**Universidade do Minho**
Escola de Engenharia

Ana Rita Oliveira Antunes

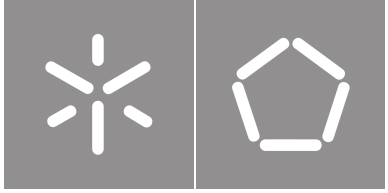
# Drowsy Driving Monitorization Using Statistical and Machine Learning Techniques

July, 2024

**Drowsy Driving Monitorization Using Statistical and Machine Learning Techniques**

Ana Rita O. Antunes

UMinho | 2024



**Universidade do Minho**
Escola de Engenharia

Ana Rita Oliveira Antunes

# Drowsy Driving Monitorization Using Statistical and Machine Learning Techniques

Doctorate Thesis
Doctorate in Doctoral Program in Industrial and Systems Engineering

Work developed under the supervision of:
**Prof.ª Ana Cristina da Silva Braga**
**Prof.º Joaquim José de Almeida Soares Gonçalves**

July, 2024

This document was created with the (pdf/Xe/Lua)LaTeX processor and the NOVAthesis template (v7.1.24) [103].

# Acknowledgements

I would like to express my deepest gratitude to all those who made the completion of this thesis possible. Undertaking this project was not just an academic requirement, it was an opportunity to increase my understanding and deepened my interest in various fields of engineering. This journey has been one of profound learning and personal growth. This made me much more curious, pushed me to go deeper and seek out more information, and constantly challenged and expanded the boundaries of my knowledge.

However, like any worthwhile endeavor, this journey came with its fair share of challenges. There were highs and lows, moments of doubt, and periods of breakthrough that have all shaped this experience. Through these phases, I have learned not only about the complexities of engineering but also about resilience and the pursuit of excellence.

I am immensely thankful to my advisors, Prof.ª Ana Cristina Braga and Prof.º Joaquim J. Gonçalves, and faculty members whose guidance was invaluable. Their insights and feedback have been crucial in bringing this work to completion. I must also extend my thanks to my classmates and colleagues who supported me, offering encouragement and camaraderie throughout this process. The exchanges I had at various conferences have been invaluable. Meeting peers and experts who shared their insights and experiences generously has enriched my perspective and underscored the collaborative spirit of the academic community. I am deeply thankful to all those I encountered for their openness and the wisdom they imparted. Additionally, I am immensely grateful to my team, affectionately known as the *Dream Team*. Your constant support and friendship have meant a lot to me throughout this project.

I also want to present my gratitude to my family (Maria Araújo, João Antunes and Gonçalo Antunes) and boyfriend (Filipe Silva), who provided unwavering support, love, and patience throughout this journey. Your belief in my abilities and your constant encouragement kept me motivated. Thank you for not letting me fall.

Lastly, but certainly not least, I extend my heartfelt gratitude to Ana Pereira. Your unwavering support, guidance, and encouragement have been instrumental in my journey. I am deeply thankful for the invaluable assistance you provided and for helping me discover my path.

## STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the Universidade do Minho.

# Abstract


## Drowsy Driving Monitorization Using Statistical and Machine Learning Techniques

Sleep is crucial to people's health and well-being, that is essential for cognitive function, emotional regulation, and physical health. Therefore, the quality of sleep directly influences daily performance, influencing our emotional stability and memory consolidation to decision making. Despite its importance, inadequate sleep contributes to an array of health problems and diminished quality of life. Consequently, sleep-related issues impact not only driver safety but also that of passengers, pedestrians, and other road users. This thesis seeks to comprehend sleep disorders among Portuguese drivers, with the objective of addressing prevailing knowledge gaps across all districts. A questionnaire-based approach, covering sleep disorders provided valuable insights. Among the findings, a significant portion of drivers reported poor sleep quality (60.2%) with a subset experiencing excessive daytime sleepiness (38.8%). Additionally, the study addresses the alignment between circadian rhythms and work schedules, acknowledging its impact on productivity. The analysis reveals that while the majority of drivers have work schedules aligned with their natural circadian rhythms, those whose schedules diverge from this alignment tend to experience increased daytime sleepiness. Furthermore, the study aims to improve road safety by creating affordable solutions that seamlessly integrate into driving routines, with a specific focus on combating drowsiness while driving. Driving simulations were conducted, and data were collected using a wearable device (Empatica E4), from which heart rate variability data was acquired. This data was then utilized for the classification and prediction of drowsiness. The study addressed challenges associated with subjective drowsiness classification (awake or drowsy) and used multivariate statistical process control techniques to improve the reliability of these classifications. Therefore, a comprehensive analysis revealed promising results, with the Ensemble Tree (ET) model emerging as the best classifier. Additionally, regression models, particularly the XGBoost (XGB), exhibited the ability to predict drowsiness with a lead time of two minutes, outperforming the ET model. Notably, the superior volume of data on drowsy events raised concerns regarding the model's ability to recognize wakefulness accurately. After adding new data, the model was able to correctly distinguish between those who are awake and those who are sleepy. These findings are significant and promising in terms of their potential to predict drowsiness in advance.



**Keywords:** drowsiness at the wheel, feature selection, machine learning, sleep disorders, statistic

# Resumo

## Monitorização da Sonolência ao Volante Com a Utilização de Técnicas Estatísticas e de Machine Learning

O sono é importante na saúde e no bem-estar das pessoas, sendo fundamental para funções cognitivas, regulação emocional e saúde física. A qualidade do sono influencia o desempenho diário, afetando a estabilidade emocional e a consolidação da memória para a tomada de decisões, que podem contribuir para problemas de saúde e diminuição da qualidade de vida. Consequentemente, impactam a segurança dos condutores, assim como a de passageiros, peões e outros utilizadores da estrada. Deste modo, esta tese analisa os distúrbios do sono, nos condutores portugueses, procurando preencher lacunas existentes na literatura, e que tenha em consideração os distritos de Portugal. O desenvolvimento de um questionário, sobre distúrbios de sono, forneceu informações valiosas, em que uma parte significativa apresenta má qualidade do sono (60,2%) e, 38,8% tem sonolência excessiva diurna. Além disso, o estudo compara os ritmos circadianos e os horários de trabalho e resultou que na maioria dos condutores existe sincronização. No entanto, nos restantes condutores foi identificada a prevalência de sonolência excessiva diurna. Nesse sentido, o estudo visa melhorar a segurança rodoviária, criando soluções acessíveis que se integrem facilmente nas rotinas de condução, abordando especificamente o problema da sonolência ao volante. Através de simulações de condução e recolha de dados usando um dispositivo inteligente (Empatica E4), foram obtidos dados da variabilidade do batimento cardíaca e utilizados para classificação e previsão de sonolência. O estudo abordou desafios na classificação subjetiva de sonolência (acordado ou sonolento) e enfatizou a importância das técnicas de controlo estatístico multivariado para melhorar as classificações. Através de uma análise analítica foram atingidos resultados promissores, onde o modelo *Ensemble Tree* (Extremely Randomized Trees (ET)) obteve os melhores resultados. O modelo de regressão *XGBoost* (eXtreme Gradient Boosting (XGB)) demonstrou a capacidade de prever a sonolência dois minutos antes, superando o modelo ET. Com mais informações sobre eventos sonolentos, surgiu a preocupação de o modelo não estar a distinguir o estado acordado de sonolento. Após a adição de novos dados, o modelo foi capaz de distinguir corretamente entre aqueles que estão acordados e aqueles que estão sonolentos. Assim, estes resultados são bastante promissores para prever a sonolência antecipadamente.

**Palavras-chave:** aprendizagem automática, distúrbios de sono, estatística, seleção das variáveis, sonolência ao volante

# Contents

# List of Figures

# List of Tables

# Acronyms

**AdaBoost** Adaptive Boosting *(pp. 40, 63, 74, 95)*

**BMI** Body Mass Index *(pp. 63, 65)*

**csi** cardiac ssympathetic index *(p. 12)*
**CV** Coefficient of Variation *(pp. 61, 71)*
**cvi** cardiac vagal index *(p. 12)*

**ECG** Electrocardiogram *(pp. 9, 17, 18, 21)*
**EEG** Electroencephalogram *(pp. 9, 10)*
**EMG** Electromyogram *(p. 10)*
**EOG** Electrooculogram *(pp. 9, 10)*
**ESS** Epworth Sleepiness Scale *(pp. 14, 46, 47, 49, 53, 54, 59, 92)*
**ET** Extremely Randomized Trees *(pp. vi, 38, 39, 63, 74–78, 83–86, 88, 89, 95)*

**FN** False Negatives *(pp. 27, 28)*
**FP** False Positives *(pp. 27, 28)*

**HF** High-Frequency *(pp. 11, 12)*

**kNN** K-Nearest Neighbor *(pp. 19, 21, 37, 38, 63, 74, 75, 77, 78, 95)*
**KSS** Karolinska Sleepiness Scale *(pp. 13, 16, 60, 66, 70, 88)*

**LDA** Linear Discriminant Analysis *(pp. 34, 63, 74, 95)*
**LF** Low-Frequency *(pp. 11, 12)*
**LR** Logistic Regression *(pp. 32, 33, 74, 95)*

**MAE** Mean Absolute Error *(pp. 29, 83, 89)*
**ME** Morningness-Eveningness *(pp. 15, 59)*
**MESSi** Morningness-Eveningness-Stability-Scale improved *(pp. 16, 46, 47, 49, 53, 92, 94)*
**MSE** Mean Square Error *(pp. 29, 42, 83, 89)*
**MSQ** Mini Sleep Quality *(p. 15)*

**NB** Naïve Bayes *(pp. 19, 35–37, 63, 74, 75, 77, 78, 95)*
**NREM** Non-Rapid Eye Movement *(pp. 8, 9)*

**PCA** Principal Component Analysis *(pp. 17, 43–45, 61, 67, 68, 88)*
**PERCLOS** Percentage of time Eyelids Closure *(p. 19)*
**PPG** Photoplethysmography *(pp. 10, 16–18, 21, 61)*
**PSQI** Pittsburgh Sleep Quality Index *(pp. 46, 47, 49, 51–53, 59, 92)*

**REM** Rapid Eye Movement *(pp. 8, 9)*
**RMSE** Root Mean Square Error *(pp. 29, 83, 89)*

**SPE** Squared Prediction Error *(pp. 44, 62, 68, 69, 71)*
**SQS** Sleep Quality Scale *(p. 15)*
**SSQ** Subjective Sleep Quality *(p. 52)*
**SVM** Support Vector Machine *(pp. 18, 19, 21, 30–32, 40, 63, 74, 75, 77, 78, 88, 95)*
**SWAI** Sleep-Wake Activity Inventory *(p. 14)*
**SWS** Slow-Wave Sleep *(p. 10)*

**TN** True Negative *(pp. 27, 28)*
**TP** True Positive *(pp. 27, 28)*

**ULC** Upper Limit Control *(p. 44)*
**ULF** Ultra-Low-Frequency *(p. 11)*

**VLF** Very-Low-Frequency *(p. 11)*

**XGB** eXtreme Gradient Boosting *(pp. vi, 41, 42, 63, 74–78, 83–87, 89, 95)*

# 1

# Introduction

*"The more I learn, the more I realize how much I don't know." By Albert Einstein.*

This section offers a comprehensive overview of the research, delving into its underlying motivation, articulating the precisely defined objectives of the PhD thesis, elucidating the employed methodology, highlighting significant contributions made to the scientific community, and providing a clear outline of the thesis structure.

## 1.1 Research Motivation

Understanding the importance of sleep is essential to comprehend the complex nature of human health. Sleep is fundamental for both physical and psychological health, and it has an important role in shaping cognitive processes, emotional regulation, and sustained attention. The way these systems interact provides the basis for essential functions such as learning, memory consolidation, long-term productivity, and concentration, as evidenced by research studies [145, 89].

The consequences of insufficient sleep extend far beyond individual experiences and affect safety, health, and general quality of life in larger domains. Insufficient sleep has been unequivocally linked to diminished productivity and an increased risk of workplace accidents, particularly in contexts such as driving, as elucidated in the study [145]. This underscores the societal significance of sleep deprivation, urging a closer assessment of its effects on individual and communal well-being. Moreover, the profound impact of sleep disturbances extends into the realm of public health, being associated with various diseases. Research, exemplified by the study [46], has underscored the association between sleep disruption, including an increased risk of hypertension, diabetes, obesity, depression, heart attacks, and strokes. Recognizing sleep disturbance as a significant public health challenge accentuates the urgency to unravel its intricate connections, thereby fostering an approach to individual and societal well-being. In conjunction

with the broader understanding of sleep patterns, two studies [10, 32] focused on the sleep quality and habits of the general Portuguese population. The participant selection process for these studies involving Portuguese adults (1119 participants) was not explicitly detailed. Nevertheless, the findings revealed significant aspects of sleep patterns and disorders in the population. A noteworthy 28.2% of participants were identified as having sleep disorders, while a substantial 54.8% reported experiencing poor sleep quality. The sleep difficulties index underscored that a majority of participants encountered sleep-related issues, with a discernible trend of worsening as individuals aged. Concerning sleep duration, the study reported that 62.9% of participants slept between 6 to 8 hours per day, and 44% required 1 to 15 minutes to initiate sleep. Furthermore, a substantial portion (51.4%) faced difficulties falling asleep, while 75.3% of adults reported spontaneous awakening. Additionally, a significant 80.5% experienced wakefulness during the night. These findings collectively contribute to a nuanced understanding of sleep patterns and challenges within the general Portuguese population, emphasizing the need for tailored interventions and awareness programs to address the identified sleep-related issues. Adding to the complexity of sleep dynamics, the circadian rhythm and work schedule emerge as factors to consider. The circadian rhythm serves as our internal twenty-four-hour clock, orchestrating the timing of waking and sleeping preferences. Beyond these basic functions, it intricately regulates various rhythmic patterns, including emotions, moods, body temperature, metabolic rate, and hormone releases. Each individual possesses a unique circadian rhythm, where some experience a vigil peak in the morning, characterizing them as morning types who tend to perform optimally during daylight hours. Conversely, eveningness individuals find their peak productivity later in the day, with a preference for sleeping and waking up later [167]. However, when the circadian rhythm lacks coordination, adverse consequences unfold, affecting rest cycles, physiological functions, and individual behavior. This lack of synchronization can lead to sleep disorders, particularly when influenced by work schedules that disrupt the natural alignment of the circadian rhythm [175].

Shifting to the road safety context, a comprehensive study in 2014, conducted by [33], investigated the prevalence of excessive daytime sleepiness among Portuguese truck drivers, involving 714 participants. This research aimed not only to quantify the prevalence but also to identify driving habits associated with drowsiness that could potentially contribute to road accidents. The urgency of such investigation was underscored by Portugal's higher-than-average road accident mortality rate, ranking seventh in Europe at that time. The findings revealed alarming statistics, with 85.7% of participants admitting to driving while experiencing sleepiness, and 15% reporting such instances occurring more than three times per week. Moreover, 15.4% acknowledged falling asleep at the wheel within the preceding five years. The gravity of the situation extended to near-miss accidents, with 42.5% confirming their proximity, while 16.3% admitted to having had accidents attributable to drowsiness. These revelations emphasize the critical need to address sleep-related issues among truck drivers to enhance road safety and mitigate the risks associated with drowsy driving. In 2015, the authors [67] recognized the imperative to explore the relationship between road accidents and sleepiness within the Portuguese population. The challenge arose from the lack of data collection regarding sleepiness as a factor in reported accidents, prompting the need to address this gap in understanding. To bridge this knowledge gap, a random selection of regular drivers (900

participants), driving at least once per week, was reached by phone. Results from this survey indicated that 23.1% of drivers experienced at least one episode of drowsiness while driving, with 3.1% acknowledging falling asleep at the wheel at least once. Disturbingly, only 2.1% narrowly avoided accidents due to drowsiness, and 0.67% actually had an accident. Simultaneously, the European Sleep Research Society conducted a comprehensive study across 19 countries in the same year [68]. Utilizing an anonymous online questionnaire, involving 12 434 participants aged 17 and above, the study aimed to evaluate the prevalence and consequences of falling asleep at the wheel. Portugal ranked fourth among European countries with the highest frequency of drowsiness behind the wheel. Among the 1 093 Portuguese participants, 22.3% admitted to falling asleep while driving, and 1.7% had experienced accidents due to drowsy driving. An important observation highlighted by the study was that many drivers persist in their activities despite being aware of their drowsiness, and one out of every six accidents resulting from this condition led to serious injury or death.

The existing literature underscores a notable gap in research concerning sleep disorders within the Portuguese population, particularly with representative samples encompassing all districts. Consequently, there exists a critical imperative to conceptualize and execute an exhaustive study, conducting thorough statistical analysis to comprehensively assess sleep quality and excessive daytime sleepiness among Portuguese drivers. Recognizing the intricate interplay between circadian rhythm and work schedules becomes imperative for comprehending and mitigating sleep-related challenges. This acknowledgment underscores the necessity for interventions tailored to individual circadian preferences to optimize both well-being and productivity. This concerted effort is indispensable for acquiring nuanced insights into the prevalence and intricate patterns of sleep-related disorders, laying the foundation for tailored therapeutic interventions. Concurrently, addressing the issue of drowsiness at the wheel demands the identification of reliable, cost-effective solutions that seamlessly integrate with the driver's activity. Current methodologies involve the utilization of movement sensors within the wheel and cameras strategically positioned in the rear-view mirror. These systems adeptly recognize a driver's drowsy state and dispatch timely alerts. However, a significant drawback lies in the often prohibitive cost of these methods, rendering them inaccessible to a broader demographic. This economic barrier creates a division, where only individuals with greater financial means can benefit from such technology [122].

The investigation of drowsiness at the wheel has garnered attention in recent years, yet there remain aspects that require refinement and exploration. In light of this, the study aims to address the following research questions:

Q1: What are the sleep quality and prevalence of excessive daytime sleepiness among Portuguese drivers?

Q2: What is the influence of the circadian rhythm on drowsiness at the wheel?

Q3: What is the best classifier to predict drowsiness at the wheel?

