Advanced Intelligent Systems* 4 (1 2022-01), p. 2100099. ISSN: 2640-4567. DOI: 10.1002/AISY.202100099. URL: https://onlinelibrary.wiley.com/doi/full/10.1002/aisy.202100099https://onlinelibrary.wiley.com/doi/abs/10.1002/aisy.202100099https://onlinelibrary.wiley.com/doi/10.1002/aisy.202100099 (cit. on p. 4).

[20] *Protocol P155 - Protocol of 2002 to the Occupational Safety and Health Convention, 1981*. URL: https://www.ilo.org/dyn/normlex/en/f?p=NORMLEXPUB:12100:0::NO::P12100_INSTRUMENT_ID:312338 (cit. on p. 4).

[21] N. S. Lop et al. "Work-related to musculoskeletal disorder amongst Malaysian construction trade workers: Bricklayers". In: *AIP Conference Proceedings* 1891 (1 2017-10), p. 020087. ISSN: 0094-243X. DOI: 10.1063/1.5005420. URL: https://aip.scitation.org/doi/abs/10.1063/1.5005420 (cit. on p. 4).

[22] T. R. Hales and B. P. Bernard. "EPIDEMIOLOGY OF WORK-RELATED MUSCULOSKELETAL DISORDERS". In: () (cit. on pp. 5, 7).

[23] Y. Roquelaure. *Musculoskeletal disorders and psychosocial factors at work*, p. 82. ISBN: 9782874525063 (cit. on p. 5).

[24] *Musculoskeletal health*. URL: https://www.who.int/news-room/fact-sheets/detail/musculoskeletal-conditions (cit. on p. 5).

[25] *OSH in figures: Work-related musculoskeletal disorders in the EU - Facts and figures | Safety and health at work EU-OSHA*. URL: https://osha.europa.eu/en/publications/osh-figures-work-related-musculoskeletal-disorders-eu-facts-and-figures (cit. on pp. 5, 6).

[26] Y. Roquelaure et al. "Revue du Praticien (La)". In: *J B Bailliere et Fils* 64 (3 2014), pp. 350–357. URL: https://hal.univ-angers.fr/hal-03391121 (cit. on p. 6).

[27] B. P. Bernard et al. "Musculoskeletal Disorders and Workplace Factors A Critical Review of Epidemiologic Evidence for Work-Related Musculoskeletal Disorders of the Neck, Upper Extremity, and Low Back". In: (1997). URL: http://www.cdc.gov/niosh (cit. on pp. 6, 15).

[28] "Main symptoms include SWELLING TINGLING NUMBNESS PAIN What can be done to prevent MSDs at the workplace?" In: (). DOI: 10.2802/166756 (cit. on p. 6).

[29] Z. S. Maman et al. "A data-driven approach to modeling physical fatigue in the workplace using wearable sensors". In: *Applied Ergonomics* 65 (2017-11), pp. 515–529. ISSN: 0003-6870. DOI: 10.1016/J.APERGO.2017.02.001 (cit. on pp. 6, 17, 48, 50).

[30] J. Dul et al. "A strategy for human factors/ergonomics: Developing the discipline and profession". In: *Ergonomics* 55 (4 2012-04), pp. 377–395. ISSN: 00140139. DOI: 10.1080/00140139.2012.661087 (cit. on p. 7).

[31] *Ergonomics - Environment, Health and Safety*. URL: https://ehs.unc.edu/workplace-safety/ergonomics/ (cit. on p. 7).

[32] *Manual Material Handling Ergonomics | Risk Management Insurance | Colorado State University*. URL: https://rmi.colostate.edu/ergonomics/industrial-ergonomics/manual-material-handling/ (cit. on p. 7).

[33] L. M. Biga et al. *11.1 Describe the roles of agonists, antagonists and synergists*. 2019-09 (cit. on p. 8).

[34] M. Schenkman, V. Rug0, and D. E. Cartaya. "THE JOURNAL OF ORTHOPAEDIC AND SPORTS PHYSICAL THERAPY Kinesiology of the Shoulder Complex". In: (1987). URL: www.jospt.org (cit. on pp. 8, 9).

[35] J. Feger and J. Jones. "Biceps brachii muscle". In: *Radiopaedia.org* (2013-05). DOI: 10.53347/RID-23204 (cit. on p. 9).

[36] *Trapezius Muscle Middle Back Pain, Shoulder Stretches, Trapezius Pain*. URL: https://my.clevelandclinic.org/health/body/21563-trapezius-muscle (cit. on p. 9).

[37] F. Deng and C. Hacking. "Trapezius muscle". In: *Radiopaedia.org* (2015-11). DOI: 10.53347/RID-41023 (cit. on p. 9).

[38] C. J. D. Luca. "MYOELECTRICAL MANIFESTATIONS OF LOCALIZED MUSCULAR FATIGUE IN HUMANS". In: () (cit. on p. 9).

[39] R. V. Hockey. *The psychology of fatigue*. Cambridge University Press, 2013-05 (cit. on p. 9).

[40] E. Simonson and P. C. Weiser. "Psychological aspects and physiological correlates of work and fatigue". In: (1976), p. 445 (cit. on p. 9).

[41] M. Fiorenza et al. "Neuromuscular Fatigue and Metabolism during High-Intensity Intermittent Exercise". In: *Medicine and Science in Sports and Exercise* 51 (8 2019-08), pp. 1642–1652. ISSN: 15300315. DOI: 10.1249/MSS.0000000000001959. URL: https://journals.lww.com/acsm-msse/Fulltext/2019/08000/Neuromuscular_Fatigue_and_Metabolism_during.10.aspx (cit. on p. 9).

[42] P. J. Fadel. "Reflex control of the circulation during exercise". In: *Scandinavian journal of medicine science in sports* 25 Suppl 4 (2015-12), pp. 74–82. ISSN: 1600-0838. DOI: 10.1111/SMS.12600. URL: https://pubmed.ncbi.nlm.nih.gov/26589120/ (cit. on p. 9).

[44] *Normal Variations of the Respiratory System Notes: Diagrams Illustrations | Osmosis*. URL: https://www.osmosis.org/notes/Normal_Variations_of_the_Respiratory_System#page-1 (cit. on p. 9).

[45] U. R. Acharya. *Advances in cardiac signal processing*. Ed. by R. A. U. et al. 2007th ed. Springer, 2007-04 (cit. on p. 10).

[46] *Anatomy and Function of the Heart's Electrical System | Johns Hopkins Medicine*. URL: https://www.hopkinsmedicine.org/health/conditions-and-diseases/anatomy-and-function-of-the-hearts-electrical-system (cit. on p. 10).

[47] *The ECG Leads, Polarity and Einthoven's Triangle – The Student Physiologist*. URL: https://thephysiologist.org/study-materials/the-ecg-leads-polarity-and-einthovens-triangle/ (cit. on p. 10).

[48] F. Shaffer and J. P. Ginsberg. "An Overview of Heart Rate Variability Metrics and Norms". In: *Frontiers in Public Health* 5 (2017-09), p. 258. ISSN: 22962565. DOI: 10.3389/FPUBH.2017.00258/BIBTEX (cit. on pp. 10, 11).

[49] *Lead systems – how an ECG works | CardioSecur*. URL: https://www.cardiosecur.com/magazine/specialist-articles-on-the-heart/lead-systems-how-an-ecg-works (cit. on p. 10).

[50] L. C. M. Vanderlei et al. "Basic notions of heart rate variability and its clinical applicability". In: *Brazilian Journal of Cardiovascular Surgery* 24 (2 2009), pp. 205–217. ISSN: 0102-7638. DOI: 10.1590/S0102-76382009000200018. URL: http://www.scielo.br/j/rbccv/a/Yh54M3tJK4tgWD5PSGcnmPK/abstract/?lang=en (cit. on p. 11).

[51] S. O. Ismaila, K. T. Oriolowo, and O. G. Akanbi. "Cardiovascular strain of sawmill workers in South-Western Nigeria". In: *International journal of occupational safety and ergonomics : JOSE* 19 (4 2013), pp. 607–611. ISSN: 1080-3548. DOI: 10.1080/10803548.2013.11077016. URL: https://pubmed.ncbi.nlm.nih.gov/24321639/ (cit. on pp. 11, 32).

[52] *View of An Ergonomic Evaluation of Douglas Fir Manual Pruning in New Zealand*. URL: https://journals.lib.unb.ca/index.php/IJFE/article/view/30157/1882525459 (cit. on pp. 11, 32).

[53] M. Korshøj et al. "Heart rate during work and heart rate variability during the following night: A day-by-day investigation on the physical activity paradox among blue-collar workers". In: *Scandinavian Journal of Work, Environment and Health* 47 (5 2021), pp. 387–394. ISSN: 1795990X. DOI: 10.5271/SJWEH.3965 (cit. on pp. 11, 32).

[54] D. A. Burton, K. Stokes, and G. M. Hall. "Physiological effects of exercise". In: *Continuing Education in Anaesthesia Critical Care Pain* 4.6 (2004-12), pp. 185–188. ISSN: 1743-1816. DOI: 10.1093/bjaceaccp/mkh050. eprint: https://academic.oup.com/bjaed/article-pdf/4/6/185/894114/mkh050.pdf. URL: https://doi.org/10.1093/bjaceaccp/mkh050 (cit. on p. 11).

[55] S. L. Balasubramaniam et al. "Age-related ranges of respiratory inductance plethysmography (RIP) reference values for infants and children". In: *Paediatric Respiratory Reviews* 29 (2019-02), pp. 60–67. DOI: 10.1016/j.prrv.2018.03.010. URL: https://doi.org/10.1016/j.prrv.2018.03.010 (cit. on pp. 11, 32).

[56] A. Strang et al. "Measures of respiratory inductance plethysmography (RIP) in children with neuromuscular disease". In: *Pediatric Pulmonology* 53 (9 2018-09), pp. 1260–1268. ISSN: 1099-0496. DOI: 10.1002/PPUL.24134. URL: https://onlinelibrary.wiley.com/doi/full/10.1002/ppul.24134https://onlinelibrary.wiley.com/doi/abs/10.1002/ppul.24134https://onlinelibrary.wiley.com/doi/10.1002/ppul.24134 (cit. on p. 12).

[57] *Electromyography (EMG) | Johns Hopkins Medicine*. URL: https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/electromyography-emg (cit. on p. 12).

[58] H. A. Yousif et al. "Assessment of Muscles Fatigue Based on Surface EMG Signals Using Machine Learning and Statistical Approaches: A Review". In: *IOP Conference Series: Materials Science and Engineering* 705.1 (2019-11), p. 012010. DOI: 10.1088/1757-899x/705/1/012010. URL: https://doi.org/10.1088/1757-899x/705/1/012010 (cit. on p. 12).

[59] M. S. Amin, S. T. H. Rizvi, and M. M. Hossain. *A Comparative Review on Applications of Different Sensors for Sign Language Recognition*. 2022-04. DOI: 10.3390/jimaging8040098 (cit. on p. 12).

[60] R. A. Fernandez et al. "Natural user interfaces for human-drone multi-modal interaction". In: *2016 International Conference on Unmanned Aircraft Systems, ICUAS 2016* (2016-06), pp. 1013–1022. DOI: 10.1109/ICUAS.2016.7502665. URL: https://www.researchgate.net/publication/304816895_Natural_user_interfaces_for_human-drone_multi-modal_interaction (cit. on p. 12).

[61] *Machine learning, explained | MIT Sloan*. URL: https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained (cit. on p. 13).

[62] S. Kotsiantis. "Supervised Machine Learning: A Review of Classification Techniques". In: *Informatica* 31 (3 2007), pp. 249–268. ISSN: 1854-3871. URL: https://www.informatica.si/index.php/informatica/article/view/148 (cit. on p. 13).