## 1.2 Research Objectives

The central aim of this PhD project is to develop intelligent models capable of preemptively detecting drowsiness or insufficient wakefulness by analyzing biometric data from a wearable device. While identifying drowsiness is crucial, the primary focus is on prevention, aiming to anticipate in advance, while the driver is still at full capacity, the imminent decline in alertness. This proactive approach aims to reduce accidents and prevent fatalities. To achieve this, a non-intrusive and cost-effective wearable device will be integrated into the driver's activities. By detecting and predicting drowsiness while driving, it aims to alert the driver in advance of any signs of drowsiness manifesting within a short period of time.

Furthermore, acknowledging a literature gap concerning sleep quality and excessive daytime sleepiness among the Portuguese driving population, the project plans to conduct a comprehensive survey across different districts to ensure representation. This extensive survey will not only provide valuable insights into the sleep patterns and challenges specific to drivers but will also contribute to a broader understanding of sleep disorders within the larger Portuguese population. The project also recognizes the crucial importance of taking into account the driver's circadian rhythm. When there is a misalignment between the circadian rhythm and work schedules, it can adversely affect productivity and increase the risk of excessive daytime sleepiness. To address this, the project endeavors to identify distinct circadian rhythm clusters among passenger drivers. This identification process aims to facilitate the aligning of schedules with individual circadian preferences, promoting optimal productivity, and mitigating the risk of daytime sleepiness for enhanced overall well-being.

To fulfill these overarching goals, specific objectives have been outlined:

- Evaluate the sleep quality and excessive daytime sleepiness of Portuguese drivers;
- Evaluate the impact of circadian rhythm on the driver;
- Identify the physiological signals collected by the wearable device capable of detecting and predicting drowsiness.

The ultimate goal of the project is to develop a non-intrusive, cost-effective solution capable of effectively detecting and predicting drowsiness at the wheel. The implementation of intelligent models is expected to lead to a reduction in drowsiness-related incidents, contributing to fewer road accidents, serious injuries, and fatalities. Moreover, by gaining insights into sleep disorders, sleep quality, excessive daytime sleepiness, and circadian rhythm among the Portuguese driving population, the project aspires to enhance passenger safety, improve drivers' overall quality of life, and elevate their job performance.

## 1.3 Research Methods and Contributions

Research methods "*are the how for building systematic knowledge*" [126], intended to be systematic and bias-free. They aim to generate findings that are as similar to reality as possible, avoiding hidden biases

that can influence the conclusion under study. Therefore, the present thesis follows a quantitative research approach since it aims to extract as much information as possible about the population from a selected sample, considering statistical methods and mathematical models.

This research process is associated with a deductive approach as it involves the application of existing theories of drowsiness at the wheel and improvements were proposed to achieve better results. For the application of the deductive approach, three essential steps are taken into consideration, starting with the **hypothesis formulation** based on the theory and explaining how the concepts will be developed. Our research questions were defined based on the problems identified in the literature, where drowsiness at the wheel continues to be a problem, and the identification of the best classifier to predict drowsiness will be achieved by driving simulations. Meanwhile, the characterization of the sleep quality and excessive daytime sleepiness of the Portuguese drivers, and the influence of the circadian rhythm on drowsiness will be answered through questionnaires. **Hypothesis testing** is the next step of the analysis, which consists of gathering data, developing and testing models that will explain the observed events and promote the probable solution to the highlighted problem. In our research, the main goal is the prediction of drowsiness, and the data will be extracted by a low-cost wearable device, and machine learning algorithms will be applied. Regarding the remaining study questions, statistical analysis will be undertaken to extract as much information as possible. Then, **hypothesis validation** is the last step, where the results must be analyzed to either confirm the theory or indicate the need to modify it [70, 126].

In terms of contributions to the research community, this PhD project offers several noteworthy aspects. Firstly, it includes the development of a comprehensive study encompassing Portuguese drivers from all districts, integrating diverse questionnaires to assess sleep quality, excessive daytime sleepiness, other sleep disorders, and personal information. The construction of this study is meticulously detailed, providing transparency on the questionnaires used, inquiries into personal information, sample size determination, the disclosure process and a detailed statistical analysis. Additionally, the evaluation of work schedules and circadian rhythms constitutes a substantial contribution, providing a nuanced exploration of the challenges encountered by Portuguese drivers and addressing critical gaps in the existing literature. This evaluation not only seeks to understand the prevailing work schedules but also aims to identify if these schedules are aligned with the natural circadian rhythms of individuals. By exploring this aspect, the project aims to understand if the work schedules in Portugal support both productivity and well-being. This exploration provides insights into possible areas where workplace practices could be improved. Another noteworthy contribution is the identification of a practical and cost-effective mathematical model capable of predicting the onset of drowsiness within a short time, using a wearable device. This model not only provides a valuable solution for early detection but also plays an important role in recognizing crucial features that distinguish drowsy and awake states. By identifying these distinctive features, the project adds significant value to the field, offering insights that can inform the development of effective interventions and preventive measures to address drowsiness among drivers.

Throughout the course of the PhD project, a key contribution to the literature lies in the development of six detailed articles, each thoroughly explaining the attained results. Furthermore, the dissemination of

these results at two conferences, the International Conference on Computational Science and Its Applications (ICCSA) and the International Symposium on Digital Forensics and Security (ISDFS), has fostered discussions and knowledge sharing within the academic community. Overall, these contributions collectively contribute to advancing the understanding of sleep-related issues among Portuguese drivers and provide practical solutions for real-world application. The articles developed are as follows:

> Antunes AR, Braga AC, Gonçalves J. Drowsiness detection using multivariate statistical process control. In Computational Science and Its Applications–ICCSA 2022 Workshops: Malaga, Spain, July 4–7, 2022, Proceedings, Part I 2022 Jul 23 (pp. 571-585). Cham: Springer International Publishing.
>
> Antunes AR, Meneses MV, Gonçalves J, Braga AC. An Intelligent System to Detect Drowsiness at the Wheel. In 2022 10th International Symposium on Digital Forensics and Security (ISDFS) 2022 Jun 6 (pp. 1-6). IEEE.
>
> Antunes AR, Braga AC, Gonçalves J. Drowsiness Transitions Detection Using a Wearable Device. Applied Sciences. 2023 Feb 18;13(4):2651.
>
> Antunes AR, Silva JP, Braga AC, Gonçalves J. Feature Selection Optimization for Heart Rate Variability. In2023 11th International Symposium on Digital Forensics and Security (ISDFS) 2023 May 11 (pp. 1-6). IEEE.
>
> Meneses MV, Antunes AR, Gonçalves J. Driver Drowsiness Classification using Machine Learning and Heart Rate Variability. In2023 11th International Symposium on Digital Forensics and Security (ISDFS) 2023 May 11 (pp. 1-6). IEEE.
>
> Antunes AR, Braga AC, Gonçalves M, Gonçalves J. Sleep Disorders in Portugal Based on Questionnaires. In International Conference on Computational Science and Its Applications 2023 Jul 1 (pp. 97-113). Cham: Springer Nature Switzerland.

## 1.4 Thesis Structure

This thesis is divided into seven chapters, organized as follows: Chapter 2 offers an in-depth exploration of drowsiness, covering essential concepts such as sleep stages and various methods for drowsiness detection. It includes an overview of available questionnaires designed to identify sleep disorders and a review of related works in the literature.

Chapter 3 begins with a comprehensive exploration of fundamental concepts in statistics and machine learning, laying the foundation for a deeper understanding of subsequent methodologies. Following this, the chapter examines the theoretical of various machine learning models, providing a discussion on the advantages and disadvantages associated with each one. This critical analysis equips readers with

valuable insights to better understand the model construction. Finally, the chapter concludes with an overview of multivariate statistical process control.

In Chapter 4, the focus shifts to an examination of sleep disorders among Portuguese drivers. The chapter details the procedural steps, encompassing the explanation of questions and questionnaires, the determination of the sample, the disclosure process, and subsequent pre-processing steps. A comprehensive statistical analysis follows, providing insights into sleep quality, excessive daytime sleepiness, and circadian rhythm. Our study brings new contributions to the literature since it integrates four questionnaires for the evaluation of sleep disorders, and the procedure is clearly described. Even though it covers all the districts of Portugal, the large number of responses achieved has filled the gap in the literature.

Chapter 5 delineates the procedures employed, elucidating the driving simulation and the relevant questionnaires. The chapter also articulates the analysis procedures used for classifying and predicting drowsiness at the wheel. It concludes with a participant description, summarizing personal information, driving experience, and observed drowsiness stages during the simulation. This chapter serves as a guideline for replicating and understanding the steps taken for drowsiness detection and prediction.

Chapter 6 addresses the modeling of drowsy driving, explaining the development of drowsiness detection through a multivariate approach. This innovative methodology sets itself apart from existing methods by incorporating additional components to mitigate subjectivity in drowsiness classification. The chapter further details the identification of physiological signals capable of detecting drowsiness through feature selection optimization with machine learning algorithms. Evaluation metrics are thoroughly analyzed to achieve optimal results with a minimal number of features. Following this, forecasting models were implemented to predict physiological signals, aiming to verify if it is possible to anticipate drowsiness in advance.

Lastly, Chapter 7 provides a comprehensive review and discussion of the main findings of this research, along with recommendations for future studies.

# 2

# State Of Art

*"Be less curious about people and more curious about ideas." By Marie Curie.*

This chapter presents an overview of the theoretical background of sleep for a better understanding of the thesis subject. Firstly (Section 2.1), it introduces the sleep stages and the existing methods for detecting drowsiness. Followed by the available questionnaires to evaluate sleep disorders (Section 2.2) such as daytime sleepiness, obstructive sleep apnea, sleep quality, and circadian rhythm. Section 2.3 presents an overview of similar works for drowsiness detection and a summary is presented (Section 2.4) about all the information collected.

## 2.1 Drowsiness Background

Sleep is essential for the performance of different activities, and due to brain activity, it is divided into two phases: non-rapid eye movement (Non-Rapid Eye Movement (NREM)) and rapid eye movement (Rapid Eye Movement (REM)). NREM phase is divided into four stages: $S_1$, $S_2$, $S_3$, and $S_4$. The transition between awake and sleep entry is defined as the $S_1$ stage. However, some researchers [19, 128] define it as the drowsiness state since it is not necessarily considered as a sleep onset [74]. Light sleep is denominated as $S_2$ since the body starts to relax, and the heart rate and respiration begin to decrease. However, $S_3$ and $S_4$ stages are considered the deepest sleep due to decreased brain activity. It is very difficult to awaken the individual in these two stages. Besides that, in the REM stage, the brain activity is similar to the awake stage and it is where the dreams occur [167, 28, 141]. Therefore, sleep is composed of ninety-minute cycles, with the NREM phase being the first to occur, followed by the REM phase, in each cycle.

The hypnogram is commonly used to evaluate sleep defragmentation and is one of the pieces of information extracted in the polysomnography exam. This exam considers different sensors, located in different parts of the body, to diagnose respiratory and sleep perturbations by extracting different signals.

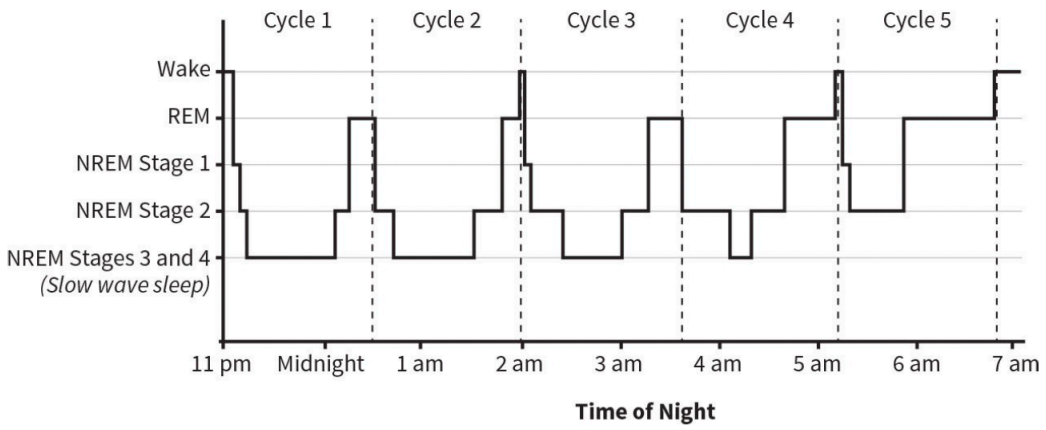



Figure 1: Hypnogram: Sleep Architecture [167].

It collects brain activity, eye movement, and muscle activity [167, 145]. Figure 1 illustrates an individual hypnogram, also known as the sleep architecture, with five cycles (in this case). On the vertical axis are the different sleep stages, and on the horizontal axis there is the time of the night. Moreover, it is possible to identify that the transition between the NREM to REM phases changes from cycle to cycle. In the first cycle, there is a passage in all the sleep stages and this changes over time and also its duration.

After understanding the sleep phases and stages, it is important to explore how drowsiness can be detected. Thus, the detection methods commonly used can be divided into four techniques: behavioral, vehicle-based, physiological, and subjective measures [132]. Behavioral approaches analyze the drivers' behavior to determine how sleepy they are. The most frequent features, including eye closure ratio, eye blinking, head position, facial expressions, and yawning, must be extracted using cameras and computer monitoring [57, 173]. This measure is not intrusive to the driver's activity, however lightning, brightness, and road conditions can be a serious problem [20]. Some cameras have a better performance during the day and others during the night [141]. Besides that, individuals with glasses can also be an issue in this type of system, since it can be difficult to identify correctly the position of the eyes [144].

Vehicle-based technique, as the name implies, uses the information present in the vehicle's intelligent systems. Additionally, vehicle tactics are employed to monitor driving behaviors and find drivers who are becoming drowsy or fatigued. The features that are frequently used include frequent lane changes, speed, steering wheel angle, and grip force data gathered from sensors in the steering wheel, accelerator, or brake pedal [79, 100]. Due to its vulnerability to outside influences, such as the marking of the lines, on the road, the weather, and the lighting, this technique can be helpful but is only partially effective in identifying tiredness. On the other hand, drowsiness detection only occurs when the driver already presents signs of drowsiness. This means these measures may not help prevent traffic accidents due to late detection [141].

Physiological techniques are the most reliable to detect drowsiness, since it is possible to alert the driver in advance. The most known and used signals are the electroencephalogram (Electroencephalogram (EEG)), electrocardiogram (Electrocardiogram (ECG)), electrooculogram (Electrooculogram (EOG)),

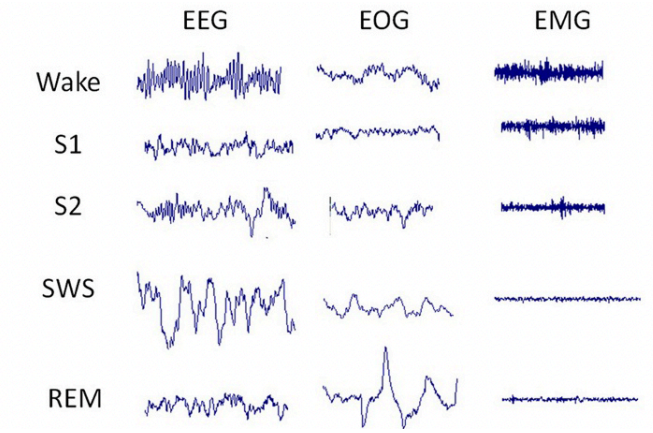



Figure 2: EEG, EOG, and EMG Signals for Each Sleep Stage [124].

electromyogram (Electromyogram (EMG)), and photoplethysmography (Photoplethysmography (PPG)). These signals take into consideration the brain activity, heart rate, eye activity, muscle activity, and changes in blood volume [86, 93, 9]. EEG, EOG, and EMG signals are the most commonly used to determine the respective sleep phases [139]. Figure 2 shows the waves for these signals in different sleep stages. Slow-Wave Sleep (SWS) represents deep sleep, and is associated with the $S_3$ and $S_4$ stages. Using this visualization, it is clear that the waves are decreasing during the course of sleep.

Heart rate is one of the measures applied to detect drowsiness, since there is a decrease in the heartbeat when the individual goes from awake to a drowsy state [98]. It is known that there are changes in the heart rate due to physical or mental stress, exercise, respiration, and metabolic changes [88]. Another metric well-known is heart rate variability, since it evaluates the heartbeat variation. Thus, the variability of the heartbeat contains information about the autonomic nervous system to convey the alertness of the individual. Importantly, it is worth noting that the activity of the autonomic nervous system undergoes alterations when an individual presents stress, fatigue, and drowsiness [110]. Sympathetic activity increases with the alertness of the individual leading to a decrease in parasympathetic activity. Contrarily, an increase in parasympathetic activity and/or a decrease in sympathetic activity describe extreme relaxation situations [110, 65]. Heart rate variability is the difference between the peaks of the R-wave, also known as the R-R interval, measured in milliseconds. This means it is the time difference between successive heartbeats [21]. One example of the R-R interval can be seen in Figure 3. This metric is also powerful for the prognostic indicators after myocardial infarction, chronic coronary disease, congestive heart failure, alcoholism, and diabetic neuropathy [88].

Therefore, heart rate variability contains information about different domains like time, frequency, and non-linear [63]. In the time domain, the aim is to quantify the variability in the interbeat interval measurements over a defined period that may range from two minutes to twenty-four hours. Different descriptive statistics can be extracted, like the mean, median of heart rate variability, and even the minimum, and maximum heart rate ( Table 1).

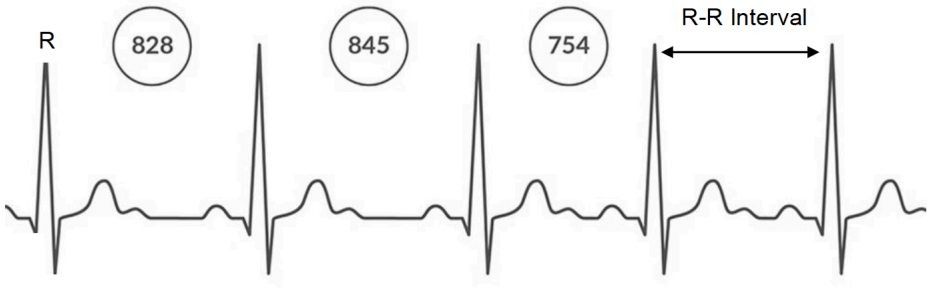



Figure 3: R-R interval for the Heart Rate [21].