[63] *"1.10. Decision Trees"*. "en". "https://scikit-learn.org/stable/modules/tree.html". "Accessed: 2022-10-12" (cit. on p. 13).

[64] https://www.datacamp.com/tutorial/random-forests-classifier-python. Accessed: 2022-10-12 (cit. on p. 13).

[65] R. G. Brereton and G. R. Lloyd. "Support Vector Machines for classification and regression". In: *The Analyst* 135.2 (2010), pp. 230–267. DOI: 10.1039/b918972f. URL: https://doi.org/10.1039/b918972f (cit. on p. 13).

[66] *Python Logistic Regression Tutorial with Sklearn Scikit | DataCamp*. URL: https://www.datacamp.com/tutorial/understanding-logistic-regression-python (cit. on p. 13).

[67] P. "Majumder. *"Gaussian Naive Bayes"*. 2020=, howpublished = "https://iq.opengenus.org/gaussian-naive-bayes/", note = "Accessed: 2022-10-12", language = "en" (cit. on p. 13).

[68] E. Allibhai. *Building a k-Nearest-Neighbors (k-NN) model with scikit-learn*. en. https://towardsdatascience.com/building-a-k-nearest-neighbors-k-nn-model-with-scikit-learn-51209555453a. Accessed: 2022-10-12. 2018-09 (cit. on p. 13).

[69] J. Tang, A. Henderson, and P. Gardner. "Exploring AdaBoost and Random Forests machine learning approaches for infrared pathology on unbalanced data sets". In: *The Analyst* 146.19 (2021), pp. 5880–5891. DOI: 10.1039/d0an02155e. URL: https://doi.org/10.1039/d0an02155e (cit. on p. 14).

[70] E. Valero et al. "Analysis of construction trade worker body motions using a wearable and wireless motion sensor network". In: *Automation in Construction* 83 (July 2017), pp. 48–55. ISSN: 09265805. DOI: 10.1016/j.autcon.2017.08.001. URL: http://dx.doi.org/10.1016/j.autcon.2017.08.001 (cit. on p. 15).

[71] *ISO - ISO 11228-1:2003 - Ergonomics — Manual handling — Part 1: Lifting and carrying*. URL: https://www.iso.org/standard/26520.html (cit. on p. 15).

[72] K. Jørgensen, B. R. Jensen, and M. Kato. "Fatigue development in the lumbar paravertebral muscles of bricklayers during the working day". In: *International Journal of Industrial Ergonomics* 8 (3 1991-11), pp. 237–245. ISSN: 0169-8141. DOI: 10.1016/0169-8141(91)90035-K (cit. on p. 16).

[73] P. C. Winwood et al. "Development and Validation of a Scale to Measure Work-Related Fatigue and Recovery: The Occupational Fatigue Exhaustion/Recovery Scale (OFER)". In: *Journal of Occupational and Environmental Medicine* 47.6 (2005-06), pp. 594–606. DOI: 10.1097/01.jom.0000161740.71049.c4. URL: https://doi.org/10.1097/01.jom.0000161740.71049.c4 (cit. on p. 16).

[74] H. J. Michielsen et al. "Examination of the Dimensionality of Fatigue". In: *European Journal of Psychological Assessment* 20.1 (2004-01), pp. 39–48. DOI: 10.1027/1015-5759.20.1.39. URL: https://doi.org/10.1027/1015-5759.20.1.39 (cit. on p. 16).

[75] E. OCCHIPINTI. "OCRA: a concise index for the assessment of exposure to repetitive movements of the upper limbs". In: *Ergonomics* 41.9 (1998-09), pp. 1290–1311. DOI: 10.1080/001401398186315. URL: https://doi.org/10.1080/001401398186315 (cit. on p. 16).

[76] T. R. WATERS et al. "Revised NIOSH equation for the design and evaluation of manual lifting tasks". In: *Ergonomics* 36.7 (1993-07), pp. 749–776. DOI: 10.1080/00140139308967940. URL: https://doi.org/10.1080/00140139308967940 (cit. on p. 16).

[77] S. Hignett and L. McAtamney. "Rapid Entire Body Assessment (REBA)". In: *Applied Ergonomics* 31.2 (2000-04), pp. 201–205. DOI: 10.1016/s0003-6870(99)00039-3. URL: https://doi.org/10.1016/s0003-6870(99)00039-3 (cit. on p. 16).

[78] L. McAtamney and E. N. Corlett. "RULA: a survey method for the investigation of work-related upper limb disorders". In: *Applied Ergonomics* 24.2 (1993-04), pp. 91–99. DOI: 10.1016/0003-6870(93)90080-s. URL: https://doi.org/10.1016/0003-6870(93)90080-s (cit. on p. 16).

[79] M. Bortolini et al. "Motion Analysis System (MAS) for production and ergonomics assessment in the manufacturing processes". In: *Computers &amp Industrial Engineering* 139 (2020-01), p. 105485. DOI: 10.1016/j.cie.2018.10.046. URL: https://doi.org/10.1016/j.cie.2018.10.046 (cit. on p. 16).

[80] K. B. Martinez, M. Nazarahari, and H. Rouhani. "K-score: A novel scoring system to quantify fatigue-related ergonomic risk based on joint angle measurements via wearable inertial measurement units". In: *Applied Ergonomics* 102 (2022-07), p. 103757. ISSN: 0003-6870. DOI: 10.1016/J.APERGO.2022.103757 (cit. on p. 16).

[81] S. Schettino et al. "Forest harvesting in rural properties: Risks and worsening to the worker's health under the ergonomics approach". In: *International Journal of Industrial Ergonomics* 82 (2021-03), p. 103087. DOI: 10.1016/j.ergon.2021.103087. URL: https://doi.org/10.1016/j.ergon.2021.103087 (cit. on p. 16).

[82] J. N. Souza-Talarico et al. "Association between awakening cortisol levels and cardiovascular performance in sugarcane rural workers". In: *Stress and Health* 38.1 (2021-07), pp. 102–110. DOI: 10.1002/smi.3077. URL: https://doi.org/10.1002/smi.3077 (cit. on p. 16).

[83] Z. Arman et al. "Physiological workload evaluation by means of heart rate monitoring during motor-manual clearcutting operations". In: *International Journal of Forest Engineering* 32.2 (2021-01), pp. 91–102. DOI: 10.1080/14942119.2021.1868238. URL: https://doi.org/10.1080/14942119.2021.1868238 (cit. on p. 16).

[84] Z. B. Sibiya, P. Ackerman, and S. A. Ackerman. "Productivity and workload analysis between manual and motor-manual pruning of iPinus patula/i at two different lifts". In: *International Journal of Forest Engineering* 32.3 (2021-04), pp. 191–201. DOI: 10.1080/14942119.2021.1906568. URL: https://doi.org/10.1080/14942119.2021.1906568 (cit. on p. 16).

[85] P. Kirk and R. Parker. "An Ergonomic Evaluation of Douglas Fir Manual Pruning in New Zealand". In: *Journal of Forest Engineering* 7.2 (1996-01), pp. 51–58. DOI: 10.1080/08435243.1996.10702685. URL: https://doi.org/10.1080/08435243.1996.10702685 (cit. on p. 16).

[86] A. Samani et al. "Following ergonomics guidelines decreases physical and cardiovascular workload during cleaning tasks". In: *Ergonomics* 55.3 (2012-03), pp. 295–307. DOI: 10.1080/00140139.2011.640945. URL: https://doi.org/10.1080/00140139.2011.640945 (cit. on p. 16).

[87] I. K. Widana et al. "Anthropometric measures for better cardiovascular and musculoskeletal health". In: *Computer Applications in Engineering Education* 29.3 (2020-09), pp. 550–561. DOI: 10.1002/cae.22202. URL: https://doi.org/10.1002/cae.22202 (cit. on p. 16).

[88] E. Ahonen et al. "The physical strain of dairy farming". In: *Ergonomics* 33.12 (1990-12), pp. 1549–1555. DOI: 10.1080/00140139008925353. URL: https://doi.org/10.1080/00140139008925353 (cit. on p. 17).

[89] S.-J. Lee et al. "Exercise and cardiovascular load in workers with high occupational physical activity". In: *Archives of Environmental &amp Occupational Health* 75.6 (2019-08), pp. 339–345. DOI: 10.1080/19338244.2019.1657059. URL: https://doi.org/10.1080/19338244.2019.1657059 (cit. on p. 17).

[90] L. Silva et al. "Respiratory Inductance Plethysmography to Assess Fatigability during Repetitive Work". In: *Sensors (Basel, Switzerland)* 22 (11 2022-06). ISSN: 1424-8220. DOI: 10.3390/S22114247. URL: https://pubmed.ncbi.nlm.nih.gov/35684868/ (cit. on pp. 17, 32, 35, 48).

[91] M. Aubier. "Respiratory muscle fatigue". In: *Intensive Care Medicine* 15.S1 (1989-02), S17–S20. DOI: 10.1007/bf00260877. URL: https://doi.org/10.1007/bf00260877 (cit. on p. 17).

[92] R. K. Mehta et al. "Comparison of objective and subjective operator fatigue assessment methods in offshore shiftwork". In: *Journal of Loss Prevention in the Process Industries* 48 (2017-07), pp. 376–381. ISSN: 0950-4230. DOI: 10.1016/J.JLP.2017.02.009 (cit. on p. 17).

[93] J. Hu and J. Min. "Automated detection of driver fatigue based on EEG signals using gradient boosting decision tree model". In: *Cognitive Neurodynamics* 12 (4 2018-08), pp. 431–440. ISSN: 18714099. DOI: 10.1007/S11571-018-9485-1/TABLES/2. URL: https://link.springer.com/article/10.1007/s11571-018-9485-1 (cit. on p. 17).

[94] L. H. Du et al. "Detecting driving fatigue with multimodal deep learning". In: *International IEEE/EMBS Conference on Neural Engineering, NER* (2017-08), pp. 74–77. ISSN: 19483554. DOI: 10.1109/NER.2017.8008295 (cit. on p. 17).

[95] M. Darbandy et al. "A new approach to detect the physical fatigue utilizing heart rate signals". In: *Research in Cardiovascular Medicine* 9 (1 2020), p. 23. ISSN: 2251-9572. DOI: 10.4103/rcm.rcm_8_20 (cit. on pp. 17, 49, 50).

[96] F. Nasirzadeh et al. "Physical Fatigue Detection Using Entropy Analysis of Heart Rate Signals". In: *Sustainability 2020, Vol. 12, Page 2714* 12 (7 2020-03), p. 2714. ISSN: 2071-1050. DOI: 10.3390/SU12072714. URL: https://www.mdpi.com/2071-1050/12/7/2714/htmhttps://www.mdpi.com/2071-1050/12/7/2714 (cit. on pp. 17, 49, 50).

[97] A. Lambay et al. "A Data-Driven Fatigue Prediction using Recurrent Neural Networks". In: *HORA 2021 - 3rd International Congress on Human-Computer Interaction, Optimization and Robotic Applications, Proceedings* (2021-06). DOI: 10.1109/HORA52670.2021.9461377 (cit. on pp. 17, 48, 50).

[98] W. Umer et al. "Physical exertion modeling for construction tasks using combined cardiorespiratory and thermoregulatory measures". In: *Automation in Construction* 112 (2020-04), p. 103079. ISSN: 0926-5805. DOI: 10.1016/J.AUTCON.2020.103079 (cit. on pp. 17, 48–50).