Table 1: Heart Rate Variability time-domain measures.

| **Metrics** | **Description** |
|---|---|
| mean_nni | Mean of the R-R intervals. |
| median_nni | Median of the R-R intervals. |
| range_nni | Difference between the maximum and the minimum of the R-R intervals. |
| sdnn | Standard Deviation of the R-R intervals. |
| sdsd | Standard deviation of differences between adjacent R-R intervals. |
| rmssd | Square root of the mean of the sum of the squares of differences between adjacent R-R intervals. |
| nni_50 | Number of pairs of successive NN intervals that differ by more than 50 milliseconds. |
| pnni_50 | Proportion (percentage) of NN50 divided by the total number of NN intervals. |
| nni_20 | Number of pairs of successive NN intervals that differ by more than 20 milliseconds. |
| pnni_20 | Proportion (percentage) of NN20 divided by the total number of NN intervals. |
| cvsd | rmssd divided by mean_nni. |
| cvnni | sdnn divided by mean_nni. |
| mean_hr | Heart rate mean. |
| min_hr | Heart rate minimum. |
| max_hr | Heart rate maximum. |
| std_hr | Standard deviation of the heart rate. |

On the other side, the frequency domain takes into consideration four bands to estimate the distribution of absolute or relative power, known as ultra-low-frequency (Ultra-Low-Frequency (ULF)), very-low-frequency (Very-Low-Frequency (VLF)), low-frequency (LF), and high-frequency (High-Frequency (HF)) [150]. Using these frequencies, it is possible to estimate the autonomic nervous system, where the sympathetic and parasympathetic activity influences the Low-Frequency (LF) band. Note that if the LF band is normalized, the sympathetic activity is dominant. By contrast, power in the HF band is thought to be parasympathetic. The LF/HF ratio measures the balance of the sympathetic and parasympathetic nervous systems [166]. According to [154] this balance tends to decline during the onset of sleep. Moreover, the total power density spectral can also be estimated considering the sum of VLF, LF, and HF bands [150]. Table 2 presents the heart rate variability frequency-domain measures explained above.

Moreover, the non-linear domain is applied to quantify the impressibility of the time series and the

Table 2: Heart Rate Variability Frequency-domain Measures.

| **Metrics** | **Description** |
|---|---|
| power_vlf | Variance in heart rate variability in the very low frequency. |
| power_lf | Variance in heart rate variability in the low frequency |
| power_hf | Variance in heart rate variability in the high frequency |
| total_power | Total power density spectral. |
| lf_hf_ratio | lf/hf ratio. |

relationship between the features cannot be depicted as a straight line. For this domain, the Poincaré plot is widely used, where a scatter plot is created by comparing each R-R interval to the previous interval. The standard deviation is computed for each R-R interval and the previous interval, denominated as $SD1$ and $SD2$, respectively. Basically, the $SD1$ corresponds to the width of the ellipse and measures the short-term heart rate variability, in milliseconds. This metric is widely correlated with HF power, which represents parasympathetic activity. Conversely, $SD2$ represents the ellipse's length, which measures both short and long-term heart rate variability, also in milliseconds. This measure is strongly related to LF power, representing sympathetic activity [150]. From the R-R interval fluctuation, two components representing the length of the transverse axis, T, and the longitudinal axis, L, were extracted. Equation 2.1 and 2.2 present the formulation of these two components based on $SD1$ and $SD2$.

$$T = 4 \times SD1 \tag{2.1}$$

$$L = 4 \times SD2 \tag{2.2}$$

In order to estimate the sympathetic and parasympathetic activities the cardiac sympathetic index (cardiac ssympathetic index (csi)) was formulated as Equation 2.3, and the cardiac vagal index (cardiac vagal index (cvi)) as Equation 2.4 [163]. Later, a modification for the $csi$ metric was suggested on the research [82] to emphasize the $L$ component, and the formulation can be expressed by Equation 2.5.

$$\text{csi} = \frac{L}{T} \tag{2.3}$$

$$\text{cvi} = \log_{10}(L \times T) \tag{2.4}$$

$$\text{modified\_csi} = \frac{L^2}{T} \tag{2.5}$$

Another commonly used non-linear metric is sample entropy, which measures signal regularity and complexity in a less biased and more trustworthy manner [150]. Table 3 summarizes the metrics for the non-linear domain.

Table 3: Heart Rate Variability Non-Linear Domain Measures.

| Metrics | Description |
|---|---|
| csi | Cardiac Sympathetic Index. |
| cvi | Cardiac Vagal Index. |
| modified_csi | Modified csi. |
| sampen | Sample entropy. |

Subjective measures to assess drowsiness are based on the driver's feedback or an individual's opinion. Different tools were developed to translate the drowsiness level into a rating scale to a measure of driver drowsiness [141, 111]. Karolinska Sleepiness Scale (Karolinska Sleepiness Scale (KSS)) [5] is the scale commonly used to classify the level of drowsiness and is divided into nine levels, from extremely alert (defined as level 1) to very sleepy, great effort keeping awake, fighting sleep (defined as level 9), as shown in Table 4. Sometimes, it is the person who classifies their drowsiness phase, and in other cases, it is an external person who classifies it according to body behavior.

Table 4: KSS levels of drowsiness and description.

| Level | Description |
|---|---|
| 1 | Extremely alert |
| 2 | Very alert |
| 3 | Alert |
| 4 | Rather alert |
| 5 | Neither alert nor sleepy |
| 6 | Some signs of sleepiness |
| 7 | Sleepy, but no effort to keep awake |
| 8 | Sleepy, some effort to keep awake |
| 9 | Very sleepy, great effort keeping awake, fighting sleep |

There is also Wierwille and Ellsworth's drowsiness scale [159] to measure the drowsiness level, which is composed of five levels, from not drowsy (level 1) to extremely drowsy (level 5). This scale has the elements to classify the drowsiness level, such as eye movement, number of blinks, and time between blinks, mouth, head, and body movements to make it easier to classify. Table 5 presents the levels and the respective characterization to facilitate the classification.

## 2.2 Sleep Disorders Questionnaires

Sleep disorder is a clinical illness that appears as insomnia due to medical, psychological, environmental, and occupational factors [147]. It can affect the well-being of individuals and also their health, however the majority of them can be effectively treated with appropriate therapies [127]. Firstly, it is essential

Table 5: Wierwille and Ellsworth's drowsiness scale [159].

| Level | Type | Characterization |
|---|---|---|
| 1 | Not drowsy | Eye movement is rapid, and the time between blinks remains stable. |
| 2 | Slightly Drowsy | Eye movement is slow. |
| 3 | Moderately Drowsy | Blinks are slow, the mouth moves, or the driver touches his/her face. |
| 4 | Significantly Drowsy | Number of blinks increases noticeably, motions unnecessary for driving, yawns are frequent, deep breathing is detected. |
| 5 | Extremely Drowsy | Eyelids are almost closed or the driver's head inclines to the front or rear. |

to understand that excessive daytime sleepiness, obstructive sleep apnea, sleep quality, and circadian rhythm due to shift work are the most known sleep disorders [44].

Daytime sleepiness can have a negative impact on one's health, which has major repercussions such as an increased chance of car accidents and workplace mishaps. This implies that the person's cognitive abilities are impaired, which may hinder their capacity to concentrate and their productivity. Sleep loss, medication side effects, illicit substance use, and obstructive sleep apnea are the main causes of daytime sleepiness [123]. As a result, the Epworth Sleepiness Scale (Epworth Sleepiness Scale (ESS)) [84] is commonly used to assess daytime sleepiness. The questionnaire includes eight questions based on different settings in which the participant must rate how likely they are to fall asleep based on their previous routine on a scale of 0 to 3. As a result, the final score can range between 0 to 24. If the ultimate score is greater than 10, the person is excessively sleepy during the day. The Portuguese version can be found at [50]. The Sleep-Wake Activity Inventory (Sleep-Wake Activity Inventory (SWAI)) [138] also assesses excessive daytime sleepiness by taking six factors into account: excessive daytime sleepiness, distress, social desirability, energy level, ability to relax, and nocturnal sleep. This questionnaire consists of fifty-nine questions answered on a Likert scale from 1 to 9, with 1 indicating that the behavior is always present and 9 indicating that it never occurs. Respondents must answer questions about their behaviors exclusively from the last seven days, with lower scores indicating more daytime sleepiness. This questionnaire has no validation in the Portuguese language.

Obstructive sleep apnea is also considered a sleep disorder characterized by bouts of partial or total upper airway blockage while sleeping [26]. This disorder has several negative consequences, including excessive daytime sleepiness (the predominant symptom), decreased alertness, and, as a result, trouble concentrating and mood disorders [3]. As a result, the STOP-Bang questionnaire [45] is used to assess the risk of developing obstructive sleep apnea. This one takes into account snoring, weariness, observed apnea, high blood pressure, body mass index, age, neck circumference, and male gender. As an outcome, there are eight dichotomous questions, and the total score ranges from 0 to 8. Those with a score of 5 or above are at a high risk of having moderate to severe obstructive sleep apnea. The STOP-Bang validation version in Portuguese can be accessed in [135]. Another option for assessing the risk of developing sleep apnea is the Berlin Questionnaire [118]. It also takes into account snoring, daytime sleepiness or weariness, and the presence of obesity or hypertension. The participant's age, weight, height, gender,

neck circumference, and ethnicity are also gathered. There are three elements to this questionnaire: snoring and apnea (five questions), daytime sleepiness (four questions), and hypertension or obesity (one question). The responses in each category are used to determine whether a person is at high or low risk of developing sleep apnea. If the participant has a positive score in two or more categories, the participant is at high risk, otherwise the individual is not. The Berlin validation version in Portuguese is available at [165]. With thirty-two questions, the Quebec Sleep Questionnaire [94] also utilized for the identification of obstructive sleep apnea. Hypersomnolence, diurnal symptoms, nocturnal symptoms, emotions, and social interactions are the five domains of this questionnaire. Each question is rated on a 1 to 7 Likert scale, and the final result is obtained by computing the mean score (1 to 7) within each area and summing the results. Smaller scores indicate a greater impact on the participant's life as a result of sleep apnea. This questionnaire's Portuguese validation is available in [146].

In the medical field, sleep quality can also be a sleep disorder and is often used to analyze several sleep metrics such as total sleep time, sleep onset delay, fragmentation, total wake time, sleep efficiency, and sleep disruptive events [90]. Poor sleep quality may influence an individual's mood, attention, memory, and tiredness, according to the study [52]. The Mini Sleep Quality (Mini Sleep Quality (MSQ)) [180] assesses both insomnia and excessive daytime drowsiness with ten questions. It employs a seven-point Likert scale ranging from never (1) to always (7). Finally, the subjects can be classified as having good sleep quality (total sum of scores between 10 and 24), mild sleep difficulties (25 to 27), moderate sleep qualities (28 to 30), or severe sleep difficulties (above 31). The MSQ questionnaire is only available in Brazilian Portuguese [56]. To assess the prior month's sleep quality, the Pittsburgh Sleep Quality Index (PSQI) [31] is widely employed. This one includes nineteen questions grouped into seven categories: subjective sleep quality, sleep latency, sleep duration, habitual sleep efficiency, sleep disruptions, use of sleeping drugs, and daytime dysfunction. Each component has a score ranging from 0 to 3, and the overall score (0 to 21) is the sum of the component scores. Subjects with a final score of more than 5 had poor sleep quality. The Portuguese version of this questionnaire can be assessed in [83]. Sleep Quality Scale (Sleep Quality Scale (SQS)) [172] is another questionnaire used to assess sleep quality during the past month. It consists of twenty-eight elements separated into six categories: daytime dysfunction, restoration after sleep, difficulty falling asleep, difficulty getting up, contentment with sleep, and problems maintaining sleep. A four-point Likert scale was utilized, with values ranging from 0 (few) to 3 (almost always), with higher scores suggesting poor sleep quality. There is no Portuguese validation version of this questionnaire.

Lastly, the circadian rhythm also plays a significant role in performance and sleep patterns. Horne and Ostberg [75] created the Morningness-Eveningness (Morningness-Eveningness (ME)) questionnaire, which consists of nineteen questions with a final score ranging from 16 to 86. Subjects with classifications ranging from 16 to 30 are classified as definitely evening, 31 to 41 as relatively evening, 42 to 58 as neither type, 59 to 69 as moderately morning, and 70 to 86 as definitely morning. As the prior questionnaire was so long, five questions were extracted to determine the circadian rhythm. This study aims to save researchers and participants time and effort, according to [37]. A Portuguese adaptation version

was made in 2002 [155], with the number of questions lowered to 16 and the time scales adjusted. In 2016 [133] an updated scale to quantify circadian rhythm was designed to address the shortcomings of the Horne and Ostberg questionnaire. It is known as Morningness-Eveningness-Stability-Scale improved (Morningness-Eveningness-Stability-Scale improved (MESSi)) and it was planned to improve the four and five Likert scales, remove self-assessment questions, add at least two dimensions, one for morningness and one for eveningness, and minimize the number of questions. As a result, the questionnaire created contains fifteen questions on a Likert range of 1 to 5 points. Furthermore, it enables the separate identification of three factors, such as distinctness, and morning and evening influence. It should be noted that the component distinctness is utilized to determine whether or not the subject experiences significant variations in performance and mood throughout the day. Each factor contains five questions, with higher values representing more changes throughout the day, morning, and evening preferences. For this questionnaire, the Portuguese version can be found at [137].

## 2.3 Similar Works

Over the years, researchers have looked into ways to detect drowsiness while driving. One of the earliest studies, dating back to 1972, explored how changes in heart rate can indicate when a driver is feeling tired or sleepy. This discovery paved the way for further investigations into heart rate variability and its connection to drowsiness [120]. Therefore, this section will be about similar studies found in the literature, focusing on drowsiness detection or prediction methods that use affordable, non-invasive devices. Studies validating the substitution of medical devices with wearable ones in comparable driving contexts can also be incorporated. It is important to highlight that the foundational concepts of machine learning modeling are introduced, for the studies that have that information, creating the way for a more in-depth exploration in the following chapter.

In 2015, a study [96] was conducted to develop a driver alertness monitoring system. Information was gathered through a smart watch (built from scratch), where the PPG sensor was located on the driver's finger. For the classification of the alertness of the driver, only twelve candidates be part of a driven simulation, with each participant undergoing an adaptation period of approximately two hours. Data was collected over two days with different schedules, each lasting eight hours. Participants' faces were recorded to assess their drowsiness levels using the KSS questionnaire. The main aim of this study was to determine the accuracy of the support vector machine model in discriminating awake and drowsy states. To simplify the process, feature selection was carried out using mutual information to identify the top eight features. A notable observation from this study was the association of power_lf, power_hf, and lf_hf_ratio with drowsiness [96]. One disadvantage of this approach is that the developed device is not practical for driver activity since the construction of the device is not finalized. However, despite identifying the relevant features and assessing the accuracy of the model, it lacks clarification on how the awake, and drowsy states were precisely defined, the number of records for each state, and a statistical comparison

between them. The drowsiness classification may be subjective, as it relied on two participants to classify the drowsiness state based on recorded videos.

In 2016, a follow-up investigation on [78] aimed to assess the PPG sensor's effectiveness in detecting heart rate among civil construction workers. The study involved eleven male workers equipped with a wearable device (on their non-dominant hand), an ECG sensor (chest belt), and a hand-held video camera to identify the causes of the data noise. Data was collected for one hour, except when it interfered with the participant's activities. Due to connectivity issues and memory constraints, the study could only be completed with seven workers. Nevertheless, the findings indicated that extracting heart rate values with the PPG sensor was reliable, even though improvements must be made [78]. Notably, this study did not consider the workers' physical state but served as a validation for using wearable devices in the driving context. The results indicate reliability for civil construction workers, suggesting promising applications for passenger drivers, given the lack of heavy work and reduced environmental noise.

A different approach [1] aimed to detect drowsiness through heart rate variability using an anomaly detection method. Data collection involved a driving simulation conducted on a simulated highway loop course, specifically during nighttime, for two hours. The simulation environment excluded other vehicles, maintaining a constant speed limit of 80 kilometers per hour. Drowsiness was defined by participant involvement in accidents during the simulation, with twenty-seven individuals taking part. An additional criterion for estimating drowsiness was considered, involving an expression-based scale divided into four levels: from the subject appearing not sleepy at all to almost falling asleep. Additionally, heart rate variability was monitored using telemetry devices, yielding eight extracted features. These features included mean_nni, sdnn, variance, rmssd, total power for the time domain, and power_lf, power_hf, and lf_hf_ratio for the frequency domain. Drowsiness classification utilized multivariate statistical process control, employing principal component analysis (Principal Component Analysis (PCA)). A single principal component integrating both time and frequency features was considered, with out-of-control points detected using Hotelling's $T^2$ and the $Q$ statistic. By applying this method, drowsiness detection was achieved with a notable success rate, accurately identifying seven out of eight events [1]. Despite yielding promising results, this method suffers from several limitations. Firstly, there is a lack of information regarding the sample size and the specific device utilized for data collection. The drowsiness classification is also subject to subjectivity, and there is inadequate information available regarding the classification process. Additionally, the absence of statistical analysis comparing feature values between awake and drowsy states is notable. Moreover, the absence of feature selection in the methodology poses a substantial limitation, potentially resulting in a time-consuming process and the loss of valuable insights.

In 2017 [43], a new approach was developed to identify the awake, stressed, fatigued, and drowsy states through physiological information. A device was developed to gather that information, also from scratch. Four experiments were developed, for twenty-eight participants, during four hours in total and considering a driving simulation. During the first thirty minutes, the participant drove in a normal city, and the information extracted was relative to the awake state. The simulated game was redefined for the stress state to have intensive traffic, with more probability of accidents and noise sound, over thirty minutes. In

the drowsiness state, the participants had to drive for over two hours, on a highway, without the presence of other cars, with monotonous music to increase the drowsiness. Using a webcam, data was manually classified as not drowsy, light, moderate, a lot, and extremely drowsy. Through this classification, the drowsiness state is the last three categories mentioned before. Lastly, the fatigue state was collected over the last thirty minutes, in a normal city road. According to the results achieved and due to the difficulty of distinguishing the drowsiness and fatigue states, these two states had to be grouped. The Support Vector Machine (Support Vector Machine (SVM)) algorithm was used to classify stress, fatigue, drowsiness, and normal states. In terms of information, thirty-eight features extracted from heart rate variability, PPG amplitude, and galvanic skin response were considered in the process. As an input for the literature, this study defines a system to detect abnormal conditions from the driver's behavior [43]. Moreover, feature selection was implemented even though there was not a comparison between the different states. The study is very detailed, nevertheless it was developed in a controlled environment, and does not consider information from new participants to classify their state. Furthermore, the classification model was iteratively developed, considering all possible combinations of stress, fatigue, drowsiness, and the normal state. This process began by modeling the normal state against all other categories collectively, followed by a similar approach for stress compared to the remaining categories. This iterative procedure was then applied to the remaining categories. Additionally, the modeling was replicated within each class, pairing the normal state with stress, the normal state with fatigue, and so forth. Finally, the last two models aggregated fatigue and drowsiness states with the normal and stress states, respectively. This indicates the necessity of developing thirteen SVM models, which can be a time-consuming endeavor. Alternatively, different machine learning algorithms can be utilized to effectively classify more than two classes.