[99] A. Aryal, A. Ghahramani, and B. Becerik-Gerber. "Monitoring fatigue in construction workers using physiological measurements". In: *Automation in Construction* 82 (2017-10), pp. 154–165. ISSN: 0926-5805. DOI: 10.1016/J.AUTCON.2017.03.003 (cit. on pp. 17, 49, 50).

[100] A. Baghdadi et al. "A machine learning approach to detect changes in gait parameters following a fatiguing occupational task". In: *https://doi.org/10.1080/00140139.2018.1442936* 61 (8 2018-08), pp. 1116–1129. ISSN: 13665847. DOI: 10.1080/00140139.2018.1442936. URL: https://www.tandfonline.com/doi/abs/10.1080/00140139.2018.1442936 (cit. on pp. 17, 48, 50).

[101] A. B. Oliveira, L. C. Silva, and H. J. Coury. "Como a variação de altura e massa da carga afetam os movimentos do membro superior durante o manuseio de caixas industriais?" In: *Revista Brasileira de Fisioterapia* 15 (6 2011-11), pp. 494–502. ISSN: 14133555. DOI: 10.1590/S1413-35552011005000017 (cit. on p. 19).

[102] *PLUX Biosignals*. URL: https://www.pluxbiosignals.com/ (cit. on p. 20).

[103] *Accelerometer (ACC) Sensor Data Sheet The biosignalsplux tri-axial Accelerometer (ACC) uses Micro-Electro-Mechanical*. 2018. URL: http://biosignalsplux.com/ (cit. on p. 22).

[104] "Electromyography (EMG) Sensor Data Sheet Muscle". In: (2017). URL: http://biosignalsplux.com/ (cit. on p. 22).

[105] Ecg. "Electrocardiography (ECG) Sensor Data Sheet". In: (2015). URL: http://biosignalsplux.com/ (cit. on p. 22).

[106] "Inductive Respiration (RIP) Sensor Data Sheet RIP 27042015". In: (2015). URL: http://biosignalsplux.com/ (cit. on p. 22).

[107] *NeuroKit/ecg$_f$indpeaks.pyatmasterneuropsychology/NeuroKitGitHub*. URL: https://github.com/neuropsychology/NeuroKit/blob/master/neurokit2/ecg/ecg_findpeaks.py (cit. on pp. 23, 24).

[108] *biosignalsnotebooks/detect.py at master · pluxbiosignals/biosignalsnotebooks · GitHub*. URL: https://github.com/pluxbiosignals/biosignalsnotebooks/blob/master/biosignalsnotebooks/biosignalsnotebooks/detect.py (cit. on pp. 23, 24).

[109] M. S. Manikandan and K. P. Soman. "A novel method for detecting R-peaks in electrocardiogram (ECG) signal". In: *Biomedical Signal Processing and Control* 7 (2 2012-03), pp. 118–128. ISSN: 1746-8094. DOI: 10.1016/J.BSPC.2011.03.004 (cit. on pp. 23, 24).

[110] M. Z. Suboh et al. "Shannon energy application for detection of ECG R-peak using bandpass filter and stockwell transform methods". In: *Advances in Electrical and Computer Engineering* 20 (3 2020), pp. 41–48. ISSN: 18447600. DOI: 10.4316/AECE.2020.03005 (cit. on pp. 23, 28).

[111] W. Xu and F. Du. "A ROBUST QRS COMPLEX DETECTION METHOD BASED ON SHANNON ENERGY ENVELOPE AND HILBERT TRANSFORM". In: *Journal of Mechanics in Medicine and Biology* 22 (3 2022-04). ISSN: 02195194. DOI: 10.1142/S0219519422400139/ASSET/IMAGES/LARGE/S0219519422400139FIGF7.JPEG. URL: www.worldscientific.com (cit. on pp. 23, 24).

[112] T. Thurner et al. "Complex-Pan-Tompkins-Wavelets: Cross-channel ECG beat detection and delineation". In: *Biomedical Signal Processing and Control* 66 (2021-04), p. 102450. ISSN: 1746-8094. DOI: 10.1016/J.BSPC.2021.102450 (cit. on p. 24).

[113] P. S. Addison. "Wavelet transforms and the ECG: a review". In: *Physiological Measurement* 26.5 (2005-08), R155–R199. DOI: 10.1088/0967-3334/26/5/r01. URL: https://doi.org/10.1088/0967-3334/26/5/r01 (cit. on p. 24).

[114] *R Wave Detection in the ECG - MATLAB Simulink*. URL: https://www.mathworks.com/help/wavelet/ug/r-wave-detection-in-the-ecg.html (cit. on p. 24).

[115] *GitHub - pistonly/modwtpy: modwt in python*. URL: https://github.com/pistonly/modwtpy (cit. on p. 24).

[116] *Maximal overlap discrete wavelet transform - MATLAB modwt*. URL: https://www.mathworks.com/help/wavelet/ref/modwt.html (cit. on p. 24).

[117] *PyWavelets - Wavelet Transforms in Python — PyWavelets Documentation*. URL: https://pywavelets.readthedocs.io/en/latest/ (cit. on p. 24).

[118] D. Zhang et al. "An ECG Signal De-Noising Approach Based on Wavelet Energy and Sub-Band Smoothing Filter". In: (2019). DOI: 10.3390/app9224968. URL: www.mdpi.com/journal/applsci (cit. on p. 27).

[119] A. A. Kostina, P. M. Tzvetkov, and A. N. Serov. "Investigation of the Method of RMS Measurement Based on Moving Averaging". In: *2020 55th International Scientific Conference on Information, Communication and Energy Systems and Technologies, ICEST 2020 - Proceedings* (2020-09), pp. 235–238. DOI: 10.1109/ICEST49890.2020.9232779 (cit. on p. 28).

[120] $scipy.signal.find_peaks_SciPyv1.9.1Manual$. URL: https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.find_peaks.html (cit. on p. 31).

[121] Y.-C. Chen and T.-C. Hsiao. "Instantaneous phase difference analysis between thoracic and abdominal movement signals based on complementary ensemble empirical mode decomposition". In: *BioMedical Engineering OnLine* 15.1 (2016-10). DOI: 10.1186/s12938-016-0233-7. URL: https://doi.org/10.1186/s12938-016-0233-7 (cit. on p. 32).

[122] D Faithfull, J. G. Jones, and C Jordan. "MEASUREMENT OF THE RELATIVE CONTRIBUTIONS OF RIB CAGE AND ABDOMEN/DIAPHRAGM TO TIDAL BREATHING IN MAN". In: *British Journal of Anaesthesia* 51 (1979), pp. 391–398. DOI: 10.1093/bja/51.5.391 (cit. on p. 32).

[123] R. Deering and J. F. Kaiser. "The use of a masking signal to improve Empirical Mode Decomposition". In: vol. IV. Institute of Electrical and Electronics Engineers Inc., 2005, pp. 485–488. ISBN: 0780388747. DOI: 10.1109/ICASSP.2005.1416051 (cit. on p. 33).

[124] D. Makowski et al. *NeuroKit2: A Python Toolbox for Neurophysiological Signal Processing*. 2020. DOI: 10.5281/ZENODO.3597887. URL: https://github.com/neuropsychology/NeuroKit (cit. on p. 33).

[125] J. T. Dennerlein et al. "Fatigue in the Forearm Resulting From Low-Level Repetitive Ulnar Deviation". In: *AIHA Journal* 64.6 (2003), pp. 799–805. DOI: 10.1080/15428110308984875. eprint: https://doi.org/10.1080/15428110308984875. URL: https://doi.org/10.1080/15428110308984875 (cit. on p. 35).

[126] J. Paulus, M. Müller, and A. Klapuri. "Audio-based music structure analysis". In: *In Proceedings of the 11th International Conference on Music Information Retrieval (ISMIR*. 2010, pp. 625–636 (cit. on p. 35).

[127] *C4S2$_S$SM*. URL: https://www.audiolabs-erlangen.de/resources/MIR/FMP/C4/C4S2_SSM.html (cit. on p. 35).

[128] J. Paulus and A. Klapuri. *AUDIO-BASED MUSIC STRUCTURE ANALYSIS*. URL: https://www.audiolabs-erlangen.de/content/05-fau/professor/00-mueller/03-publications/2010_PaulusMuellerKlapuri_STAR-MusicStructure_ISMIR.pdf (cit. on p. 35).

[129] A. Phinyomark, P. Phukpattaranont, and C. Limsakul. "Feature reduction and selection for EMG signal classification". In: *Expert Systems with Applications* 39 (8 2012-06), pp. 7420–7431. ISSN: 0957-4174. DOI: 10.1016/J.ESWA.2012.01.102 (cit. on p. 38).

[130] *Welcome to TSFEL documentation! — TSFEL 0.1.4 documentation*. URL: https://tsfel.readthedocs.io/en/latest/ (cit. on p. 38).

[131] *tsfel.feature_extraction.features — TSFEL 0.1.4 documentation*. en. https://tsfel.readthedocs.io/en/latest/_modules/tsfel/feature_extraction/features.html. Accessed: 2022-10-12 (cit. on p. 38).

[132] *tsfel.utils package — TSFEL 0.1.4 documentation*. URL: https://tsfel.readthedocs.io/en/latest/descriptions/modules/tsfel.utils.html (cit. on p. 38).

[133] *1.13. Feature selection — scikit-learn 1.1.2 documentation*. URL: https://scikit-learn.org/stable/modules/feature_selection.html (cit. on p. 38).

[134] *sklearn.feature$_s$election.SelectKBest¯scikit−learn*1.1.2*documentation*. URL: https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.SelectKBest.html#sklearn.feature_selection.SelectKBest (cit. on p. 38).

[135] S. H. Abed et al. "A modern deep learning framework in robot vision for automated bean leaves diseases detection". In: *International Journal of Intelligent Robotics and Applications* 5.2 (2021-04), pp. 235–251. DOI: 10.1007/s41315-021-00174-3. URL: https://doi.org/10.1007/s41315-021-00174-3 (cit. on p. 39).

[136] *sklearn.model$_s$election.LeaveOneOut¯scikit−learn*1.1.2*documentation*. URL: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.LeaveOneOut.html (cit. on p. 39).

[137] *4.2. Permutation feature importance — scikit-learn 1.1.2 documentation*. URL: https://scikit-learn.org/stable/modules/permutation_importance.html (cit. on p. 39).

[138] *sklearn.metrics.roc$_c$urve¯scikit−learn*1.1.2*documentation*. URL: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.roc_curve.html#sklearn.metrics.roc_curve (cit. on p. 40).