Two years later, [97], ECG (chest) and PPG (left wrist) sensors were implemented for drowsiness detection, considering the heart rate variability. The classifier to discriminate between the awake and drowsy states was constructed using a driving simulation with six participants. The awake state was defined as the driving simulation developed in the morning, whereas the drowsy state was defined as the simulation developed after lunch or dinner. In terms of duration, it was between one to two hours, and the heart rate variability features were extracted every minute. Moreover, two cameras were employed, one for capturing the participant's face and the other for monitoring their behavior. A convolutional neural network was built by utilizing a threshold recurrence plot derived from filtering with a modified rectified linear unit. Additionally, all features were derived from the recurrence plot applied to the R-R intervals [97]. One interesting remark about this work was that by using recurrence plots to distinguish awake and drowsy states, it was possible to conclude that the PPG sensor can be a better alternative than ECG. This study complements the validation of the use of the PPG sensor to detect drowsiness, although the information collected is quite small. Besides that, the way the awake and drowsy states were defined can be misleading since it does not consider the circadian rhythm of the participant. For example, evening types may not be fully awake in the morning. It's crucial to highlight that the features utilized lack detailed descriptions, making it challenging to comprehend the results, particularly in the comparison between drowsy and awake states.

Heart rate variability was once again utilized in [92] to validate the effectiveness of a wearable device as a single data source for detecting drowsiness. The study involved conducting driving simulations on a closed-loop monotonous track for a total duration of fifty-five minutes, with the initial ten minutes designated for adaptation. Autonomous cruise control was incorporated into the experiment. The Empatica E4 wristband wearable device was employed, while the participant's facial expressions were recorded for defining the ground truth of drowsiness through video rating and image processing. Thirty participants took part in the study, where their states were categorized as either not drowsy or drowsy, every five minutes. Although raters manually classified only the first minute, the state for the remaining duration was determined using image processing based on the eyelid closure time and the number of micro-sleep events. From each 5-minute interval, one minute was extracted to be rated by the observer, where the scale ranged from not drowsy to extremely drowsy, with six levels. Each stage included specific elements to consider for accurately identifying the drowsiness level. Moreover, the dataset comprised twenty-six features derived from time, frequency, and non-linear analyses of heart rate variability. A sliding window of 2 seconds was applied to increase the number of observations. To address class imbalance, the synthetic minority oversampling technique was applied to adjust the number of instances in the two classes before training the classifier in each iteration of 10-fold cross-validation. Feature selection was performed on each training set in every iteration after oversampling. Correlation-based feature subset selection was utilized for this purpose, aiming to identify a subset of features highly correlated with the output class while minimizing correlation among the features themselves. Nine machine-learning algorithms were utilized for the modeling process, including random forest, random tree, decision stump, decision table, K-Nearest Neighbor (K-Nearest Neighbor (kNN)), Bayesian Network, Naïve Bayes (Naïve Bayes (NB)), SVM, and multilayer perceptron. Among these, the K-Nearest Neighbor model emerged as the top performer, delivering superior results. Therefore, with the application of supervised machine learning, it was proved that is feasible to detect drowsiness using only physiological data [92]. Another interesting remark was that the study was developed using a wearable device available on the market, rather than developing a new one. This indicates the feasibility of implementing drowsiness classification with today's technology. However, there is a lack of identification regarding the best subset of features and a statistical analysis to compare both stages.

In the study [179], a fatigue prediction method was developed to anticipate the transition from non-fatigue to fatigue stages using physiological information. The common measure employed was Percentage of time Eyelids Closure (PERCLOS) (percentage of time eyelids closure), which assesses how frequently drivers close their eyes to detect drowsy or fatigue. The study utilized the VIRTTEX driving simulator, powered by a hydraulic Stewart platform, to gather data and it provides accurate feedback on road and tire forces. Besides that, participants wore ISCAN© eye tracking goggles to monitor eyelid closure, and a Bio-Harness 3.0 sensor recorded physiological signals such as ECG and breathing waves. Before the main drive, participants practiced lane changing and interacting with the automated system to be familiarized with the set up. After that, the simulation was conducted in daylight and good weather, where the participants followed a leading vehicle on a four-lane road with light traffic. The participants then drove for

about forty to forty-five minutes without distraction at a speed of fifty to seventy mph. The total duration of the experiment was approximately one hour for each participant. Thirty-five physiological features were extracted from the gathered data, and the mean squared error was employed to determine the optimal subset of features. The analysis, conducted using the nonlinear autoregressive exogenous network model, identified the top four features for prediction based on heart rate and breathing. Remarkably, this model accurately forecasted driver fatigue a minimum of 13.8 seconds in advance [179]. Although this study is a pioneer in predicting fatigue, the sample size is limited and only four features of heart rate variability and breathing data were considered. There is no identification of these features and how they affect in fatigue stages.

The systematic review developed in the research [30] focuses on analyzing specific relevant features for drowsiness detection in the PubMed, Embase and Cochrane databases. The review encompassed nineteen studies, examining participant demographics, including age, as well as heart rate measures to detect drowsiness, fatigue, and stress. For each study, a thorough analysis was conducted to identify the key findings and, in some cases, the most important features. These features include heart rate variability parameters such as $sdnn$, $rmssd$ and spectral components like low frequency to high frequency ratio (lf_hf_ratio). In examining drowsiness in drivers, higher heart rate variability was consistently associated with these states, while alert drivers showed lower heart rate variability but increased heart rate. Moreover, drowsy drivers exhibited increased high frequency power and a reduced lf_hf_ratio ratio, indicating high parasympathetic activity. However, one study deviated from this pattern and reported opposite results, with increased SDNN among drowsy drivers but also a higher lf_hf_ratio ratio [30].

For a comprehensive understanding of the studies mentioned above, Table 6 provides details including the authors, year of publication, intelligent devices utilized, advantages, disadvantages, and modelation employed in each study. Through this analysis, various studies aiming to find improved solutions to combat drowsiness were identified. However, there are some limitations that must be addressed. While these studies primarily focus on the classification of drowsiness and the validation of devices, it's crucial to define a methodology capable of predicting drowsiness in advance and understanding which features contribute to that state.

Table 6: Advantages and disadvantages of the similar works.

| Year | Author | Device | Advantage | Disadvantage | Modelation |
|---|---|---|---|---|---|
| 2015 | Lee, Lee, and Chung [96] | Lilypad Arduino, PPG | With the placement of a wearable device it was possible to identify the awake and drowsy state. The most important features to distinguish these two states were detailed. | The wearable device is not practical for the driver's activity. The distinction between awake and drowsy states is not explicit, nor is the number of records collected. The sample size is limited and the experience was developed under a controlled environment. There is no comparison between the awake and drowsy states. Subjectivity of the drowsiness classification. | Data was divided into drowsy and awake state to implement in the SVM model. Besides that, it was defined the training set as 70% and the testing set as 30%. The model achieved an accuracy of 95.8% |
| 2016 | Hwang et al. [78] | Basis Peak, Polar H7 Strap | A detailed comparative analysis of heart rate measurements using a PPG and an ECG sensors. The findings indicated that the PPG sensor produced similar results comparable to those of the ECG sensor. | The sample size is limited, there is no classification of the worker's state and the participants are all male. | The comparative analysis was performed using the mean difference, standard deviation of the differences, mean average percentage error (equal to 4.79%) and the Pearson correlation ($r = 0.85$). |
| 2016 | Abe et al. [1] | Telemetry device | The study introduces a novel approach to detecting driver drowsiness by integrating heart rate variability analysis with multivariate statistical process control, a method not previously explored in the literature. | The study did not fully account for real-world conditions such as varying road types, weather conditions, and traffic levels. There is also a lack of a statistical analysis to compare drowsiness and awake states, as also feature selection. Besides that, the device is intrusive since it involves electrodes. | The study employed multivariate statistical process control with one principal component. Sensitivity for Hotelling's $T^2$ was determined to be 68%, while for the $Q$ statistic it reached 94%. Specificity was observed at 61% for Hotelling's $T^2$ and 11% for the $Q$ statistic. |
| 2017 | Choi et al. [43] | Wearable device | Presents the most important features to distinguish the awake and drowsy state, as the classifier. It is proven that is possible to detect drowsiness through a wearable device. There is also information about stress and fatigue. | The drowsiness study was developed in a controlled environment, the sample size is limited and the drowsiness classification is subjective. There is no exploratory analysis to compare the two stages to understand what changes from one stage to another. The time-consuming nature of developing thirteen support vector machine models to cover all combinations of stress, fatigue, drowsiness, and the normal state. The drowsiness classification is subjective. | The dataset was partitioned into categories representing stress, fatigue, drowsiness, and normal states. Additionally, the training set comprised 27 participants, with the final participant reserved for testing purposes. Results showed that the model achieved an overall accuracy of 68.31% when using the four classes and 84.46% using three (join fatigue and drowsiness). |
| 2019 | Lee, Lee, and Shin [97] | Polar H7 Strap, Microsoft Band 2 | The awake and drowsy state was distinguished using recurrence plots. It was proved that the PPG sensor can be an alternative to the ECG sensor. A statistical analysis is presented to compare the features values for the awake and drowsy state. | The study was performed in a controlled environment. The number of participants was low, as was the number of records. There were only 203 records for the drowsy state and 138 for the awake state. Lack of description for each feature used. | The data was divided into awake and drowsy states, and a convolutional neural network was employed for classification. Evaluating the model's performance and incorporating insights from the PPG sensor data, four key evaluation metrics were obtained: (i) accuracy = 64%, (ii) precision = 74%, (iii) recall = 78%, and (iv) F-score = 71%. |
| 2020 | Kundinger, Sofra, and Riener [92] | Empatica E4 | Practical demonstration that a wearable device, present in the market, can be used to detect drowsiness, without developing one. The driving simulation procedure is detailed, as were the heart rate variability features. | The implementation of a sliding window could not been applied in the best way since the classification of drowsiness was defined in a different period. The best subset of features to classify the drowsiness and awake states was not identified, as well as the lack of comparison between the two states. | Drowsiness data was split into drowsy and non-drowsy categories, with a 10-fold cross-validation employed during the training phase. Furthermore, feature selection was performed using correlations. Notably, kNN demonstrated superior performance, achieving an accuracy rate of 92.13%. |
| 2020 | Zhou et al. [179] | VIRTTEX , ISCAN©, Bio-Harness 3.0 | The fatigue prediction was achieved with 13.8 seconds ahead of time. Besides that, it was highlighted that wearable devices can be utilized to extract physiological data with minimal invasiveness. | The study was developed in a controlled environment with a limited sample size. There is no reference of the features selected and there is only 1.6% of non-fatigue data. | The data was split into training and testing sets, with the latter involving a participant not used in the training phase. Feature selection was performed using the random forest algorithm, where the top four was selected using the mean squared error. Then, drowsiness prediction was employed using the nonlinear autoregressive exogenous network, that achieved an F1-score of 97.4%. |

## 2.4 Summary

The thorough literature review on drowsiness helped understand how sleep stages are categorized and the different methods used to detect drowsiness. It also explored four sleep disorders (sleep apnea, excessive daytime sleepiness, sleep quality and circadian rhythm), revealing the reasons behind drowsiness and the metrics used to measure it. Physiological techniques were found to be the most dependable for detecting drowsiness, despite worries about signal interference and the invasive nature of traditional measuring devices. To overcome these issues, recent studies have focused on creating wireless, non-invasive devices for signal measurement with reliable results. Furthermore, the review highlighted the prevalence of driving simulations and wearable devices in drowsiness research, underscoring their utility in real-world applications. Notably, heart rate variability has emerged as a robust metric for discriminating between wakeful and drowsy states.

Improvements are essential considering the limitations identified in the reviewed studies. Some studies were constrained by limited sample sizes, hindering the generalization of their findings. Others lacked statistical analyses to compare awake and drowsy states, which is crucial for understanding feature behavior. Additionally, feature selection was not systematically conducted in some studies, potentially overlooking important predictors of drowsiness. Moreover, some of them relied on subjective classifications of drowsiness, which may introduce bias and inconsistency. Furthermore, while many studies focused on classifying drowsiness, there is a growing recognition of the importance of predicting drowsiness before it occurs, which requires proactive measures rather than reactive responses. Finally, conducting studies in controlled environments limits the applicability of the results to real-world scenarios. Therefore, this thesis presents a unique opportunity to bridge this gap and contribute to the advancement of knowledge in drowsiness detection by addressing these critical limitations in the existing literature.

# 3

# Methodology

*"The combination of some data and an aching desire for an answer does not ensure that a reasonable answer can be extracted from a given body of data." By John W. Tukey.*

This chapter explores the statistical concepts and models essential for gaining a deeper understanding of data and reinforcing the foundations of this doctoral thesis. To begin, Section 3.1 will explore machine learning concepts and highlight the importance of statistics. Consequently, it provides in-depth explanations of the implemented algorithms and a comprehensive evaluation of their respective merits and drawbacks. Then, in Section 3.2, an in-depth analysis of eight machine learning models will be presented. Each model is examined thoroughly, encompassing key concepts, advantages, disadvantages, and real-world applications across various domains. Following this, Section 3.3 will elucidate the significance of anomaly detection and introduce the fundamentals of multivariate statistical process control using principal components analysis. A brief summary of the methodologies is described in Section 3.4.

## 3.1 Statistics vs Machine Learning

Nowadays, a lot of information is constantly monitored, and there is a need to possess the skills to analyze and derive insights from this data effectively. In the era of big data, machine learning, data science, and artificial intelligence are frequently explored and highly popular, while the importance of statistics has sometimes been undervalued. However both statistics and machine learning are ***"about learning from data"*** [72]. For this reason, it will be elucidated the fundamental concepts of both statistics and machine learning, while also highlighting the advantages of each one.

Statistics is a science that involves collecting, analyzing, interpreting, organizing data, and presenting valuable insights about the subject under study. Besides that, statistics aids decision-making and

drawing conclusions in the presence of uncertainty [47]. According with the study [112], ***"statistics is the science of data"***. Therefore, it is divided into descriptive and inferential statistics. Descriptive statistics is applied to organize and summarize data to make it easier for interpretations and to define future analysis [112]. Frequency tables, location measures (such as mean, median, mode, and quartiles), and dispersion measures (like amplitude, interquartile range, variance, and coefficient of variation) are well known to explore qualitative and quantitative data. In contrast, inferential statistics is employed on a sample (a specific group) to extrapolate insights about the population, which serves as a representation of the entire group under examination. Hypothesis testing, estimating numerical characteristics, and assessing correlations within data are among the most familiar inferential statistical techniques. Statistical modeling entails the use of statistical methods on data to discover underlying relationships by analyzing the significance of the variables [47]. It significantly impacts decision-making as statistical analysis provides objective insights rooted in data rather than relying on intuition or guesswork. The ability to quantify uncertainty and establish the reliability of conclusions through techniques like confidence intervals and hypothesis testing contributes to more robust and dependable decision-making. The capability to transform datasets into meaningful and manageable information enables efficient data management. Additionally, it guides risk assessment and supports scientific research by enabling solid conclusions from experiments and observations. Performance evaluation, comparative analysis, and data visualization all find their strength in statistics enabling the clear communication of findings. In summary, statistics stand as a versatile and indispensable tool, unlocking valuable insights and refining decision-making [156].

Machine learning is a field of computer science that uses past experiences to learn from data and apply acquired knowledge to make informed future decisions. It is an intersection of computer science, engineering, and statistics, where the primary aim of machine learning is to generalize discernible patterns or derive previously unknown rules from provided examples [47]. Besides that, it also employs the discovered patterns to forecast future data or to make other types of decisions under uncertainty [115]. Machine learning can be divided into supervised, unsupervised, and reinforcement learning. Supervised learning algorithms begin with labeled data for training purposes. This implies that for each provided dataset, the algorithm predicts an output or solution based on the relationships and dependencies it has learned from that data [53]. Defining a training set as $D = \{(x_i, y_i)\}_{i=1}^{N}$, where $x$ represents inputs (also known as features or attributes), $y$ represents the outputs (response variable) and $N$ is the number of training examples. When the response variable ($y_i$) consists of categorical values, such as gender categorized into female or male, this type of problem is commonly referred to as classification or pattern recognition. Conversely, when $y_i$ represents a real-valued scalar, it is typically referred to as a regression problem [47]. Moreover, unsupervised learning or descriptive algorithms initiate with a model that takes in unlabeled data and autonomously identifies hidden patterns or groupings within the data, operating without any external guidance or instructions [53]. In mathematical terms, the available information only concerns the inputs, denoted as $D = \{(x_i)\}_{i=1}^{N}$, and the objective is to uncover patterns within the data. This process is called knowledge discovery, and there are no metrics for evaluating errors, as in supervised learning, where it can be compare the predicted values $(\hat{y}_i)$ can be compared with the response variable $(y_i)$. In

contrast, reinforcement learning stands apart from traditional machine learning methods like supervised and unsupervised learning since it does not rely on labeled data or direct supervision. It is an agent that learns to make decisions through interactions with an environment to achieve the highest possible cumulative rewards. Basically, the agent learns through trial and error, interacting with its environment and adjusting its behavior based on the outcomes of its actions. This self-directed learning approach enables the agent to adapt and improve its decision-making abilities over time, making it well-suited for tasks where explicit guidance or labeled examples may be unavailable or impractical. Therefore, machine learning offers a range of advantages to enterprises and the research community. Starting with cost savings that stand out by automating repetitive tasks, reducing manual labor requirements, optimizing resource allocation through predictive analysis, and improving decision-making to prevent costly errors and inefficiencies. It also efficiently handles massive data volumes, enabling real-time analysis for swift responses to market dynamics and accelerating insights through data processing automation. Basically, the ability of the models to adapt to any type of area under study is a very strong point. Therefore, identifying anomalies, vulnerabilities, and threats in data through continuous monitoring is another advantage of machine learning. Additionally, real-time analysis empowers agile decision-making, requiring minimal technical expertise, and enabling people to efficiently address complex problems. Overall, machine learning optimizes operations, makes data-driven choices, and maintains competitiveness [16].

The prevalent two-step approach to classification problems involves model development or training and model testing or deployment. During the model development phase, a dataset containing input data along with actual class labels is utilized. Following the training of the model, it goes through testing against new samples or information to assess the performance of the model. Subsequently, the model is implemented for real-world applications, predicting classes for new data instances where the class labels are unknown. The simple split, also known as holdout or test sample estimation, involves dividing the data into two distinct subsets referred to as the training set and the test set (or holdout set). A common practice is to allocate two-thirds of the data as the training set, while the remaining one-third serves as the test set. The training set is utilized by the inducer or model builder to construct the classifier. Subsequently, the developed classifier is assessed on the test set to evaluate its performance. A weakness of this approach lies in its assumption that the data in the two subsets share identical properties, essentially assuming homogeneity. Given that this method relies on simple random partitioning, it may not accurately reflect the complexities present in real-world datasets, particularly when the distribution of the classification variable is skewed. To solve this restriction, stratified sampling is recommended as an improvement, with the strata defined by the output variable. While this is an improvement over the simple split, it still has some bias due to the single random partitioning. Adopting a process known as $k$-fold cross-validation becomes useful in eliminating the bias associated with a random sampling of training and holdout data samples when testing the predicted evaluation metric. This process, also referred to as rotation estimation, involves the random division of the complete dataset into $k$ mutually exclusive subsets, each approximately of equal size. The classification model passes through training and testing iteratively $k$ times. The model is trained on all but one fold at each iteration and then tested on the remaining single fold. The evaluation metric of

the model is then estimated through cross-validation by averaging the $k$ individual measures [152]. For a better understanding, Figure 4 is a possible representation of $k$-Fold Cross-Validation to classify data into a cat or dog.

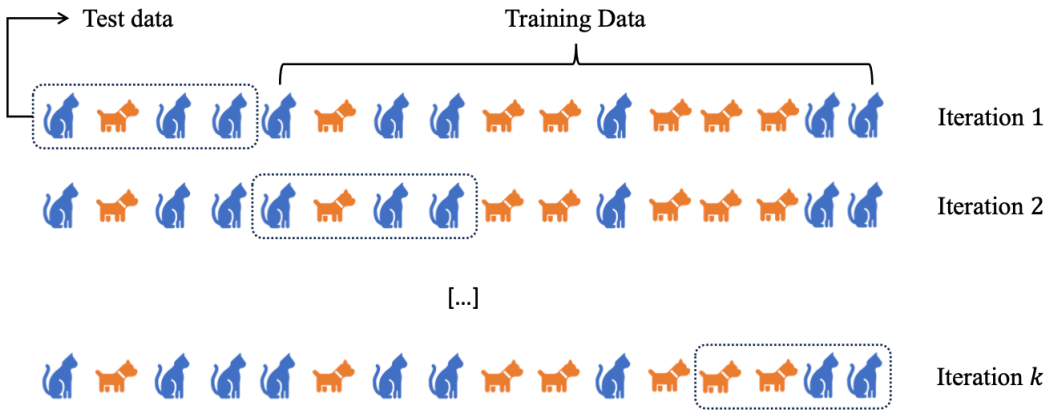



Figure 4: $k$-Fold Cross-Validation (Based on [152]).

A pivotal consideration in the training of machine learning models is the challenge of overfitting, a phenomenon where the model becomes excessively fitted to the training dataset, leading to suboptimal performance on external datasets. Overfitting significantly limits the model's generalizability. Regularization strategies are aimed at preventing models from overfitting by defining constraints for the model parameters or the performance function. The goal of adopting regularization is to establish a balance that allows the model to capture patterns in the data without being excessively sensitive to noise or variations within the training set, hence improving the model's ability to robustly generalize to new and unknown data [152]. Conversely, when a model exhibits a substantial number of errors in the training data, it indicates a condition known as high bias or underfitting. Thus, underfitting signifies the model's inadequacy in accurately predicting the labels of the data on which it was trained. Various factors can contribute to underfitting, with two primary reasons standing out: simplicity of the model and features lack of informativeness, or failure to capture the essential characteristics of the data [29].

Another concept of machine learning is the application of feature selection. Feature selection is a critical aspect of machine learning, as elucidated in [178]. This process involves the careful selection and optimization of input variables to improve the performance and interpretability of machine learning models. The significance of feature selection lies in its ability to reduce model complexity, mitigate overfitting, and improve overall model efficiency. By strategically choosing relevant features, unnecessary noise, and irrelevant information are minimized, leading to models that are not only more interpretable but also computationally more efficient. The application of feature selection is critical, especially in scenarios where datasets may contain a large number of features, many of which may not contribute significantly to predictive accuracy. In essence, the incorporation of feature selection is indispensable for optimizing the utility and performance of machine learning models in diverse and complex data environments. In the field of feature selection for machine learning optimization, the integration of genetic algorithms presents

a powerful and evolutionary-driven approach. Inspired by the principles of natural selection and genetics, genetic algorithms navigate complex solution spaces to identify optimal subsets of features. The process begins with the random generation of a population of potential solutions, where each solution represents a unique set of features. These solutions are evaluated using a fitness function, which quantifies their effectiveness in the given task. Solutions with better fitness are more likely to be selected for the following generation, reflecting natural selection. Crossover, simulating genetic recombination, involves exchanging genetic information between selected solutions. Mutation introduces random modifications, allowing for the discovery of new solution spaces. The algorithm refines its answers through recurrent generations, eventually converging on feature subsets that improve model performance. Genetic algorithms offer a robust and adaptable approach to feature selection, particularly suited for tasks where the solution space is vast and complex. Their ability to explore diverse combinations of features makes them a valuable tool in optimizing machine learning models for various domains and datasets [178].

The evaluation of classifier models on a given dataset involves the application of various assessment metrics to evaluate the effectiveness of classification and to facilitate model comparisons. In this context, inferential statistics play an important role in determining the impacts of independent variables on the inherent variability within the dependent variable. Hypothesis tests, a fundamental component of inferential statistics, are employed to extract valuable insights about the dataset under investigation. Two distinct categories of errors, known as Type I and Type II errors, come into consideration. Type I error occurs when the null hypothesis is wrongly rejected despite its truth, whereas Type II error involves the failure to reject the null hypothesis when it is, in fact, false [18]. Mitigating these errors is essential, though not always straightforward. The level of significance controls Type I error, representing the acceptable risk threshold for erroneously rejecting a true null hypothesis. Conversely, Type II error is influenced by sample size, being sensitive to the number of observations in the sample. Consequently, an increase in the number of observations tends to decrease the likelihood of committing a Type II error. Managing these errors is a critical aspect of statistical analysis, requiring a careful balance between the acceptable risk and the robustness of the sample size [142]. On the other hand, the assessment of a classifier often involves the use of a confusion matrix, a well-established tool that considers the accurate classification of positive and negative instances, denoted as True Positive (True Positive (TP)) and True Negative (True Negative (TN)) respectively [2]. This matrix also includes cases where negatives are predicted but are real positives, referred to as False Negatives (False Negatives (FN)), and instances where positives are predicted but are real negatives, termed as False Positives (False Positives (FP)) [2]. An illustrative example of a confusion matrix that establishes the actual classification against the classification predicted is provided in Table 7.

Various metrics can be derived from the confusion matrix to assess the performance of the applied model. One widely recognized metric is accuracy, which measures the correctness of classifications, Equation 3.1 [54].

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + FP + TN} \tag{3.1}$$

Table 7: Confusion Matrix Example.

| | **Real** | |
|---|---|---|
| **Predicted** | Positive | Negative |
| Positive | TP | FP (Type I Error) |
| Negative | FN (Type II Error) | TN |

However, accuracy, while fundamental, exhibits limitations, particularly in sensitive domains like healthcare, where misclassifications can lead to significant consequences [81]. In response to this constraint, additional metrics such as precision, recall rate, specificity, F-Measure, Area Under the ROC (receiver operating characteristic curve) Curve (AUC), and Type I and II errors are considered. These metrics offer a more nuanced evaluation and can be extracted from the information provided by the confusion matrix. Precision (Equation 3.2 [54]) reflects the model's ability to correctly identify positive instances, providing a measure of how trustworthy positive predictions are.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3.2}$$

Recall rate (Equation 3.3 [54]), or sensitivity, is the model's capacity to capture all actual positive instances, minimizing the risk of false negatives.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3.3}$$

Specificity (Equation 3.4 [54]), on the other hand, focuses on accurately identifying negative instances, crucial in scenarios where the consequences of false positives are substantial.

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{3.4}$$

F-Measure strikes a balance between precision and recall, offering a composite metric that considers both false positives and false negatives, Equation 3.5 [54].

$$F_1 = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \tag{3.5}$$

Moreover, AUC provides a comprehensive assessment of a model's discriminative ability across different thresholds, offering a summarized measure of its overall performance, as demonstrated in Equation 3.6 [158].

$$\text{AUC} = \frac{1}{2}(\text{Specificity} + \text{sensitivity}) \tag{3.6}$$

Lastly, Type I error, or false positive rate, measures instances where the model incorrectly identifies negatives as positives, while Type II error, or false negative rate, is for situations where the model fails to recognize true positives [143].

$$\begin{aligned} \text{Type I Error} &= \frac{FP}{FP + TN} \\ \text{Type II Error} &= \frac{FN}{FN + TP} \end{aligned} \tag{3.7}$$

By considering these metrics, the evaluation framework becomes more robust, especially in domains where misclassifications carry significant implications, such as healthcare [152, 142].

In the field of regression analysis, Mean Square Error (Mean Square Error (MSE)), Root Mean Square Error (Root Mean Square Error (RMSE)), and Mean Absolute Error (Mean Absolute Error (MAE)) stand out as widely recognized metrics. Therefore, MSE is a commonly used metric that calculates the average squared difference between predicted and actual values (Equation 3.8), where $n$ is the number of data points, $y_i$ is the actual value, and $\hat{y}_i$ is the predicted value. It can be observed that MSE places greater penalties on larger errors due to the squaring operation, thereby increasing sensitivity to outliers and amplifying their influence on the overall error [41].

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{3.8}$$

In contrast, RMSE (Equation 3.9), as the square root of MSE, offers a metric to measure the average magnitude of prediction errors. RMSE is reported in the same units as the target variable, which improves its interpretability. Its popularity arises from the balanced perspective it presents, which balances mistake sensitivity with a high level of interpretability [41].

$$RMSE = \sqrt{MSE} \tag{3.9}$$

Conversely, MAE (Equation 3.10), as a metric, computes the average absolute difference between predicted and actual values. Unlike MSE, MAE does not magnify the impact of larger errors. It is less susceptible to outliers and provides a more balanced perspective of the model's performance [41].

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{3.10}$$

These measures are necessary for evaluating the accuracy and performance of regression models. While MSE and its square root, RMSE, provide a more comprehensive understanding of the errors in a regression model by squaring the errors and considering their average, MAE offers a simpler alternative. Therefore, MAE calculates the average absolute difference between the predicted and actual values without squaring the errors. This makes MAE a straightforward and intuitive measure of model performance. Thus, MAE may be preferred in situations where extreme errors should not be overly emphasized. For example, if a model is being used in a safety-critical application where even a few extreme errors could have serious consequences, MAE might be a more appropriate choice. Additionally, MAE is less sensitive to outliers compared to MSE, making it a robust metric in datasets with outliers. It's important to note that the choice between MSE, RMSE, and MAE depends on the specific requirements and characteristics of the problem at hand. While MSE and RMSE are commonly used when large errors should be penalized more heavily, MAE provides a simpler and more direct measure of overall error magnitude [170].

## 3.2 Machine Learning Models

This section presents the mathematical formulations of various machine learning models, exploring their primary strengths and weaknesses along with their diverse applications.

### 3.2.1 Support Vector Machine

The SVM is recognized for its simplicity and adaptability in addressing classification or regression challenges. The primary goal of SVM is to discern and establish a hyperplane capable of effectively classifying data points, as stated in [129]. SVM's operate within the field of supervised learning, deriving input-output functions from a labeled set of training data. This produces functions capable of either classification, facilitating the assignment of cases to predefined classes, or regression, enabling the estimation of continuous numerical values for desired outputs. Nonlinear kernel functions are frequently employed in classification to transform input data (typically expressing highly complicated nonlinear relationships) to a high-dimensional feature space where the input data becomes linearly separable. The maximum-margin hyperplanes are then built to optimally separate the output classes in the training data. In the context of a classification-oriented challenge, it is a common observation that numerous linear classifiers (hyperplanes) have the capability to partition the data into distinct subsections, each corresponding to a specific class. Nevertheless, among these classifiers, only one hyperplane stands out by accomplishing the maximum separation between the classes. This is illustrated in Figure 5, where the hyperplane, along with the two maximum-margin hyperplanes, effectively separates the cats from the dogs.

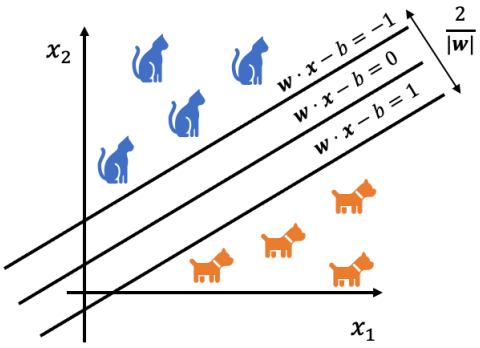



Figure 5: Support Vector Machine Separation (Based on [152]).

SVM data can have more than two dimensions, that is, two separate classes. This approach is a linear classifier, aiming to identify the $(n-1)$ hyperplane that maximizes the distance from the hyperplanes to the nearest data points, where $n$ represents the number of dimensions or class labels. The underlying assumption is that a larger margin or distance between these parallel hyperplanes improves the generalization power of the classifier, thereby improving the predictive capabilities of the SVM model. The mathematical representation of these hyperplanes involves quadratic optimization modeling. Such hyperplanes are specifically referred to as maximum-margin hyperplanes, and the linear classifier derived

from this methodology is known as a maximum-margin classifier. Let $(x_1, c_1), \ldots, (x_n, c_n)$ denote the set of data points derived from the training dataset, where $c$ represents the class label associated with the response variable and $\mathbf{x}$ as the input vector. In other words, each data point is an $m$-dimensional real vector with scaled $[0, 1]$ or $[-1, 1]$ values. Normalization and/or scaling are critical steps in preventing variables/attributes with higher variance from dominating the classification models. Considered as training data, these instances represent the correct classifications for the SVM to attain through the establishment of a dividing hyperplane expressed in Equation 3.11, where $\mathbf{w}$ is perpendicular to the separating hyperplane, crucial for optimal performance. Without it, the hyperplane is constrained to pass through the origin, limiting the solution space.

$$\mathbf{w} \cdot \mathbf{x} - b = 0 \tag{3.11}$$

The interest is to maximize the margin and take into consideration the support vectors and the parallel hyperplanes nearest to these support vectors in each class relative to the optimal hyperplane. These parallel hyperplanes can be defined through Equation 3.12, as demonstrated.

$$\begin{aligned} \mathbf{w} \cdot \mathbf{x} - b &= 1 \\ \mathbf{w} \cdot \mathbf{x} - b &= -1 \end{aligned} \tag{3.12}$$

When confronted with linearly separable training data, a strategic choice is made in selecting hyperplanes to establish clear separation without any interposing data points. Employing geometric principles, the calculated distance between these hyperplanes is defined as $\frac{2}{|w|}$, with the objective of maximizing this distance translating to the minimization of the norm of the weight vector ($|w|$). To eliminate data points, it is important to guarantee that the conditions in Equation 3.13 hold true for all instances denoted by $i$ [152].

$$\begin{aligned} \mathbf{w} \cdot \mathbf{x_i} - b &\geq 1 \\ \mathbf{w} \cdot \mathbf{x_i} - b &\leq -1 \end{aligned} \tag{3.13}$$

Moreover, there are cases, where data can not be linearly separable, and non-linear transformations, can be applied to transform data into higher-dimensional space, known as kernel function. The most well-known kernel functions are the polynomial, radial basis function, and neural network kernel, defined by the Equations 3.14, 3.15 and 3.16. The selection of the kernel function relies on the characteristics of the data and the specific problem under consideration.

$$\text{dth-Degree polynomial: } K(x, x') = (1 + \langle x, x' \rangle)^d \tag{3.14}$$

$$\text{Radial Basis Function (RBF): } K(x, x') = exp(-\gamma ||x - x'||^2) \tag{3.15}$$

$$\text{Neural Network: } K(x, x') = tanh(k_1 \langle x, x' \rangle + k_2) \tag{3.16}$$

The implementation of SVM has several advantages that make it a valuable tool for a wide range of machine-learning applications. Firstly, the management of high-dimensional data, which is pertinent in

today's data-rich environments. The capability of maximizing the margin between classes, through the identification of an optimal hyperplane, contributes to robust generalization, and consequently, mitigates the risk of overfitting. Furthermore, the application of kernel functions enables the application of non-linear problems, detecting complicated patterns that other systems may miss. Another interesting fact is the robustness to outliers since SVM considers convex optimization ensuring the achievement of the global optimum. However, there are some disadvantages to the application of the SVM model, like the sensitivity to choose the kernel function, and making an incorrect choice can lead to suboptimal results. Another drawback is the computational intensity associated with SVM training, particularly when dealing with large datasets. The algorithm's complexity can be less suitable for real-time applications. Furthermore, it can be a challenge to understand the decision boundaries, especially in high-dimensional feature spaces. Lastly, SVM do not directly provide probabilistic outputs, which can be a limitation in applications where probabilistic confidence estimates are required [153]. SVM has been applied in various research domains, particularly in brain disease analysis and cancer diagnosis. In the field of neuroscience, it has played an important role in diagnosing neurological disorders such as Alzheimer's disease, schizophrenia, and depression. Additionally, SVM facilitates neuroimaging analysis, enabling the examination of vast datasets related to cancer, which, in turn, has contributed to innovative discoveries in drug development and a deeper comprehension of cancer-related genes, as discussed in [77]. Furthermore, SVM has proven its effectiveness in breast cancer diagnosis and classification, as demonstrated in [136].