I

R Peaks Detetion

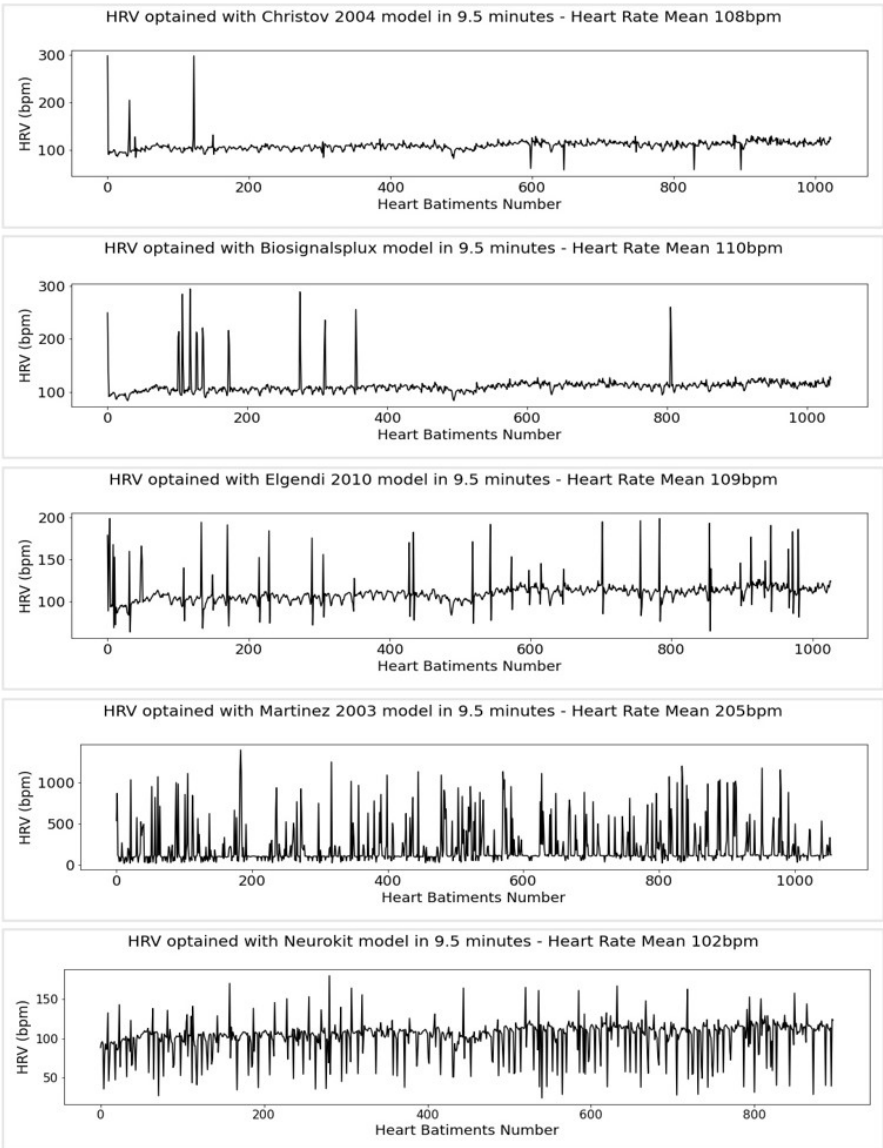

Figure I.1: These graphs represent the HRV of the same ECG signal calculated by some tested algorithms for R peak detection - Part I

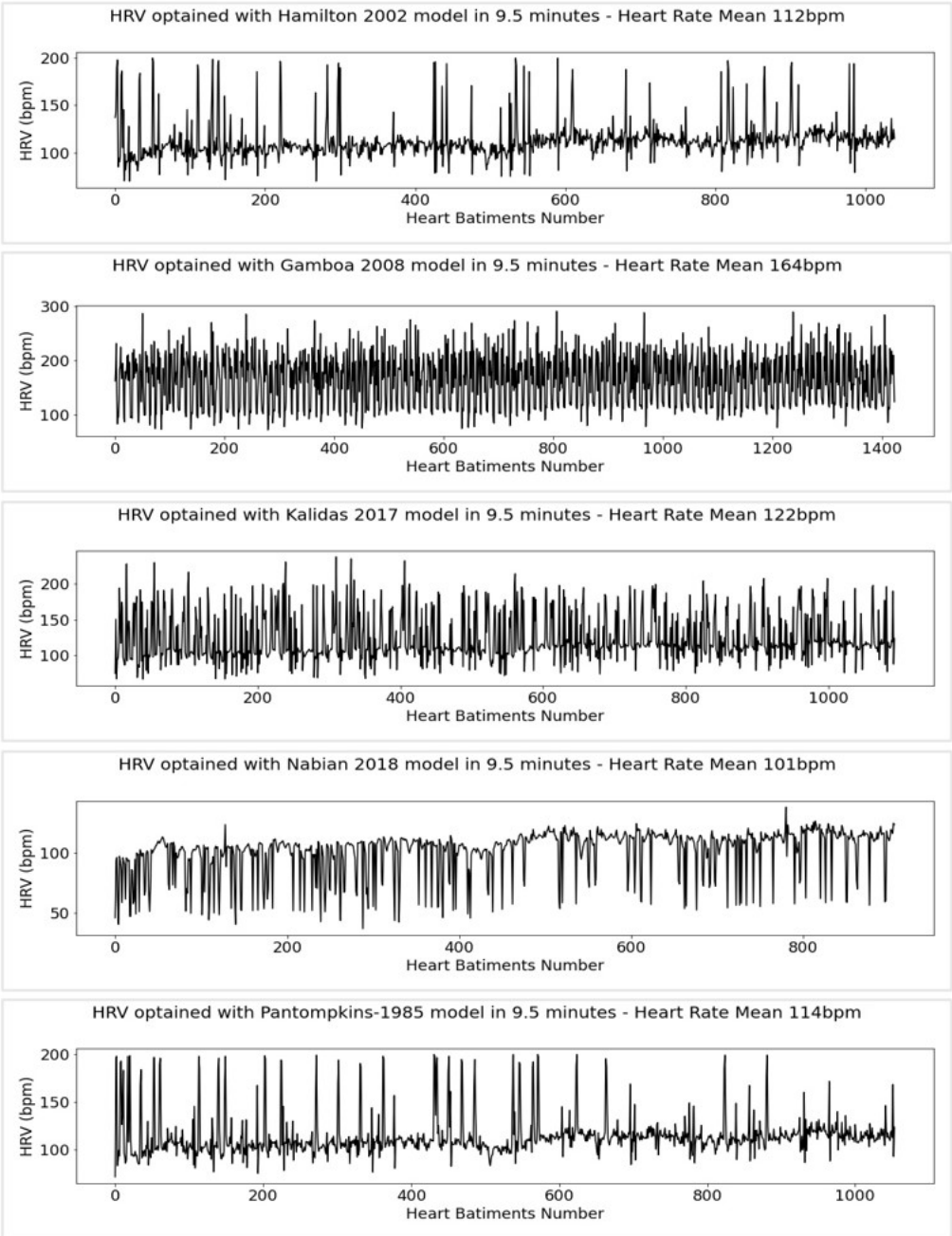

Figure I.2: These graphs represent the HRV of the same ECG signal calculated by some tested algorithms for R peak detection - Part II

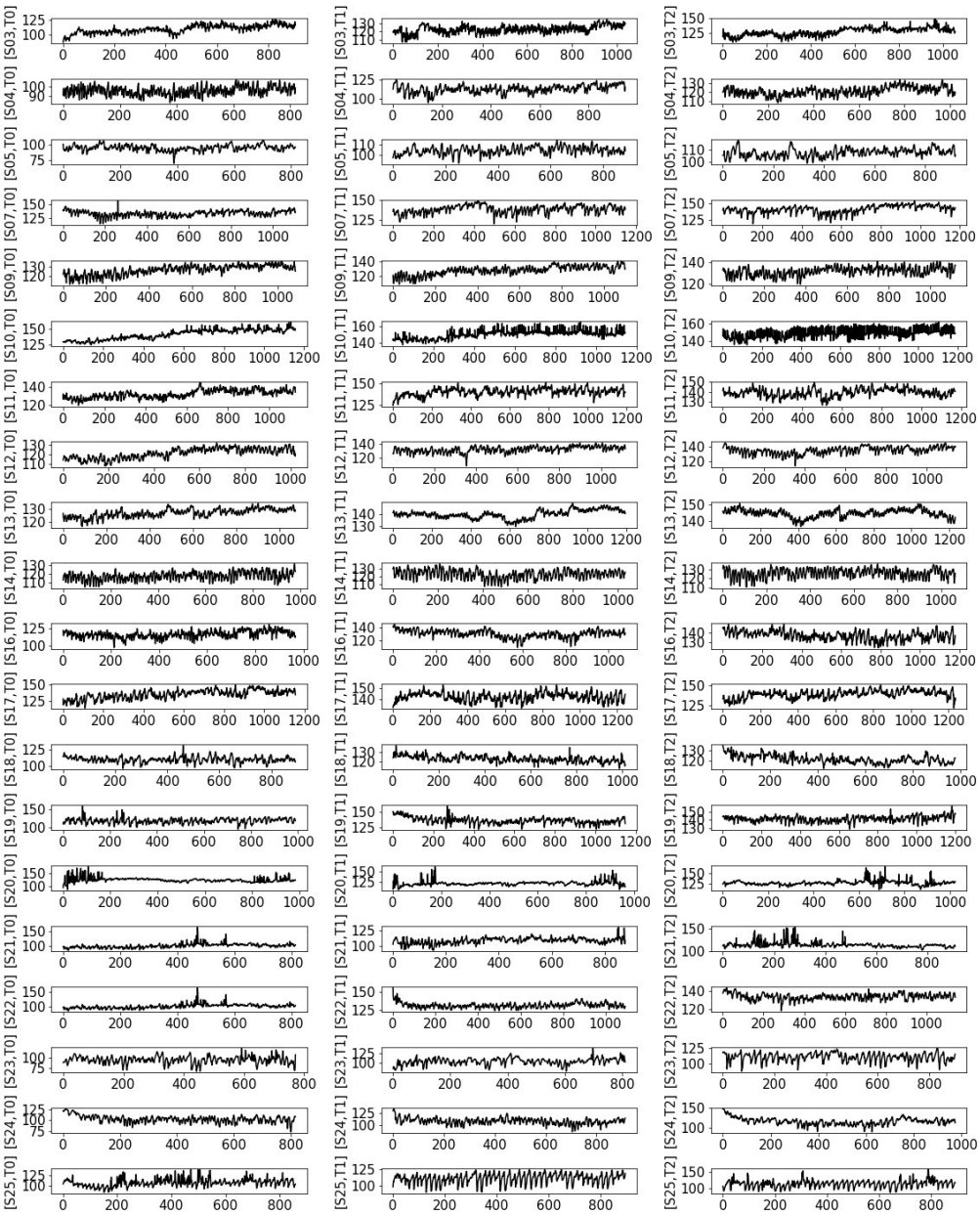

Figure I.3: HRV signal of various subjects and tasks obtained here after processing.

II

Segmentation

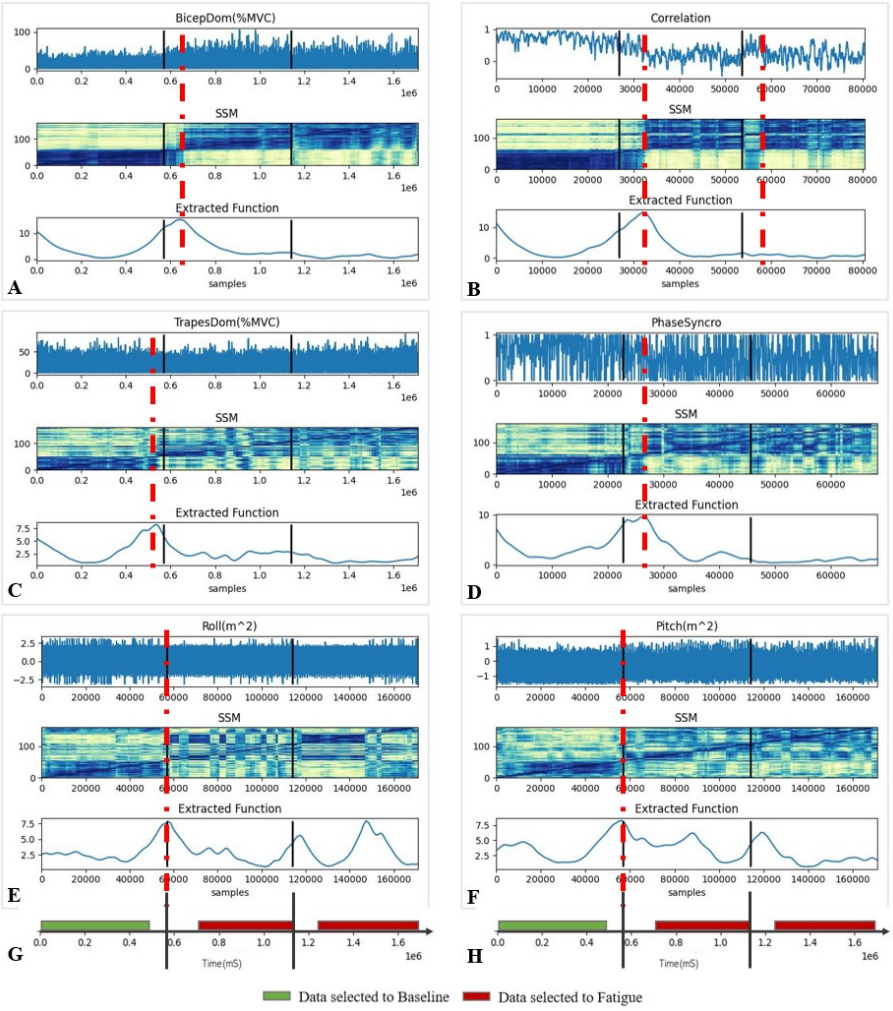

Figure II.1: Another example of SSM targeting a subject how is made in Figure 4.19