### 3.2.2 Logistic Regression

Logistic Regression (Logistic Regression (LR)) is a statistical method for modeling the relationship between a binary response variable and one or more independent variables. One example of the response variable could be the classification of an individual's state as either drowsy or awake. Let's assume a simple linear model like in Equation 3.17, where $Y_i$ represents the response variable that has 0 or 1 values, $x_i$ the independent variable, $\beta_0$ the intercept term, $\beta_1$ the regression coefficient, and $\epsilon$ the random error.

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i \tag{3.17}$$

Assuming that the response variable follows a Bernoulli distribution, then $P(Y_i = 1) = \pi_i$ and $P(Y_i = 0) = 1 - \pi_i$. Besides that, the expected value for a discrete random variable $X$ is known as $E[x]$ or $\mu$, and is computed by Equation 3.18, where $f(x)$ is denominated as probability mass function.

$$f(x_i) = P(X = x_i)$$

$$\mu = E[X] = \sum_x x \cdot f(x) \tag{3.18}$$

Therefore, since $E[\epsilon_i] = 0$, the expected value for the response variable $Y_i$ can be computed as follows,

$$E[Y_i] = \sum_{i=0}^{1} x \times f(x) = 0 \times (1 - \pi_i) + 1 \times \pi_i = \pi_i$$

and this statement is equivalent to expressing that the expected value of the response variable corresponds to the probability of the response variable attaining the value 1.

$$E[Y_i] = \beta_0 + \beta_1 x_i = \pi_i$$

Based on this information, it can establish the following constraint:

$$0 \leq E[Y_i] = \pi_i \leq 1$$

however, this can be an issue when implementing Equation 3.17, as it may result in predicted values of the response variable falling outside the interval $[0, 1]$. For that reason, the response function must be nonlinear and the logit response was defined as in Equation 3.19 [112].

$$E[Y] = \frac{e^{\beta_0+\beta_1 x}}{1 + e^{\beta_0+\beta_1 x}} = \frac{1}{1 + e^{[-(\beta_0+\beta_1 x)]}} \tag{3.19}$$

Considering the presence of multiple independent variables, the linear model can be represented as shown in Equation 3.20, where $p$ denotes the number of independent variables. In this model, the response variable $Y_i$ is expressed as a linear combination of the independent variables $x_{1i}, x_{2i}, \ldots, x_{pi}$, along with an intercept term $\beta_0$ and regression coefficients $\beta_1, \beta_2, \ldots, \beta_p$ associated with each independent variable.

$$Y_i = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_p x_{pi} + \epsilon_i \tag{3.20}$$

Here, $\epsilon_i$ represents the error term. The logit response, which transforms the linear combination of independent variables into probabilities, is expressed by Equation 3.21 [76].

$$E[Y] = \frac{e^{\beta_0+\beta_1 x_1+\ldots+\beta_p x_p}}{1 + e^{\beta_0+\beta_1 x_1+\ldots+\beta_p x_p}} = \frac{1}{1 + e^{-(\beta_0+\beta_1 x_1+\ldots+\beta_p x_p)}} \tag{3.21}$$

When considering the advantages of the LR model, it stands out for its ease of comprehension and subsequent implementation. Moreover, it exhibits computational efficiency and through the coefficients, it is possible to extract more information about the relationship between the independent and the response variables. Besides that, it also stands for presenting output probabilities, which can be useful when making decisions based on the likelihood of an event occurring. In contrast, the performance of the model can be affected by the lack of linearity between the predictors and can be sensitive to outliers or noisy data. Moreover, LR is not suitable for high-dimensional data since LR may not fare well when dealing with a large number of predictor variables, as it can lead to overfitting [117]. It is evident that LR has a wide range of applications across various fields, including the physical sciences [101], economics [27], and political sciences [87]. Nevertheless, the majority of studies found in the literature predominantly pertain to the area of healthcare. For instance, these studies encompass a wide range of topics such as the assessment of lung cancer risk, post-surgical mortality rates [24], the mental well-being of students, and the relationship between various factors known to influence mental health [119]. Furthermore, a similar investigation was conducted to compare the quality of life among COVID-19 patients and the influencing factors for COVID-19 patients [39].

### 3.2.3 Linear Discriminant Analysis

Linear Discriminant Analysis (Linear Discriminant Analysis (LDA)) is a classification model implemented for dimensionality reduction and classification. This statistical technique aims to identify a linear transformation of the original feature space to a lower one that maximizes the separation between different classes of data. In other words, when there are multiple independent features that describe the data, LDA constructs a linear combination of these features to maximize the mean differences between the response variable classes. Two measures must be calculated: *within-class* and *between-class* scatter matrices as shown in Equation 3.22 and 3.23, where $x_i^j$ is the $i$th sample of class $j$, $\mu_j$ the mean of class $j$, $c$ the number of classes, $N_j$ the number of observations of the class $j$ and $\mu$ the mean value of all classes.

$$S_w = \sum_{j=1}^{c} \sum_{i=1}^{N_j} (x_i^j - \mu_j)(x_i^j - \mu_j)^T \tag{3.22}$$

$$S_b = \sum_{j=1}^{c} (\mu_j - \mu)(\mu_j - \mu)^T \tag{3.23}$$

The objective is to improve the between-class measure while simultaneously reducing the within-class measure. One approach to achieve this goal is by maximizing the ratio between the determinant of the between-class scatter matrix and the determinant of the within-class scatter matrix.

$$\text{ratio} = \frac{det|S_b|}{det|S_w|}$$

When $S_w$ is a non-singular matrix the maximization of the ratio occurs when the column vectors of the projection matrix, $\mathbf{W}$, correspond to the eigenvectors of $\mathbf{S_w^{-1}S_b}$ [106]. The implementation of an LDA model involves three primary steps. Initially, it requires the computation of the distances between the means of different classes, followed by calculating the distances between the class means and samples within each class. Finally, the objective is to construct a lower-dimensional space that maximizes between-class variance while minimizing within-class variance. Therewith, LDA presents different advantages starting with the dimensionality reduction and the ability to maximize class separation is a key strength, leading to improved classification. It is also effective for multiclass problems, is less prone to overfitting, making it suitable for datasets with limited samples, and is robust to outliers. Conversely, LDA has limitations when the number of dimensions significantly exceeds the number of samples in the data, complicating its ability to find an effective lower-dimensional space. Moreover, in cases where classes exhibit non-linear separability, LDA struggles to differentiate between these classes due to its inherent linearity assumption [161]. LDA has found applications across diverse fields, including medical diagnosis, where it has been utilized for tasks such as staging Alzheimer's disease [99], diagnosing breast cancer [162], and detecting Parkinson's disease [7]. In the realm of marketing and customer segmentation, LDA serves as a valuable tool for enhancing various aspects, such as customer churn prediction [177], among other applications [51]. Image processing [95], biology [113], and quality control [8] are just a few examples of the diverse domains where LDA has demonstrated its versatility and utility.

### 3.2.4 Naïve Bayes

The NB model is a probabilistic classification method derived from Bayes' theorem, commonly employed in machine learning for solving classification problems. This method is suitable for cases where the response variable is characterized by nominal values, while the independent variables may encompass a combination of both numeric and nominal types. The term "Naïve" derives from its foundational assumption of independence among the independent variables. In essence, the NB classifier posits that independent variables are unrelated to one another and that the presence or absence of a particular variable among the predictors is unrelated to the presence or absence of any other variables. This independence assumption simplifies the model [152]. Let's consider the definition of conditional probability in Equation 3.24, where $A$ represents the hypothesis or response variable class and $B$ is the data or evidence.

$$P(A \cap B) = P(A|B) \times P(B) = P(B \cap A) = P(B|A) \times P(A) \tag{3.24}$$

Thus, $P(A|B)$ is the probability of class $A$ given evidence $B$, $P(B|A)$ is the probability of evidence $B$ given class $A$, $P(A)$ is the prior probability of class $A$, and $P(B)$ is the prior probability of evidence $B$. From this information, the probabilities can be computed the Equation 3.25.

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)} \text{ for } P(B) > 0 \tag{3.25}$$

Considering $E_1$, $E_2$, ..., $E_k$ that are $k$ mutually exclusive and $B$ any event, then Bayes theorem can be expressed like in Equation 3.26 [112].

$$P(Y_1|B) = \frac{P(B|E_1) \times P(E_1)}{P(B|E_1) \times P(E_1) + ... + P(B|E_k) \times P(E_k)} \tag{3.26}$$

In the context of probabilistic reasoning, the term $P(A)$ represents the prior probability of the event $A$. This prior probability remains unaffected by any specific information concerning event $B$. On the other hand, $P(A|B)$ represents the posterior probability of $A$ given the occurrence of $B$ This probability is often referred to as the posterior probability, as it relies on and derives from the particular value of $B$ specified. Conversely, $P(B|A)$ signifies the conditional probability of $B$, given the occurrence of $A$. This probability is commonly known as the likelihood. Finally, $P(B)$ represents the prior probability of $B$, which is also designated as the evidence. According to these terminologies, the posterior probability of $A$ given $B$ can be denoted as Equation 3.27. These terminologies and concepts play a fundamental role in probabilistic reasoning and decision-making processes [152].

$$Posterior = \frac{Likelihood \times Prior}{Evidence} \tag{3.27}$$

In alignment with other machine learning methodologies, the NB algorithm embraces a structured two-phase model known as training and testing. **Training Phase:** During this initial phase, the model and its associated parameters are estimated using the available training data. This step serves as the foundation upon which the classifier or predictor is built.

1. Gather data, clean the noise, and organize the information collected where the rows represent the cases and columns the variables under study.

2. Computation of the prior probabilities for all class labels associated with the response variable. This process establishes the baseline probabilities for each potential outcome, independent of any predictor variables. Subsequently, the likelihoods are determined for all independent variables, considering their various potential values concerning the response variable. This computation enables an assessment of how each independent variable relates to the potential outcomes of the response variable, thereby aiding in the modeling and analysis of the relationships between variables. In this context, the estimation of probabilities plays a critical role. To ensure the accuracy and reliability of these estimates, it is imperative to employ an appropriate methodology. When dealing with independent variables of a categorical nature, the estimation of likelihood (conditional probability) is derived from the proportion of training samples corresponding to a specific variable value in relation to the classes of the response variable. For numerical variables, a different approach is adopted. It involves the computation of both the mean and variance for each independent variable within each class. Subsequently, the likelihood is determined by applying the following formula, Equation 3.28, where $c$ is the number of classes of $Y$ and $v$ is an independent variable.

$$P(x = v|c) = \frac{1}{\sqrt{2\pi\sigma_c^2}} e^{-\frac{(v-\mu_c)^2}{2\sigma_c^2}} \tag{3.28}$$

**Testing Phase:** In the subsequent phase, known as testing, the trained model is applied to new, unseen cases for classification. This phase evaluates the model's performance on real-world data and assesses its ability to generalize from the training set to new instances. Utilizing the probabilities generated during the training phase, it becomes possible to assign a new sample to a class label through the application of the following Equation 3.29

$$P(C|F_1, ..., F_n) = \frac{P(C)P(F_1, ..., F_n|C)}{P(F_1, ..., F_n)} \tag{3.29}$$

Since the denominator remains constant (identical for all class labels), it can be omitted from the equation, resulting in the following simplified formula, which essentially represents the joint probability.

$$classify(f_1, ..., f_n) = argmax p(C = c) \prod_{i=1}^{n} p(F_i = f_i|C = c) \tag{3.30}$$

This structured approach to model development and evaluation is a fundamental practice within the field of machine learning, including the NB method. It ensures that the model is not only trained effectively but also capable of making accurate predictions on novel data points. The NB classification technique possesses several advantageous characteristics that make it a valuable tool in data classification. It is notably robust, demonstrating the ability to perform effectively even when data deviates from the model's

assumptions. Is less prone to overfitting, making it suitable for small sample data. Estimating fewer parameters increases computational efficiency. One limitation is the zero observations problem, wherein the model may not be reliable when training data lack specific cases that are theoretically possible. In response, it is recommended to assign low probabilities to such cases instead of zero probabilities. Additionally, model performance can be impacted by the presence of highly correlated features in the dataset, as correlated features may exert undue influence [169]. Therefore, NB has been applied in many fields, such as medical domain applications, including cases related to severe head injuries, breast cancer, heart disease, thyroid issues, liver disease, and abdominal conditions. Additionally, has been applied in email spam filtering, even when selected features exhibit dependencies, software defect prediction, and cybersecurity [169].

### 3.2.5 K-Nearest Neighbors

The kNN algorithm stands as a non-parametric learning approach and retains the entirety of its training examples within memory. With a previously unseen example denoted as $x$, the kNN algorithm identifies the $k$ training instances that are most proximate to $x$ and subsequently provides either the majority label (in the context of classification) or the average label (for regression tasks) [29]. In other words, the kNN algorithm stands out as one of the most simpler machine learning techniques. In scenarios involving classification-type, the algorithm assigns a case to a particular class based on the majority vote of its neighbors. Specifically, the case is categorized according to the class that is most prevalent among its $k$ nearest neighbors, with $k$ representing a positive integer. When $k$ is equals to 1, the case is assigned to the class of its closest neighbor. One important choice involves establishing the similarity measure. In the context of the kNN algorithm, the similarity measure assumes the form of a mathematically distance metric. When presented with a new case, kNN generates classifications based on the outcomes of $k$ neighbors that are closest in distance to that specific point. Consequently, it becomes imperative to define a metric that quantifies the distance between the new case and the instances within the dataset. The Euclidean distance stands out as one of the most widely recognized metrics, representing the linear distance between two points within a dimensional space. Another widely used metric is the Manhattan distance (Equation 3.32). It's noticeable that both of these distance measures are specific instances of the more general Minkowski distance. The Minkowski distance (Equation 3.31) is defined as follows, where $i = (x_{i1}, x_{i2}, ..., x_{ip}$ and $j = (x_{j1}, x_{j2}, ..., x_{jp})$ represent two p-dimensional data objects, such as a new case and an example within the dataset. The parameter $k$ is a positive integer.

$$d(i,j) = \sqrt{|x_{i1} - x_{j1}|^k + |x_{i2} - x_{j2}|^k + ... + |x_{ip} - x_{jp}|^k} \tag{3.31}$$

Nevertheless, it is important to note that the Manhattan distance can be computed when setting $k = 1$, in Equation 3.32, while the Euclidean distance corresponds to $k = 2$, as indicated in Equation 3.33.

$$d(i,j) = \sqrt{|x_{i1} - x_{j1}| + |x_{i2} - x_{j2}| + ... + |x_{ip} - x_{jp}|} \tag{3.32}$$

$$d(i,j) = \sqrt{|x_{i1} - x_{j1}|^2 + |x_{i2} - x_{j2}|^2 + ... + |x_{ip} - x_{jp}|^2} \tag{3.33}$$

It is clear that these metrics are for numeric data. However, when dealing with nominal data, alternative approaches are necessary. In the case of a multi-value nominal variable, it is essential to use a suitable encoding method, such as one-hot encoding [152]. One of the main strengths of kNN is its simplicity, making it easy to understand and implement. Additionally, kNN is known for its fast training process, allowing for quick deployment in various scenarios. The algorithm also exhibits robustness to noisy training data, making it suitable for datasets with some degree of uncertainty or inconsistency. Moreover, kNN demonstrates good performance when dealing with datasets containing several labels, demonstrating its versatility in multi-class classification tasks. However, these merits are accompanied by notable drawbacks. The algorithm incurs a high computational cost, particularly as the dataset size grows, which can be a limiting factor in resource-intensive environments. Furthermore, its run-time performance may suffer, especially when applied to large datasets, leading to suboptimal efficiency. Additionally, kNN is sensitive to the local structure of the data, meaning that its performance might be influenced by the specific arrangement and distribution of data points, potentially affecting the accuracy in certain scenarios [102]. The health sector stands out as a primary domain for various applications, including disease diagnosis [164] and the discovery of new drugs through the comparative analysis of molecular structures with known properties [114]. Additionally, image analysis plays an important role, as exemplified in the detection of skin lesions [149]. Furthermore, kNN is also utilized in recommendation systems to suggest content based on the preferences of users [157], anomaly detection [40] and forecasting of customer demands for production planning [91].

### 3.2.6 Extremely Randomized Trees

The Extremely Randomized Trees, also designated as Extra-Trees (ET) algorithm, is a powerful ensemble approach based in the concept of combining a large number of decision trees. Ensembles techniques, widely employed in various applications for classification and regression problems, combines the decisions of different models and makes a decision based on that combination, which results in higher performance than the achievements of a single decision or model. The architecture of the ET algorithm consist of multiple decision trees, each featuring a root node, child or split nodes, and leaf nodes. When presented with a dataset $X$, the algorithm initiates the decision-making process at the root node. Here, ET selects a split rule using a random subset of features and a partially random cut point. This selection process is then iteratively applied at each child node until a leaf node is reached. Three key parameters characterize the ET algorithm: the number of trees in the ensemble ($k$), the number of features selected randomly at each split ($f$), and the minimum number of samples required to split a node ($n_{min}$). Understanding and tuning these parameters are crucial for optimizing the performance of the ET algorithm in various applications. The training process of the ET algorithm involves utilizing the entire learning sample to train each tree. In this approach, the top-down splitting of nodes within the tree is characterized by entirely random splits, as opposed to selecting the best splits. Unlike traditional decision tree algorithms that aim

for optimal splits, ET introduce an additional layer of randomness by employing completely random splits. This distinctive feature contributes to enhanced diversity among the individual trees in the ensemble, promoting robustness and preventing overfitting. Instead of determining the locally optimal cut-point for each attribute based on measures like information gain, the algorithm employs a distinct strategy. A random cut-point is chosen, drawn from a uniform distribution within the empirical range of the attribute within the training set of the tree. This introduces an additional layer of randomness, deviating from the conventional approach of optimizing cut-points. The selection process involves generating multiple random splits, and the split that yields the highest score among these random selections is ultimately chosen to split the node. During the testing phase, a test sample go through evaluation within each decision tree. The process involves navigating through each child node, iteratively selecting the best splits, and directing the test sample towards the appropriate right or left child node until a leaf node is reached. The classification for the test sample within each individual decision tree is determined by the leaf node it arrives at. The final prediction for the test sample is then determined as the majority of votes among the $k$ decision trees constituting the ET algorithm [66]. To gain a more comprehensive insight into the concepts and structure of the ET model, Figure 6 illustrates the workflow and underlying concepts of the ET algorithm.