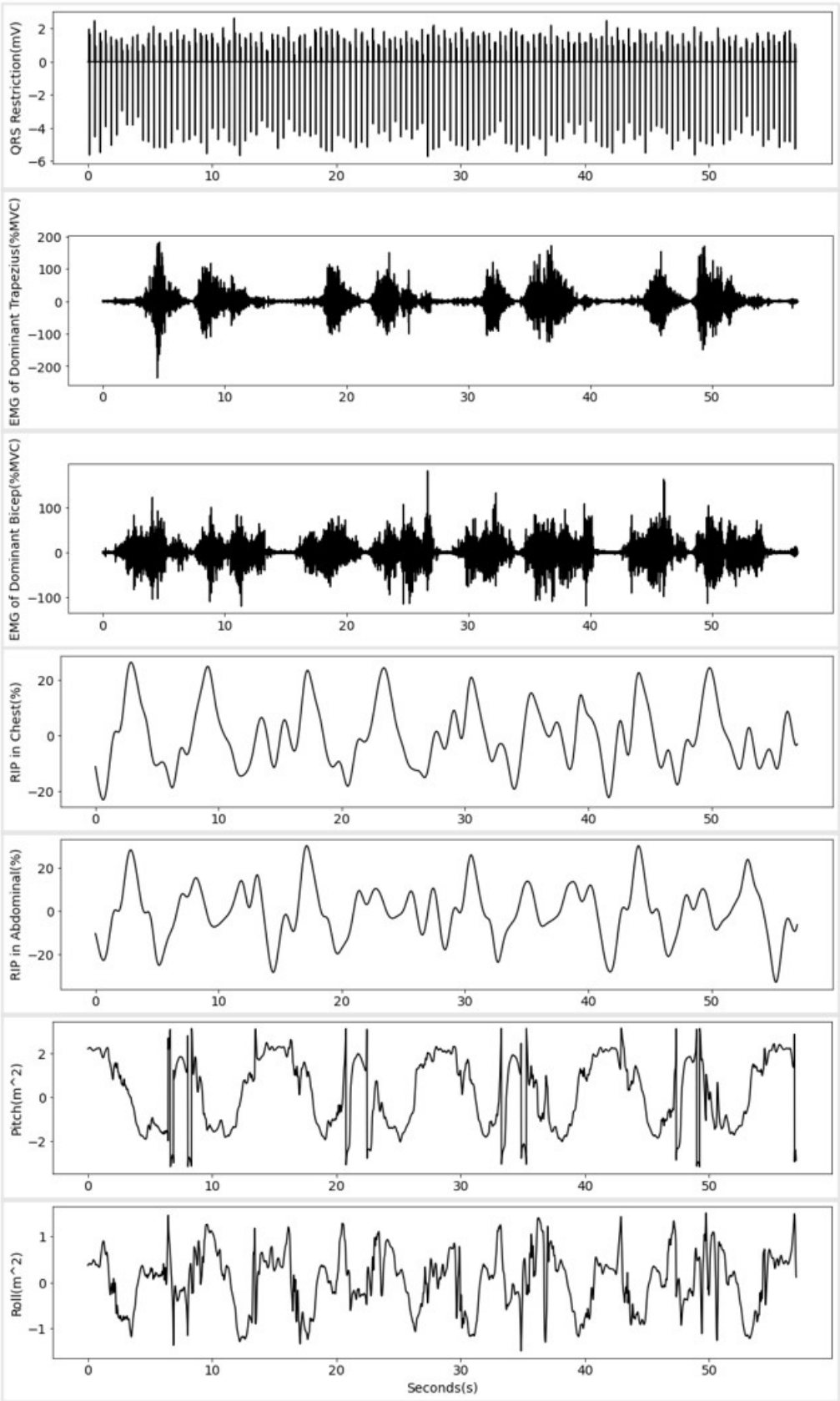

Figure II.2: All segmented signals from a sample.

III

Features Selection

Table III.1: Features selected considering each type of signal after the exclusion of features by correlation and variance

Chi2 Score	**ECG Signals**	**EMG Signals**
1º	CVL-Root mean square	Bicep-Autocorrelation
2º	CVL-Mean	Trapezius-Autocorrelation
3º	HRV-Mean	Trapezius-Median
4º	RHR-Mean	Bicep-Kurtosis
5º	CVS-Mean	Trapezius-Mean
6º	CVS-Variance	Trapezius-Amplitude
7º	HRV-Variance	Bicep-Interquartile range
8º	CVS-Interquartile range	Bicep-Skewness
9º	HRV-Mean absolute deviation	Bicep-Median
10º	HRV-Interquartile range	Bicep-Amplitude
11º	RHR-Mean absolute deviation	Bicep-Mean
12º	RHR-Interquartile range	Trapezius-Kurtosis
13º	HRV-Kurtosis	Trapezius-Interquartile range
14º	HRV-Skewness	-
Chi2 Score	**RIP Signals**	**ACC Signals**
1º	Phase Syncr.-Autocorrelation	Pitch-Autocorrelation
2º	Correlation-Autocorrelation	Roll-Autocorrelation
3º	Phase Syncr.-Kurtosis	Roll-Mean absolute deviation
4º	Correlation-Kurtosis	Pitch-Root mean square
5º	%RC-Skewness	Roll-Interquartile range
6º	PhaseSyn-Skewness	Pitch-Mean absolute deviation
7º	Correlation-Skewness	Roll-Kurtosis
8º	%RC-Interquartile range	Roll-Median
9º	%RC-Mean	Roll-Mean
10º	%RC-Kurtosis	Pitch-Skewness
11º	%RC-Median	Pitch-Interquartile range
12º	%RC-Amplitude	Pitch-Kurtosis
13º	%RC-Autocorrelation	Pitch-Median
14º	-	Pitch-Mean
15º	-	Pitch-Median absolute deviation
16º	-	Roll-Skewness

IV

Models Development

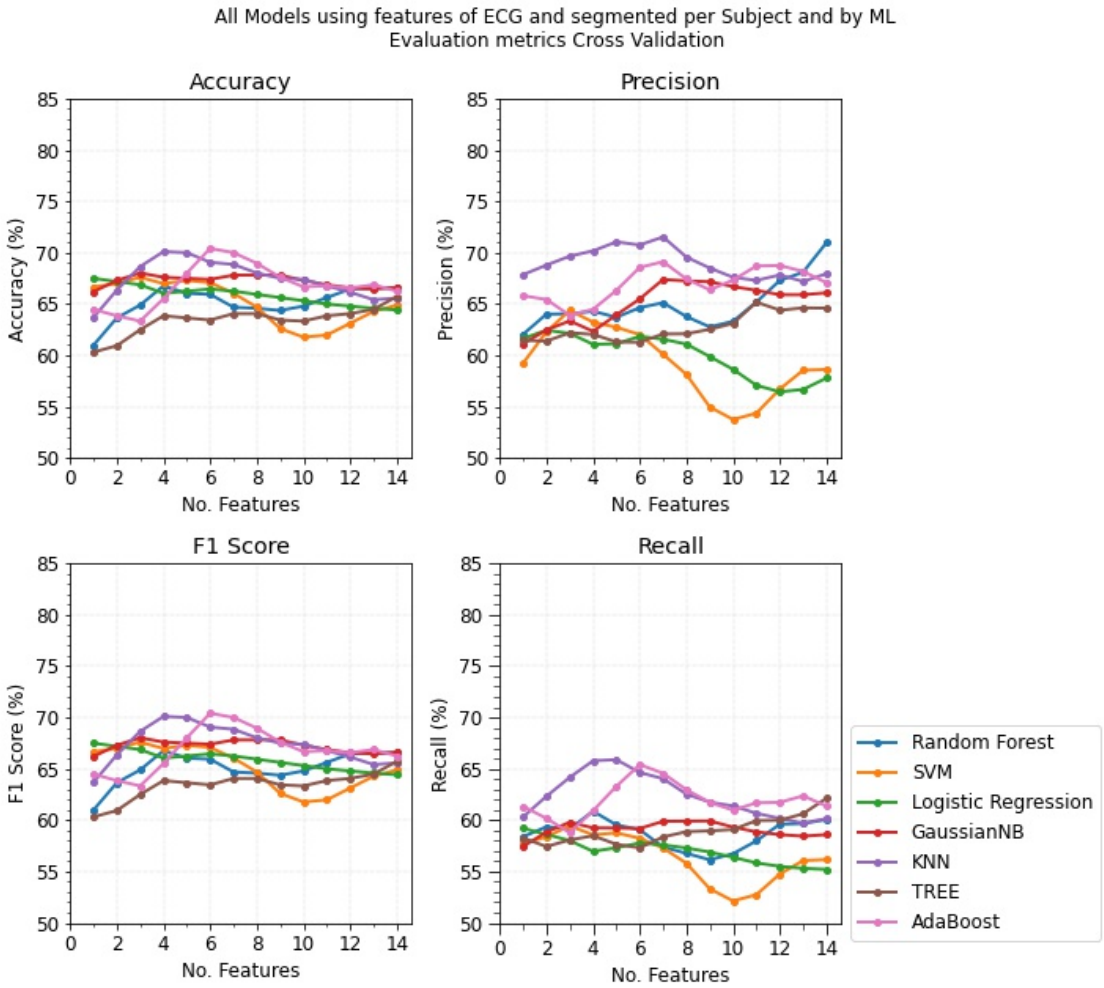

Figure IV.1: Evaluation metrics by all algorithms of classifiers trained with Leave-one-out Cross Validation and with ECG data set.

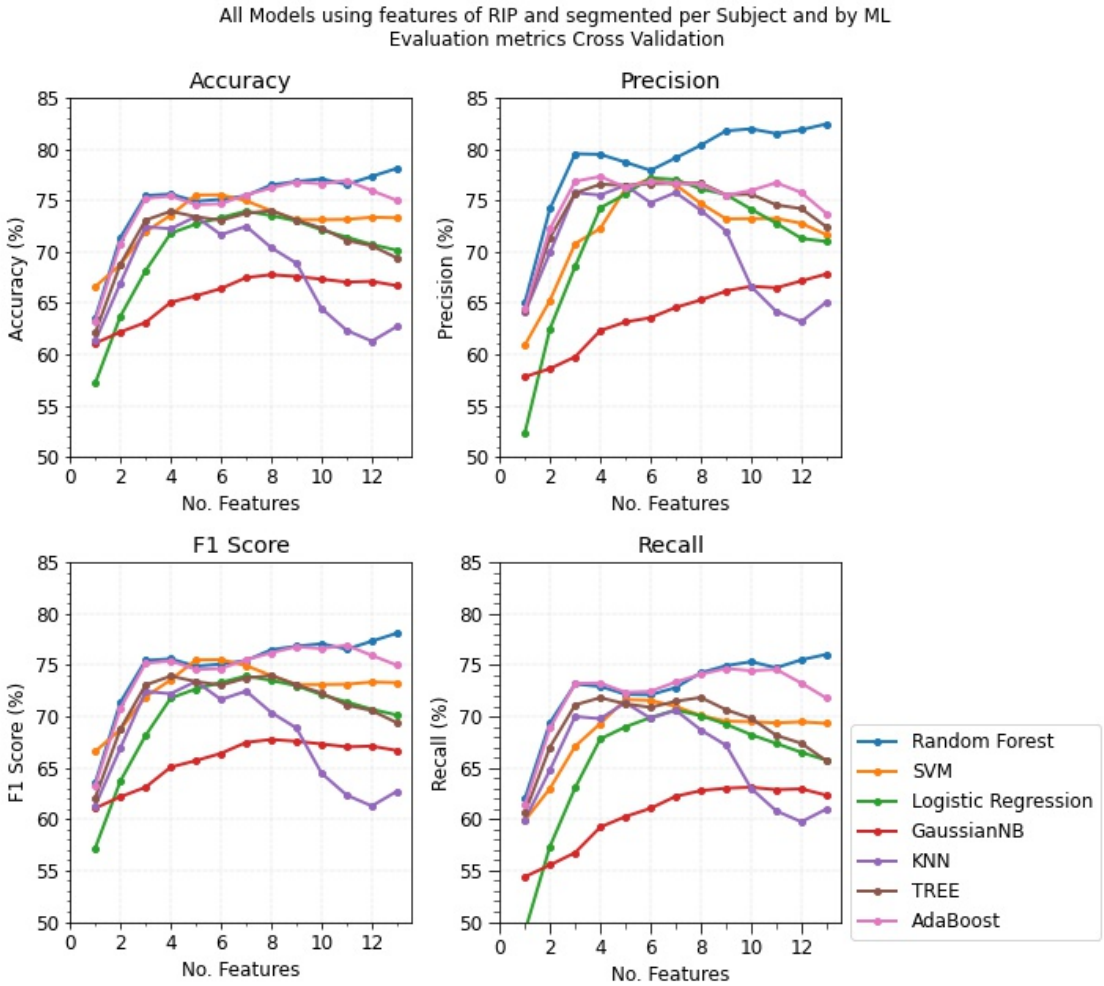

Figure IV.2: Evaluation metrics by all algorithms of classifiers trained with Leave-one-out Cross Validation and with RIP data set.

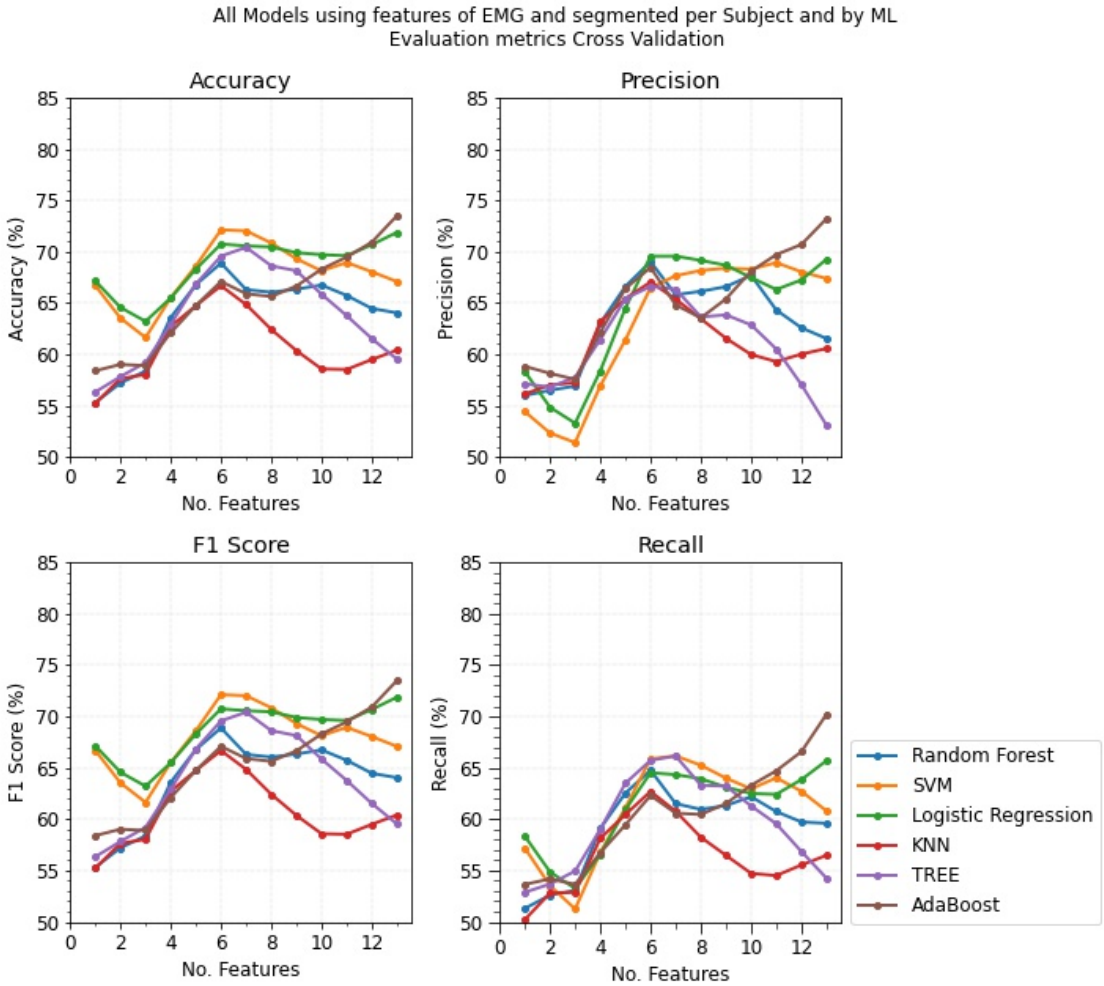

Figure IV.3: Evaluation metrics by all algorithms of classifiers trained with Leave-one-out Cross Validation and with EMG data set.

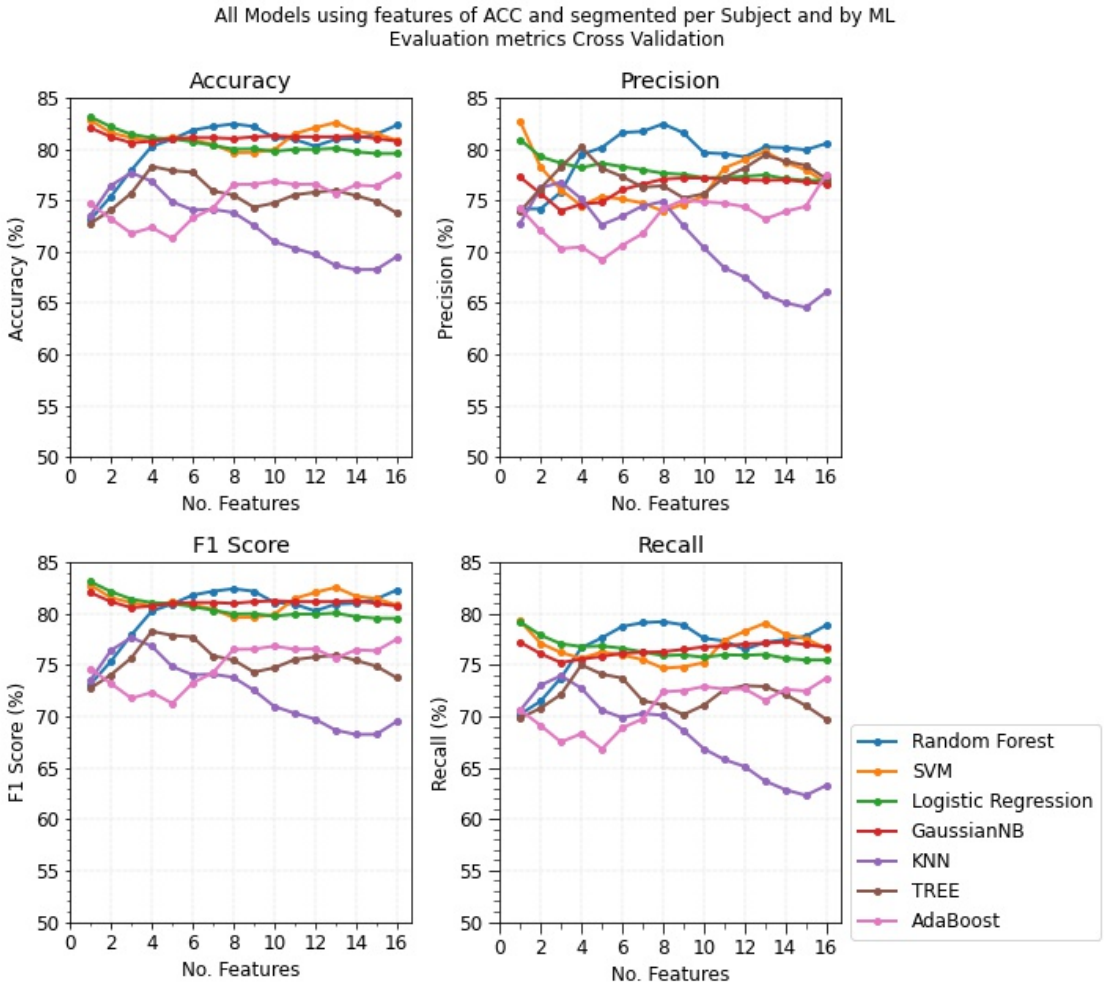

Figure IV.4: Evaluation metrics by all algorithms of classifiers trained with Leave-one-out Cross Validation and with ACC data set.