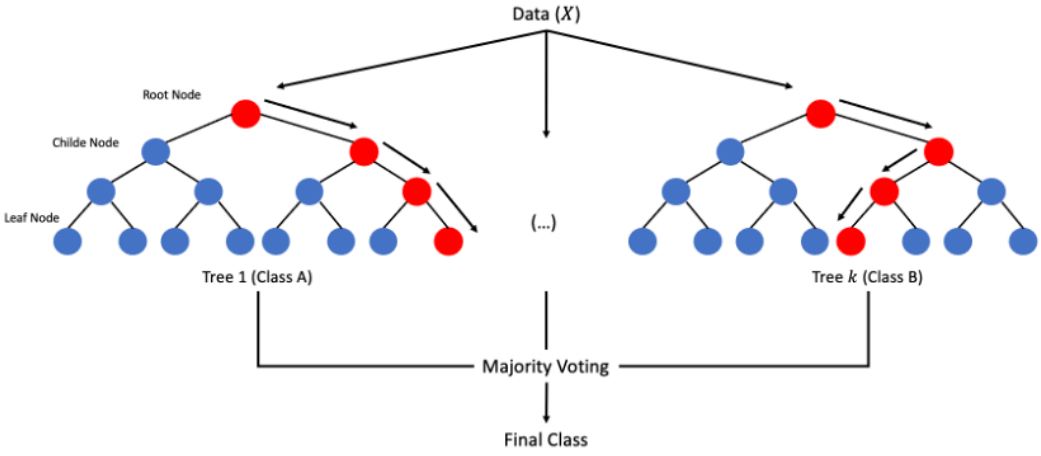



Figure 6: Extremely Randomized Trees Approach (Based on [140]).

In the case of decision tree-based ensembles, optimal performance is achieved when base learners exhibit independence, a quality facilitated by randomization during tree growth, fostering better diversity and reducing correlation. Essentially employing a divide-and-conquer approach, ensemble learning methodologies result in stable and robust classifiers for supervised machine learning tasks, mitigating factors such as noise, bias, and variance for precise predictions. However, it is crucial to acknowledge that the computational costs associated with training numerous individual classifiers can be significantly huge [140]. Furthermore, in the health domain, ET has been successfully utilized for disease diagnosis, as indicated by [17]. Additionally, it was implemented to predict potential risks to fetal well-being and to draw clinical conclusions, as exemplified in the research carried out by [22]. There are also applications for e-mail classification [151], predict the stock market prices [130], and hotel recommendations [125].

### 3.2.7 Adaptive Boosting

Adaptive Boosting (AdaBoost), short for Adaptive Boosting, is an ensemble learning method specifically designed for binary classification tasks, where the goal is to classify instances into one of two classes: positive $(+1)$ or negative $(-1)$. Boosting represents a machine learning methodology that revolves around the concept of constructing a remarkably accurate prediction rule through the fusion of numerous weak and less accurate rules [64]. The context involves having $m$ labeled training examples or the number of observations denoted as $(x_1, y_1), ..., (x_m, y_m)$, where the $x_i$ belong to a domain $X$, and the labels $y_i \in \{1, +1\}$, like in SVM model. For each round, $t = 1, ..., T$, a distribution $D_t$ is calculated, across the $m$ number of observations. Each training instance is initially assigned an equal weight. Subsequently, a specific weak learning algorithm is employed to identify a weak hypothesis $h_t : X \rightarrow \{1, +1\}$. The objective of the weak learner is to ascertain a weak hypothesis with a minimal weighted error $\epsilon_t$ relative to the distribution $D_t$. In other words, the algorithm iterates through a fixed number of rounds ($t$), and at each round a weak learner is trained on the training data, with a specific emphasis on instances that were misclassified in the preceding rounds. The error $\epsilon_t$ for the weak learner is then computed as the sum of the weights assigned to incorrectly classified instances, divided by the total weight. The weight $(\alpha_t)$ of the weak learner in the final prediction is determined based on its error, with higher accuracy resulting in a greater influence on the overall model. Subsequently, the weights of the training instances are adjusted, assigning higher weights to instances that were misclassified by the weak learner, thereby elevating their significance in subsequent rounds. The final prediction, $F(x)$, is a weighted combination of the weak learners' predictions, as expressed in Equation 3.34, where $t$ represents the number of rounds, $(\alpha_t)$ is the weight assigned to the weak learner at round $t$, and $h_t(x)$ is the prediction of the weak learner at round $t$ [148].

$$F(x) = \sum_{t=1}^{T} \alpha_t h_t(x) \tag{3.34}$$

The AdaBoost algorithm has gained popularity in machine learning due to its notable advantages. Firstly, it demonstrates a low generalization error, contributing to its effectiveness in classifications. Its simplicity of code, facilitates its widespread adoption, making it accessible to both novice and experienced practitioners. Furthermore, is computationally efficient and it is suitable for large-scale applications. Besides that, its versatility extends to complex tasks, offering a robust solution for complex learning scenarios. It is also simple to modify and integrate the technique with other machine learning algorithms. Despite these strengths, there are also limitations. In particular, it is sensitive to outliers or noise, which may impact its performance in the presence of anomalous data points. Consequently, it is important to perform a preprocessing analysis to reduce the maximum noise as possible. Additionally, a small dataset can result in a bad performance of the classification model and the algorithm's ensemble nature may lead to a complex and challenging interpretation of the models [85]. AdaBoost model demonstrates diverse applications in critical health-related domains, contributing significantly to the classification of various medical conditions. It has been employed in the classification of breast cancer, diabetic retinopathy, mental health disorders,

and thyroid conditions, as evidenced by the study [73]. Beyond health applications, it plays an important role in computer vision, where it is utilized for tasks such as human activity recognition [174]. In the domain of speech recognition, the model has proven valuable, like in the detection of depression through patient speech [25]. Additionally, it also finds relevance in natural language processing applications, exemplified by its use in predicting the risk of ICU readmission [121].

### 3.2.8 eXtreme Gradient Boosting

eXtreme Gradient Boosting (XGB) stands out as a ensemble learning technique designed to combine a sequence of weak learning models into a robust learning model, thereby achieving better overall performance of machine learning systems. An illustration of the application of a tree ensemble model is shown in Figure 7.

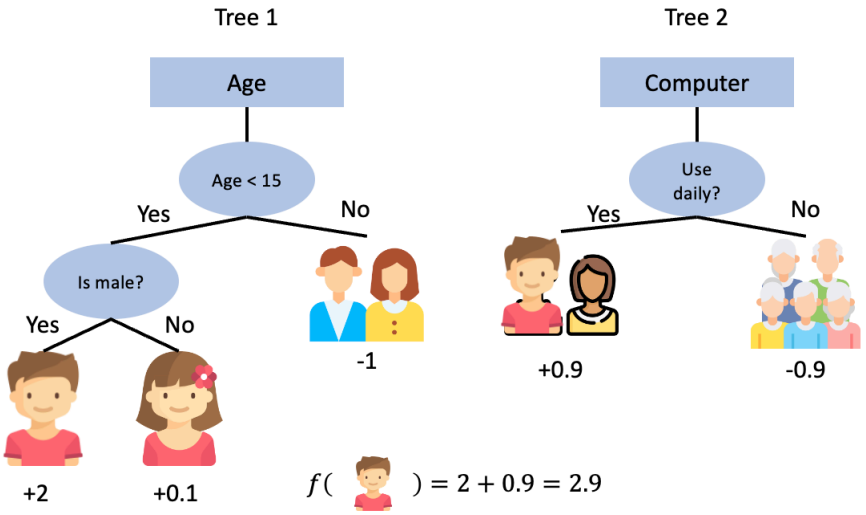



Figure 7: An Example of a Tree Ensemble Model (Based on [38]).

Thus, the XGB model's formulation concerns optimizing an objective function, which is a composite of a loss function and regularization terms. Analyzing a training dataset denoted as $(x_i, y_i)$, where $x_i$ is the feature vector for the $i$-th observation, $y_i$ represents the corresponding response variable, $n$ stands for the number of observations, and $m$ indicates the number of features. A tree ensemble model is employed to predict the response variable ($\hat{y}_i$), defined in Equation 3.35 through $K$ additive functions. Here, $\mathcal{F}$ denotes the space of regression trees, $q$ captures the structure of each tree, mapping an example to its corresponding leaf index, and $T$ stands for the number of leaves in the tree. Furthermore, $f_k$ represents an independent tree structure $q$ with associated leaf weights $w$, computed for each leaf as a continuous

score. In this context, $w_i$ signifies the weight score on the $i$-th leaf.

$$\begin{aligned} \hat{y}_i &= \phi(\mathbf{x}_i) = \sum_{k=1}^{K} f_k(\mathbf{x}_i),\ f_k \in \mathcal{F}, \\ \mathcal{F} &= \{f(\mathbf{x}) = w_{q(\mathbf{x})}\}(q : \mathbb{R}^m \rightarrow T, w \in \mathbb{R}^T) \end{aligned} \tag{3.35}$$

Decision rules within the trees are executed to categorize the leaves, determined by $q$. The ultimate prediction or classification is then obtained as the summation of the weight scores for each leaf, denoted by $w$. To acquire the set of functions utilized in the model, XGB approach involves the minimization of the regularized objective defined in Equation 3.36. A differentiable convex loss function is denoted by $l$, which quantifies the disparity between the prediction ($\hat{y}_i$) and the target ($y_i$). The selection of the loss function depends the nature of the problem at hand, such as regression or classification. In instances of regression problems, one common choice is the MSE, whereas for classification tasks, alternatives like log loss or hinge loss might be applied. The second term $\omega$ serves as a penalty for the model's complexity, specifically targeting the complexity of the regression tree functions, with $\gamma$ and $\lambda$ representing the regularization parameters. Usually, it incorporates both L1 (LASSO) and L2 (ridge) regularization terms to control the quantity of features and the scale of feature weights. The introduction of this additional regularization term is instrumental in refining the learned weights $t$ to ensure a smoother outcome, mitigating the risk of overfitting. In essence, the regularized objective is designed to favor the selection of a model characterized by straightforward yet effective predictive functions [38].

$$\begin{aligned} \mathcal{L}(\phi) &= \sum_i l(\hat{y}_i, y_i) + \sum_k \omega(f_k) \\ \text{with } \omega(f) &= \gamma T + \frac{1}{2}\lambda ||w||^2 \end{aligned} \tag{3.36}$$

XGB stands out in the field of machine learning for its performance and efficiency, often surpassing alternative algorithms, particularly in scenarios involving extensive datasets. Its strength derives from the incorporation of regularization terms within its objective function, a strategic feature that reduces overfitting and improves the model's adaptability to new data instances. Furthermore, XGB presents feature importance score, facilitating a comprehensive of each feature's impact on the model's predictions. Nevertheless, it is critical to recognize the delicate nature of XGB's hyperparameter tuning process, which can be complex and requires close attention to avoid overfitting. Despite its efficiency, XGB may incur computational intensity, particularly when dealing with extensive datasets. Additionally, the interpretability of models generated by XGB, especially when incorporating deep trees, can be a challenge [38, 105]. This model has been applied across various fields, for instance, in healthcare systems, it has been used to predict diabetes mellitus [171], aid in the medical treatment of acute bronchiolitis [107], and classify fetal health based on cardiotocogram data [131]. Additionally, the model finds applications in spam detection across social media platforms [160] and email systems [116]. It also has applications in forecasting sales volume [176], predicting stock market volatility[168], and even forecasting the spread of COVID-19[58].

## 3.3 Multivariate Statistical Process Control

Anomaly detection is employed to discern patterns within specific data that deviate from the norm, a process often referred to as identifying abnormal data. This form of analysis is crucial since abnormal data points can provide valuable insights into the process, including rare events, and enable the implementation of proactive measures. Anomaly detection is applied in various contexts, including medicine and public health, fraud detection, industrial processes, image processing, text data analysis, and sensor networks. Three established techniques rely on classification, statistical, and clustering methods. Classification methods can be computed by considering techniques such as support vector machine, Bayesian networks, rule-based, neural network approaches, among others. Regarding statistical methods, widely utilized approaches include mixture models, signal processing, process control, and principal component analysis. As for clustering methods, options encompass regular clustering and co-clustering [4, 36].

The process control method serves the purpose of monitoring process performance and promptly detecting anomalies when they arise. As a result, an anomaly can occur when the process deviates from its expected and predefined behavior. Hence, various visual techniques can be employed to comprehend the behavior of the analyzed process, whether focusing on a single feature (univariate analysis) or multiple features (multivariate analysis), exemplified by the use of control charts [60]. This method does come with certain drawbacks. It can potentially lead to misleading outcomes, and interpreting the results can be challenging due to the presence of strong correlations between attributes, a phenomenon often referred to as collinearity. Additionally, other issues arise, including dealing with a high number of attributes, referred to as dimensionality, coping with data noise caused by external factors, and addressing missing data.

A solution that has emerged to streamline the process analysis, reduce dimensionality concerns, and address attribute collinearity is PCA. This technique converts a multitude of features into lower-dimensional spaces, wherein the principal components are represented as linear combinations of the original features. These components are orthogonal and proficient at preserving the essential information from the original data. The first principal components exhibit a higher variability from the original data [60]. Therefore, in mathematical terms, the matrix $\mathbf{X} \in \mathbb{R}^{N \times M}$, represents the data under analysis, taking into account the number of observations (*N*) and the features (*M*) of the process. To enhance the quality of results, it is imperative that the values within $\mathbf{X}$ are standardized, ensuring they possess a mean of zero and a variance of one. In terms of the number of principal components, *R*, to take into consideration it depends on the goal of the analysis. If the objective is to reduce collinearity among the features, the *R* value can match the count of the original features. However, when prioritizing dimensionality reduction, the selection of principal components should favor those with higher variability. Thus, the number of components can be expressed as $R \leq M$, and the principal components are defined by Equation 3.37 as follows:

$$\mathbf{T} = \mathbf{XP} + \mathbf{E} \tag{3.37}$$

In this equation, the scores matrix is denoted as $\mathbf{T} \in \mathbb{R}^{N \times R}$, the loading matrix as $\mathbf{P} \in \mathbb{R}^{M \times R}$, and the unexplained information by the PCA method is known as an error ($\mathbf{E} \in \mathbb{R}^{N \times R}$). Furthermore, the

estimation of the original features $\left(\hat{\mathbf{X}}\right)$ and the residuals $\left(\hat{\mathbf{E}}\right)$ can be expressed using Equation 3.38 and Equation 3.39, respectively. This formulation provides a comprehensive understanding of the PCA process and its constituent matrices[80, 6, 60].

$$\hat{\mathbf{X}} = \mathbf{T}\mathbf{P}^{\mathrm{T}} \tag{3.38}$$

$$\hat{\mathbf{E}} = \mathbf{X} - \hat{\mathbf{X}} \tag{3.39}$$

Therefore, in the context of anomaly detection, two key statistical metrics can be calculated: the Squared Prediction Error (Squared Prediction Error (SPE)), also referred to as the *Q* statistic, and Hotelling's $T^2$. The SPE statistic quantifies the disparity between the original data and the *R*-dimensional subspace, as expressed in Equation 3.40.

$$SPE = \sum_{i=1}^{M} (\mathbf{x}_i - \hat{\mathbf{x}}_i)^2 \tag{3.40}$$

On the other hand, the process's stability is evaluated through Hotelling's $T^2$ statistic, as defined in Equation 3.41. In this equation, $\Lambda \in \mathbb{R}^{R \times R}$ represents the covariance matrix of $\mathbf{T}$, where the score vector for the $i$th observation is represented as $\mathbf{t}_\mathbf{i}^{\mathbf{T}} = t_{i1}, t_{i2}, \ldots, t_{iR}$, and the eigenvalues of the $R$ components are denoted as $\lambda_r$.

$$T_i^2 = \mathbf{t}_\mathbf{i}^{\mathbf{T}-\mathbf{1}} \mathbf{t}_\mathbf{i} = \sum_{\mathbf{r}=\mathbf{1}}^{\mathbf{R}} \frac{\mathbf{t}_\mathbf{r}^2}{\lambda_\mathbf{r}} \tag{3.41}$$

Finally, anomalies can be detected by comparing the computed statistics with their respective upper limit controls (Upper Limit Control (ULC)s). For the SPE statistic, the ULC is determined using Equation 3.42, taking into account the significance level ($\alpha$), sample mean ($b$), and variance values ($v$):

$$ULC(SPE) = \frac{v}{2b} \chi^2 \left( \frac{2b^2}{v}, \alpha \right) \tag{3.42}$$

On the other hand, for Hotelling's $T^2$ statistic, the ULC is established through Equation 3.43, considering the $F$-distribution and a $100(1-\alpha)\%$ confidence level:

$$ULC(T^2) = \frac{R(N^2-1)}{N(N-R)} \times F_{R,(N-R),\alpha} \tag{3.43}$$

These ULCs serve as thresholds to identify anomalies in the data, providing a basis for effective anomaly detection.

## 3.4 Summary

In today's data-driven world, the ability to evaluate and derive insights from massive amounts of data is essential. While machine learning, data science, and artificial intelligence are frequent in the attention, the

significance of statistics should not be understated. This thesis has reviewed the core principles of both statistics and machine learning, emphasizing their shared foundation in learning from data. Statistics, as an exact science, enables the collection, analysis, and interpretation of data, offering valuable insights for decision-making. Meanwhile, machine learning uses previous experiences to make intelligent decisions, demonstrating adaptability across multiple domains. The evaluation of these models involves addressing challenges such as overfitting and underfitting, and statistical metrics play a crucial role in assessing their performance. The thesis emphasizes the complex understanding required for robust model evaluations in a variety of applications, including classification problems to regression analysis. Furthermore, it also emphasizes the importance of feature selection in improving model interpretability and efficiency.