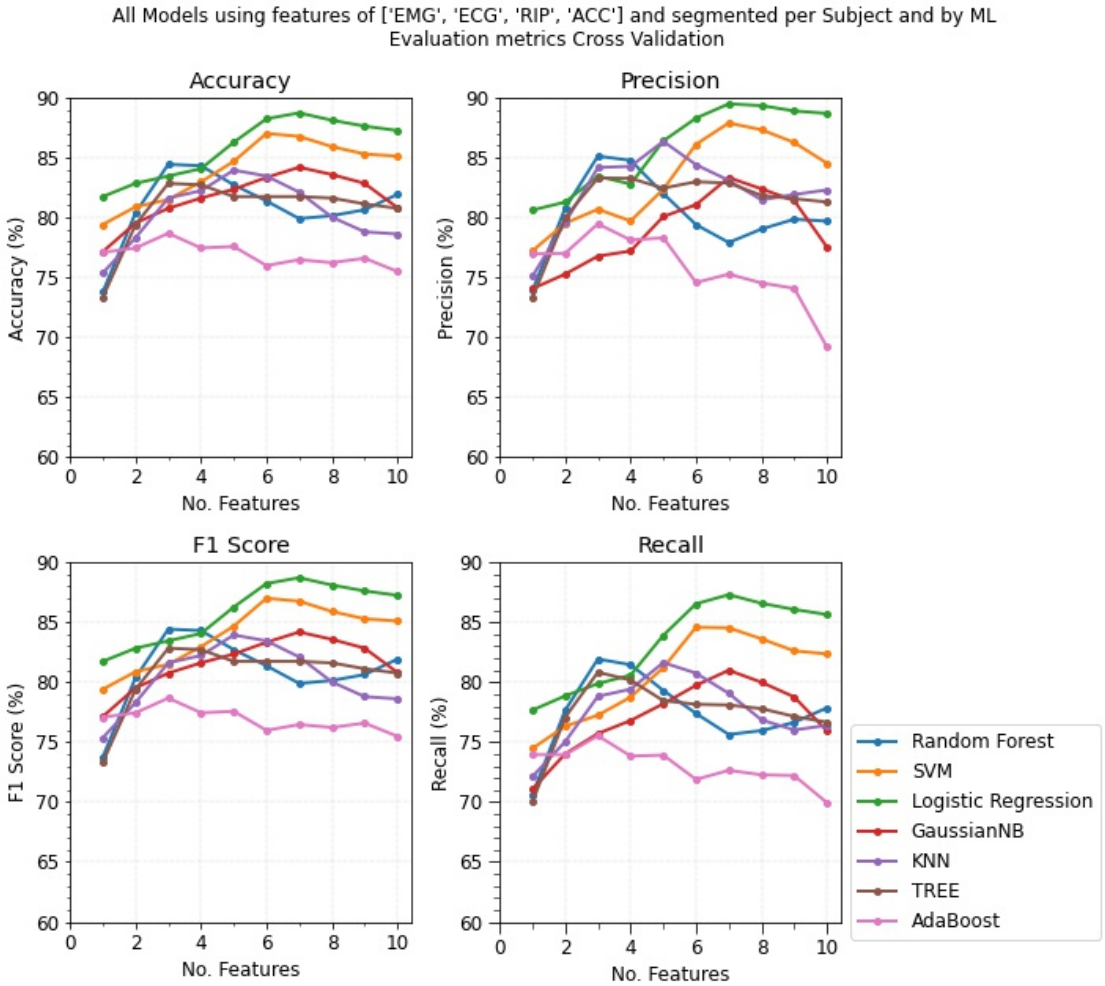

Figure IV.5: Evaluation metrics by all algorithms of classifiers trained with Leave-one-out Cross Validation and with All data sets.

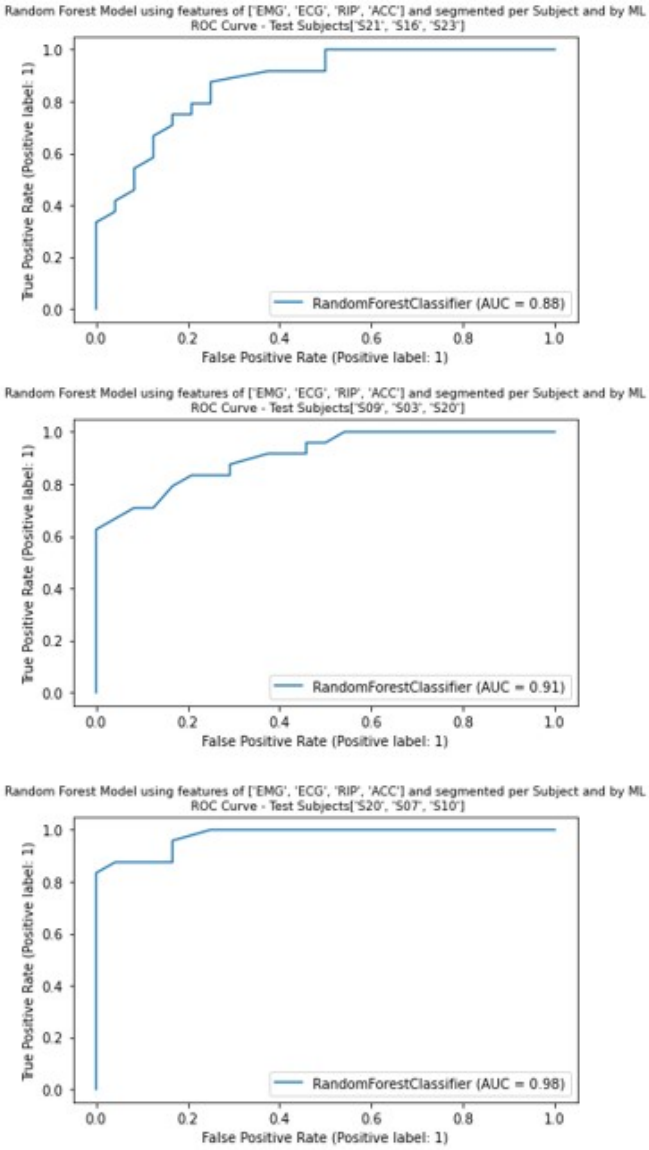

Figure IV.6: Receiver Operating Characteristic (ROC) curve for the Random Forest model that uses all signal types. For the various graphs different test subjects were used to evaluate the classifier. The area under the curve for different test data was revealed to be a good model.

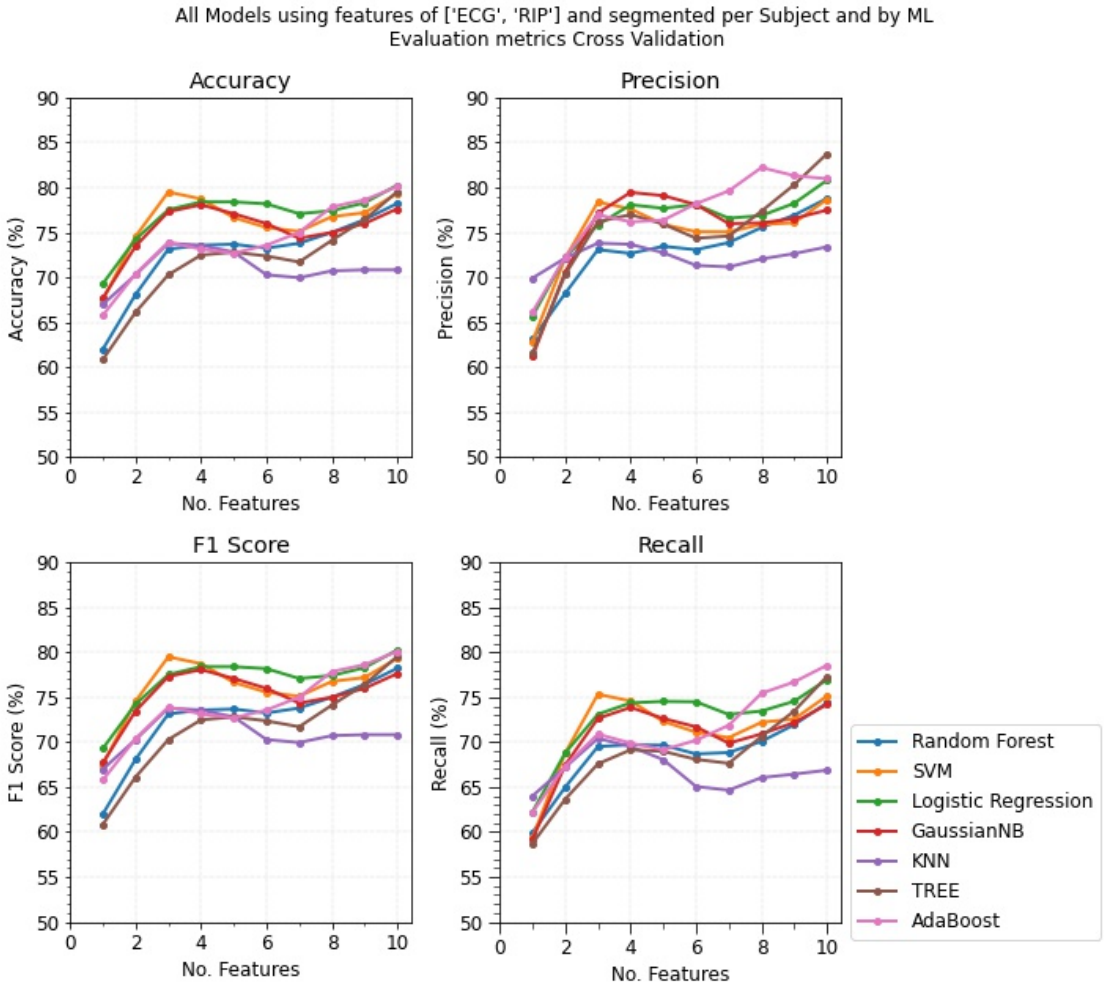

Figure IV.7: Evaluation metrics by all algorithms of classifiers trained with Leave-one-out Cross Validation and with ECG and RIP data sets.

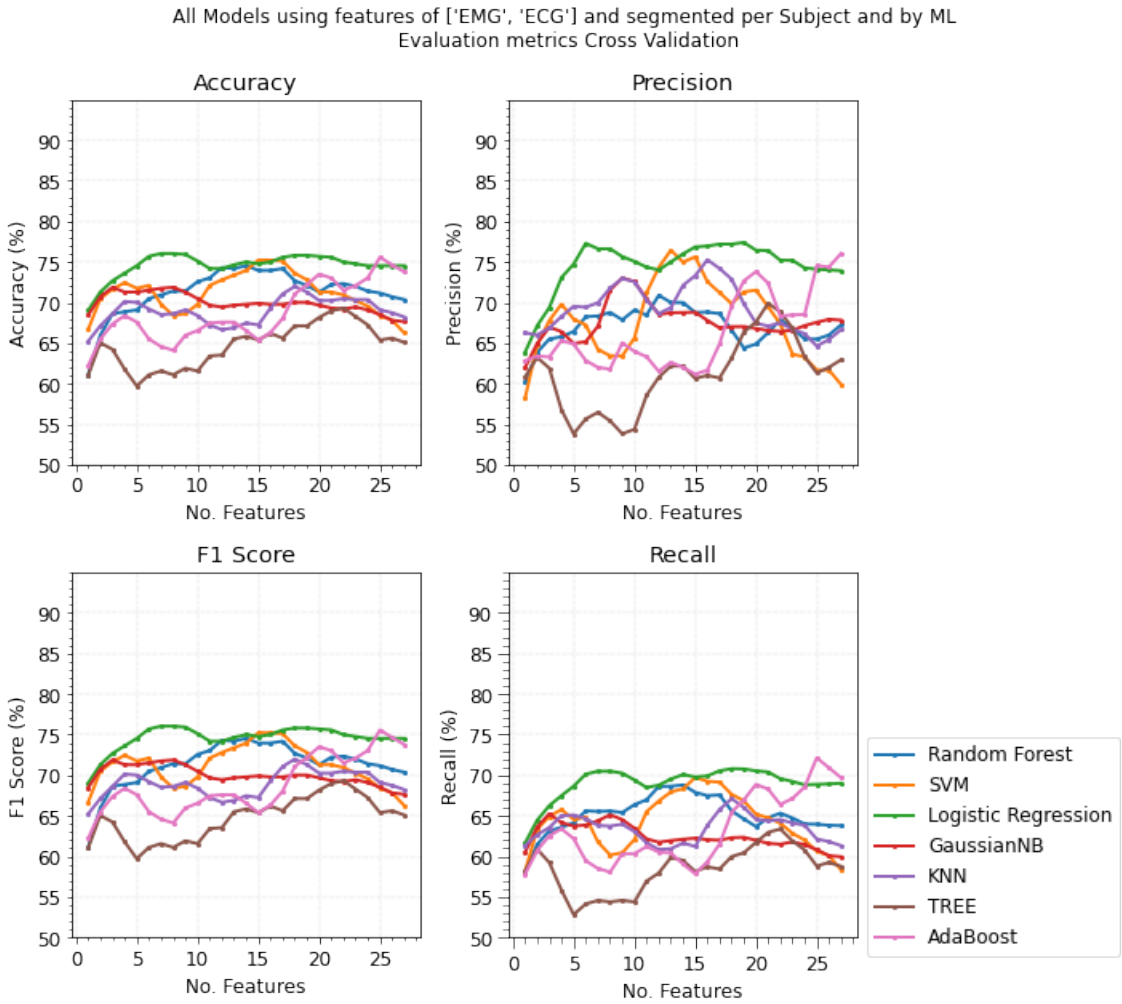

Figure IV.8: Evaluation metrics by all algorithms of classifiers trained with Leave-one-out Cross Validation and with EMG and ECG data sets.

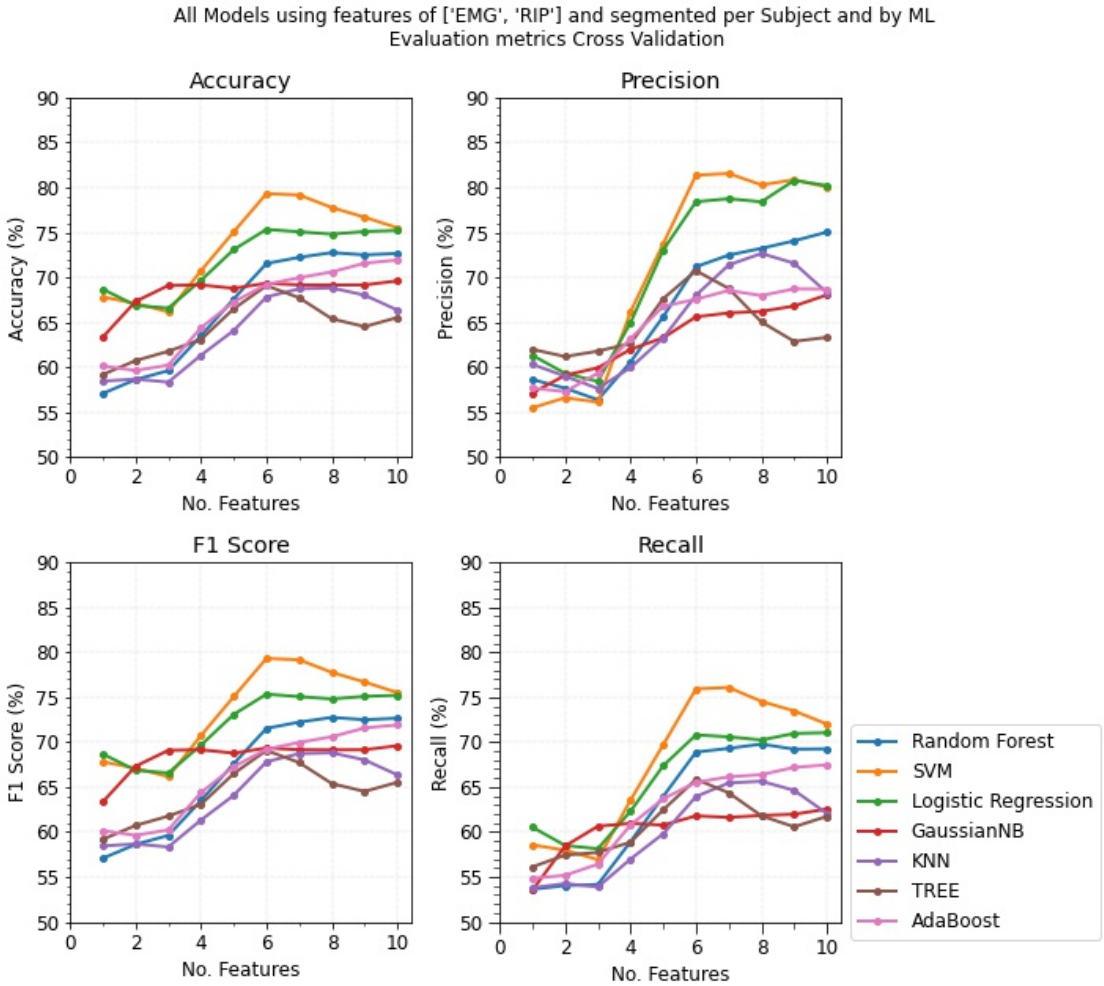

Figure IV.9: Evaluation metrics by all algorithms of classifiers trained with Leave-one-out Cross Validation and with EMG and RIP data sets.

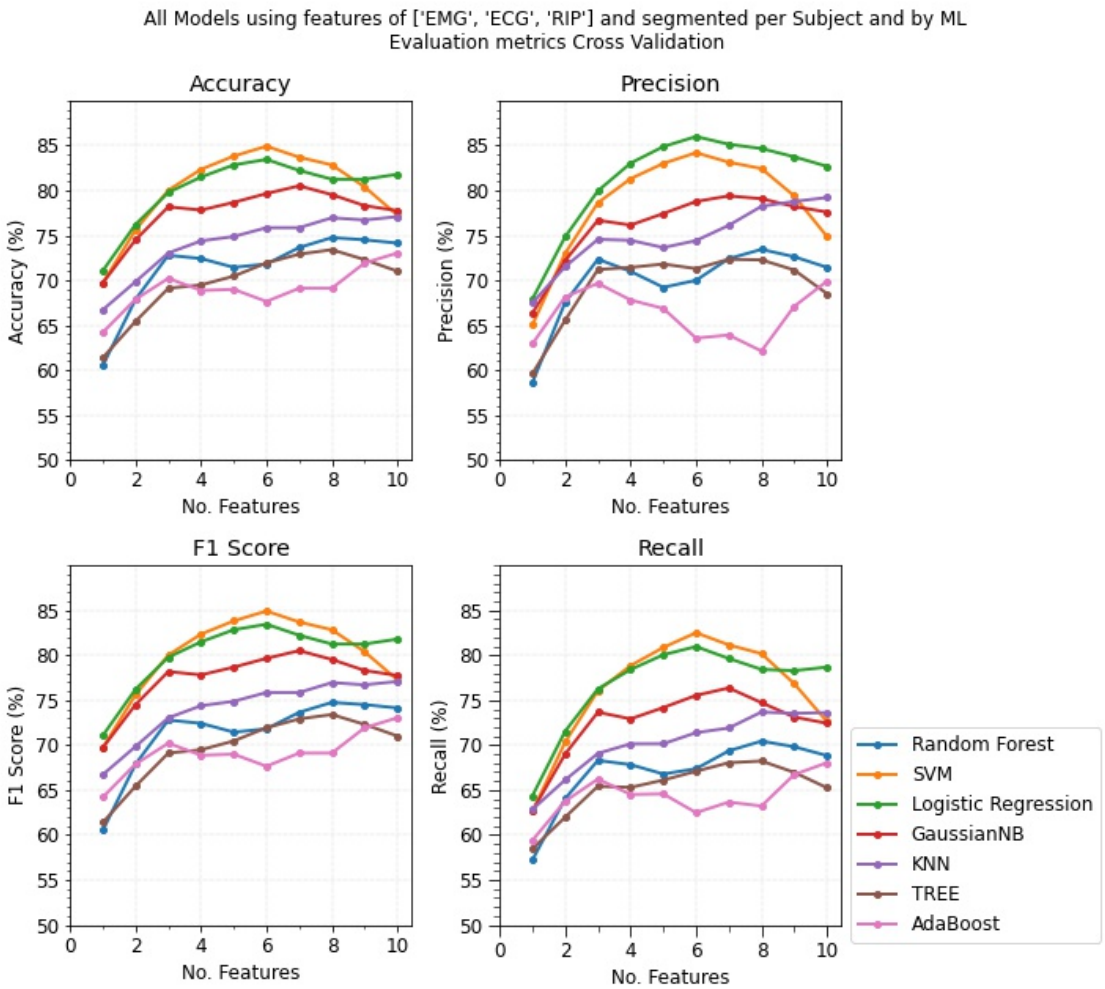

Figure IV.10: Evaluation metrics by all algorithms of classifiers trained with Leave-one-out Cross Validation and with EMG, ECG and RIP data sets.

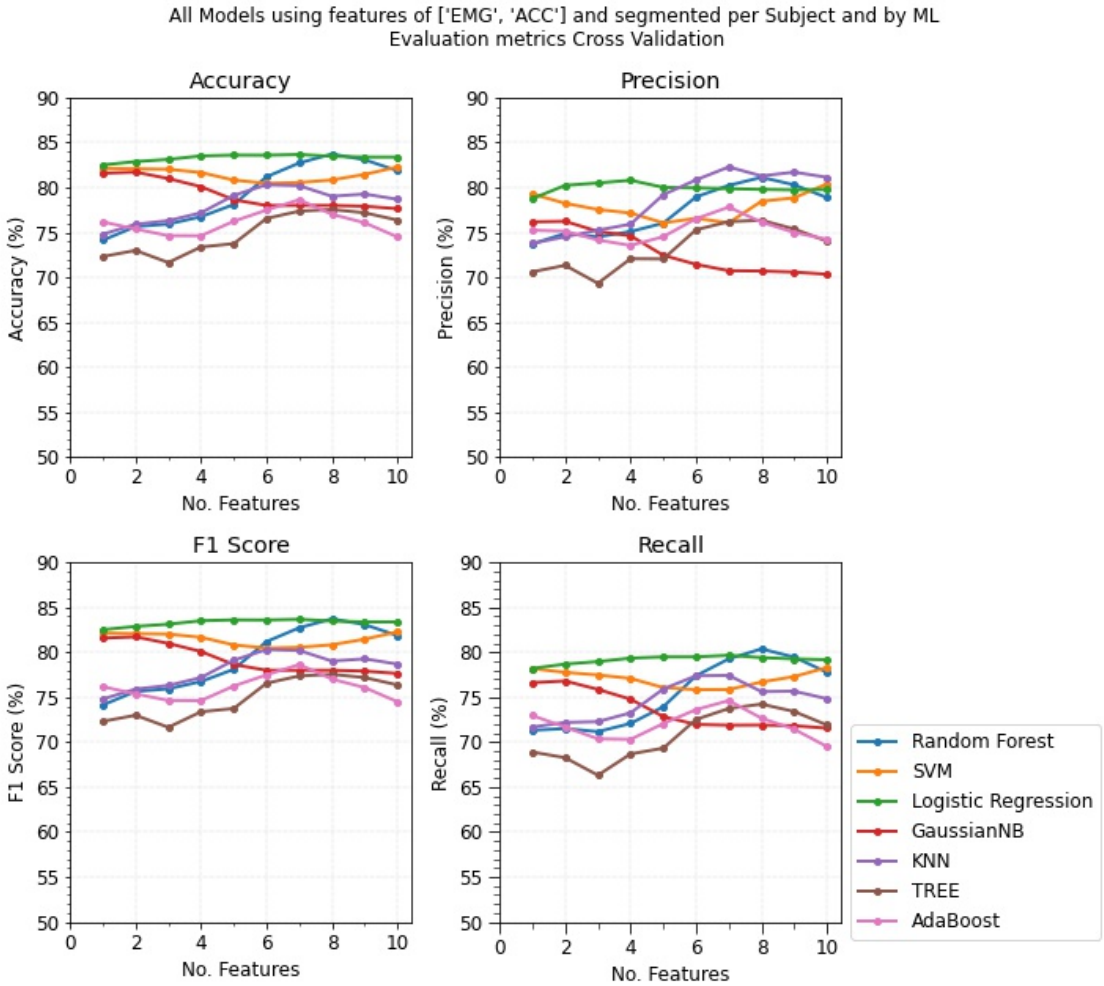

Figure IV.11: Evaluation metrics by all algorithms of classifiers trained with Leave-one-out Cross Validation and with EMG and ACC data sets.