Machine learning models, with their varied strengths and weaknesses, find applications across a multitude of domains. In healthcare, these models contribute to disease diagnosis and risk assessment, enhancing our understanding of conditions such as cancer, Alzheimer's, and cardiovascular diseases. Beyond healthcare, they play an important role in finance, where they aid in stock market prediction and risk assessment, as well as in cybersecurity, ensuring robust threat detection. Customer segmentation and recommendation systems improve user experiences in e-commerce and content consumption platforms. Image analysis, which is critical in domains such as remote sensing and medical imaging, takes advantage of these models' adaptability for pattern identification and classification. Machine learning models' versatility extends to sectors as diverse as marketing, economics, and political sciences, demonstrating their wide use and influence in dealing with real-world difficulties across businesses.

Finally, this thesis has addressed the vital field of anomaly detection, explaining its significance and use in a variety of disciplines. The exploration of three foundational techniques - classification, statistical methods, and clustering - provided a comprehensive understanding of their roles and challenges. The subsequent focus on the process control technique emphasized its value in monitoring and discovering anomalies quickly, despite its acknowledged drawbacks. PCA was introduced as a solution, bringing out a detailed method that addressed difficulties of dimensionality, collinearity, and data noise. The full explanation of statistical metric calculation provides a practical framework for effective anomaly detection. As a result, this thesis adds new insights to the field by emphasizing the ongoing evolution of approaches to provide effective anomaly identification in complex and dynamic datasets.

# 4

# Sleep Disorders Characterization of Portuguese Drivers.

The present chapter is about the sleep disorders of Portuguese drivers, where Section 4.1 presents the procedure developed to gather all the information needed for the study. The sample description (Section 4.2) is extensively described, with the number of answers obtained by regions of Portugal, as well as personal information and driving habits. Then, using the necessary procedures, a statistical analysis (Section 4.3) was carried out in order to extract more information about the sleep disorders of Portuguese drivers. Section 4.4 introduces an approach to visualizing the responses through the implementation of a dashboard. This tool serves as a dynamic platform to disseminate the study's findings effectively. By leveraging the dashboard's interactive features, any individual can gain deeper insights into the research outcomes, fostering a more comprehensive understanding of the results achieved. Finally, in Section 4.5 it is presented a brief summary of the present chapter.

## 4.1 Procedure

In order to evaluate the sleep disorders of the Portuguese drivers it was required to choose which questionnaires would be used to assess drowsiness during the day, the risk of developing obstructive sleep apnea, the quality of sleep, and circadian rhythm. The following questionnaires were selected based on the recommendations of a sleep physician and their availability in Portuguese language: ESS, STOP-Bang, Pittsburgh Sleep Quality Index (PSQI), and MESSi. The creation of the final questionnaire was thus the following stage. It was divided into five parts and conducted using the LimeSurvey platform. The first one concerns private data, including sex, age, height, weight, geographic distribution, occupation, and inquiries about one's driving habits as well as work schedule and weekly hours. The following were the driving questions.

4.1: Do you have a driver's license? Yes or no.

4.2: Do you usually drive? Yes or no.

4.3: What type of vehicle do you usually drive? Light passenger car, light goods vehicle, or other.

4.4: What type of course do you usually take? Long or short distance.

4.5: How many years do you have a driver's license? 0 to 5 years, or more than 5 years.

4.6: In the last year, have you ever driven with drowsiness that forced you to stop? Yes or no.

4.7: In the last year, have you ever fallen asleep at the wheel? Yes or no.

4.8: In the last year, have you ever come close to having a traffic accident because you fell asleep at the wheel? Yes or no.

4.9: In the last year, have you had any traffic accidents because you fell asleep at the wheel? Yes or no.

4.10: In the last year, have you had any car accidents for any other reason? Yes or no.

The following sections were concerned with the ESS, STOP-Bang, PSQI, and MESSi sleep disorders questionnaires.

The third phase involved identifying the target market and calculating the sample size to ensure that it was representative of the Portuguese population. As a result, it was decided to concentrate on the Portuguese drivers population, per region, since it would be challenging to obtain information on infants, kids, and teenagers. The NUTS II classification system divides Portugal's regions into the North, Center, Lisbon Metropolitan Area, Alentejo, Algarve, Autonomous Region of Azores, and Madeira. The representative sample of the Portuguese drivers population, by region, was determined using stratified sampling. This method of stratification makes use of probabilities, and after dividing the population into distinct strata—exclusive, homogeneous segments—a stratum is defined as a sampling for each stratum. In addition, the size of the stratum in the sample is determined from the fraction of the stratum in the population. The method used here is referred to as proportionate stratified sampling [48]. A confidence range for the population proportion was taken into consideration in order to determine the sample size, by region. This interval can be represented as Equation 4.1, and Equation 4.2 provides the estimation error. Understanding all of the factors is crucial, thus $\hat{p}_i$ is the estimated percentage, $z_{\left(1-\frac{\alpha}{2}\right)}$ is the selected z-value, also referred to as the crucial value for the standard Normal distribution, $\alpha$ is the threshold of significance, and $n_i$ is the sample size that will be used to identify the relationship [112].

$$\hat{p}_i \pm z_{\left(1-\frac{\alpha}{2}\right)} \times \sqrt{\frac{\hat{p}_i \times (1-\hat{p}_i)}{n_i}} \tag{4.1}$$

$$\text{Estimation Error} = z_{\left(1-\frac{\alpha}{2}\right)} \times \sqrt{\frac{\hat{p}_i \times (1-\hat{p}_i)}{n_i}} \tag{4.2}$$

The sample size by region can be calculated using Equation 4.3 when the level of significance ($\alpha$) and the estimate error have been established.

$$n_i = \left( \frac{z_{\left(1-\frac{\alpha}{2}\right)} \times \sqrt{\hat{p}_i \times (1-\hat{p}_i)}}{\text{Estimation Error}} \right)^2 \tag{4.3}$$

In this study, it was decided to take into account both the estimation error and degree of significance as 5%. Therefore the final expression can be calculated as in Equation 4.4.

$$n_i = \left( \frac{z_{\left(1-\frac{0.05}{2}\right)} \times \sqrt{\hat{p}_i \times (1-\hat{p}_i)}}{0.05} \right)^2 = \left( \frac{1.96 \times \sqrt{\hat{p}_i \times (1-\hat{p}_i)}}{0.05} \right)^2 \tag{4.4}$$

The National Institute of Statistics of Portugal (known as INE) reported that there were around 8.6 million adults in Portugal, in 2021[1]. There is no information about how many drivers there are in Portugal and, for that reason, the proportions to be achieved, for each region, were computed based on the Portuguese adults. After determining the overall number of Portuguese drivers and in each region, the proportion was calculated by dividing the two. In light of this, the Northern, Lisbon Metropolitan Area, and the Center are the areas with the highest proportion of adults. The total number of adults ($N_P$), the corresponding proportion (P), and the minimal representative sample size ($N_S$) that must be attained are listed in Table 8.

Table 8: Adult Population by region and respective proportion.

| **Region** | $\mathbf{N}_P$ | **P** | $\mathbf{N}_S$ |
|---|---|---|---|
| North | 3 012 972 | 35% | 350 |
| Center | 1 898 999 | 22% | 264 |
| Lisbon Metropolitan Area | 2 323 345 | 27% | 303 |
| Alentejo | 592 869 | 7% | 99 |
| Algarve | 358 763 | 4% | 61 |
| Autonomous Region of Azores | 197 084 | 2% | 34 |
| Autonomous Region of Madeira | 212 533 | 2% | 37 |
| Total | 8 596 565 | 100 | 1 149 |

It was finally time to compile the responses. On December 3rd, 2021, the Portuguese adult population's sleep disorder questionnaire became available and it was closed a year later. Social media sites like Facebook and LinkedIn were taken into consideration for diffusion, along with email correspondence with businesses, academic institutions, and town halls in different regions. After one year of disclosure,

[1]Available in `https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_indicadores&indOcorrCod=0007307&contexto=bd&selTab=tab2`, accessed: 2021-10-26

two lines had to be removed after responses were analyzed since the participants consistently assigned the same number to each question. Then, it was necessary to implement data pre-processing to correct some of the values in the height, time to go to bed, and time to be awake features. Rather than introducing 23:00, the hours to go to bed or wake up had discrepancies. The STOP-Bang, ESS, and PSQI guidelines were taken into consideration for the classification of sleep disorders. However, the MESSi questionnaire only has three factors (distinctness, morning affect, and eveningness), and higher values correspond to more alterations during the day, as well as morning and evening preferences, respectively. Therefore, the circadian rhythm was set as the morning type if the morning component was bigger than the evening factor. It was categorized as an evening type if the evening factor outweighed the morning factor. If not, it was set to be a form of intermediate circadian rhythm.

## 4.2 Sample Description

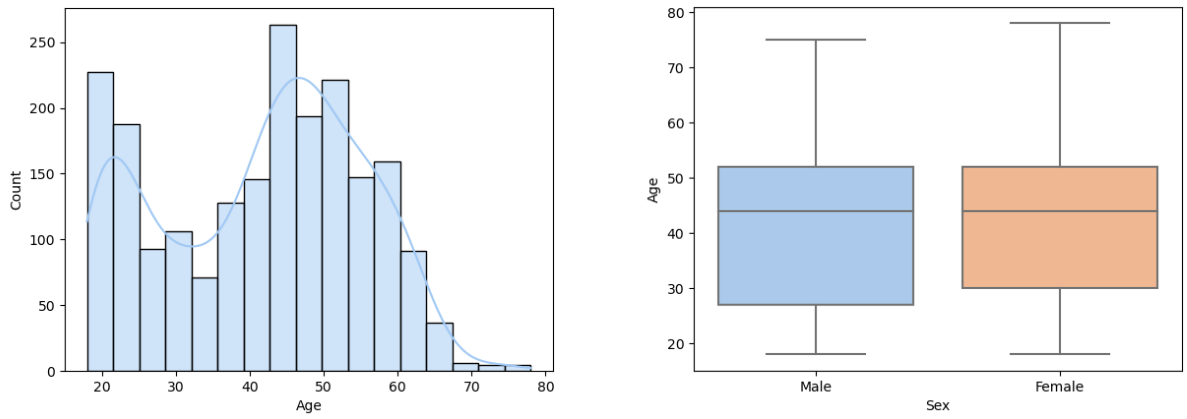
With 2 087 complete answers, response levels that were significantly higher than the initial minimum value were attained. The results are shown in Table 9, where it is clear that the North, the Lisbon Metropolitan Area, and the Center had more replies than the other regions. The Azores, Madeira, the Alentejo, and the Algarve were the next most popular regions. Note that the values achieved are about the Portuguese adults.

Table 9: Sample Stratification by Region.

| **Region** | $N$ |
|---|---|
| North | 468 |
| Center | 338 |
| Lisbon Metropolitan Area | 455 |
| Alentejo | 159 |
| Algarve | 97 |
| Autonomous Region of Azores | 297 |
| Autonomous Region of Madeira | 273 |

The results were as follows: 73.93% of the participants were female, 34.49% were of normal weight, 28.94% were overweight, 20.36% were at risk of being overweight, 13.90% were obese, and 2.31% were underweight. The median age of the participants is 44 years old, which indicates that 50% of them are under that age. The lowest and highest ages were 18 and 78, respectively. Figure 8(a) depicts the age distribution, which shows that the majority of participants are young (aged 18 to 29) and middle-aged (aged 40 to 60). Using the age by sex, Figure 8(b), it appears that the ages are similar between males and females, despite the fact that there are more females in the research.

In relation to driving habits, Table 10 shows the results of the driving questions. As a result, the majority of participants (91.57%) have a driver's license, generally drive (92.26%) light passenger cars (95.41%), short course routes (77.82%), and have more than 5 years of driving experience (88.60%). Furthermore,




(a) Age Distribution

(b) Age by Sex

Figure 8: Age of the Portuguese Adults.

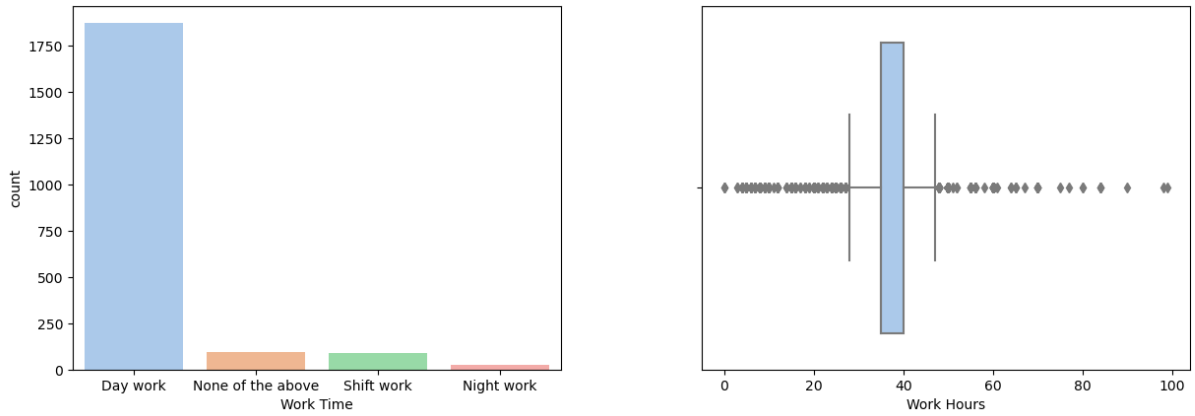
12.37% of participants were forced to pull over owing to drowsiness, up from 2.21% the previous year. Of those who fell asleep, 74.36% were on the verge of a traffic collision, and 20.69% were actually involved in one. In contrast, only 6.07% of people with a driver's license and who usually drive have been in an accident for any cause other than tiredness behind the wheel.

Table 10: Answers about driving.

| Question | Options | Values | Question | Options | Values |
|---|---|---|---|---|---|
| 4.1 | Yes<br>No | 91.57%<br>8.43% | 4.6 | Yes<br>No | 12.37%<br>87.63% |
| 4.2 | Yes<br>No | 92.26%<br>7.74% | 4.7 | Yes<br>No | 2.21%<br>97.79% |
| 4.3 | Light passenger car<br>Light goods vehicle<br>Other | 95.41%<br>2.95%<br>1.64% | 4.8 | Yes<br>No | 74.36%<br>25.64% |
| 4.4 | Short Course<br>Long Course | 77.82%<br>22.18% | 4.9 | Yes<br>No | 20.69%<br>79.31% |
| 4.5 | 0 to 5 years<br>More than 5 years | 11.40%<br>88.60% | 4.10 | Yes<br>No | 6.07%<br>93.93% |

Another piece of data gathered was the work schedule and its duration, as shown in Figure 9. The majority of participants (89.70%) work during the day, 4.36% work shifts, 1.29% work at night, and 4.65% do not fit into any of the previous alternatives. In terms of weekly labor hours, the majority of participants work between 35 and 40 hours, which represents the public and private sectors. Although three students defined the work hours as zero, others with less than 35 hours per week may be part-time workers. There

are also 25% of participants who work more than 40 hours every week.




(a) Work Time

(b) Work Hours

Figure 9: Work Information.

Concerning the sleep disorders of the drivers (Table 11), 22.14% have a high risk of developing sleep apnea, 38.78% have excessive daytime sleepiness, and 60.23% have bad sleep quality. The PSQI questionnaire revealed that the majority of individuals (50%) have 6 to 8 hours of sleep per night. Aside from that, 50% of them require 10 to 30 minutes to fall asleep. In terms of circadian rhythm, approximately, 60% is morning type, 34.75% evening type, and the remaining drivers are defined as intermediate. For the distinctness effect, the first quartile ($Q_1$) was equaled to 14, the second ($Q_2$) equaled to 16, and the third ($Q_3$) equaled to 19. Thus, 25% of the participants have a distinctness effect of less than 14, 50% have a distinctness effect of 14 to 17, and the remaining 25% have a distinctness effect of 19 or higher. Those who have higher values have considerable fluctuations in performance and mood throughout the day.

Table 11: Sleep Disorders Results.

| Questionnaire | Options | Values |
|---|---|---|
| STOP-Bang | Low Risk | 77.86% |
| | High Risk | 22.14% |
| ESS | Normal | 61.22% |
| | Excessive | 38.78% |
| PSQI | Bad Quality | 60.23% |
| | Good Quality | 39.77% |
| MESSi | Morning Type | 60.02% |
| | Evening Type | 34.75% |
| | Intermediate | 5.23% |

## 4.3 Statistical Analysis

In this study, it was achieved a higher value of excessive daytime sleepiness (38.78%) when compared to two Portuguese studies developed in 2014 [33] and 2015 [68], which obtained 20% and 9.11%, respectively. The study developed by [33], on the other hand, had a high risk of developing obstructive sleep apnea for 29% of the participants, whereas in [67] just 10% of the individuals had a high risk. In terms of sleep quality, 35.30% [67] and 54.8% [10] have poor sleep quality. However, in our investigation, the high risk of developing sleep apnea reached 22.14%, and 60.23% has bad sleep quality. It is also feasible to compare the sleep length and the time it takes to fall asleep. According to [10], 62.9% of adults sleep between 6 and 8 hours per night, and 44% require 1 to 15 minutes to fall asleep. According to our findings, 50% of adults sleep 6 to 8 hours, per day, and it takes 10 to 30 minutes to fall asleep. One interesting remark is that 49.45% of them fall asleep in less than 15 minutes. The majority of the results obtained are higher than those provided in other studies. It is also vital to recognize that the responses are subjective since they are based on how the person interprets the question. Nevertheless, it is crucial to note that the sample sizes differ and were retrieved in various ways, which can influence the outcome. This study has an important contribution to the Portuguese population, and also to the sleep disorders field since it considers different validated instruments to evaluate four sleep issues. It is noteworthy to mention that this study has a sample size larger than what is typically available, especially in Portugal.

Although, more critical information can be extracted from those results. According to [61] people are unaware of their sleep shortage and the consequences of sleep deprivation. Based on this information, the question that arises is: **The Portuguese drivers are aware of their sleep quality?** For this question, two factors must be considered: subjective sleep quality (Subjective Sleep Quality (SSQ)), as indicated by each participant in the PSQI questionnaire, and the PSQI final result. All participants were asked to rate their sleep quality on a four-point scale: very good, fairly good, fairly bad, and very bad. The alternatives “very good” and “good” were combined and classified as good sleep quality for this analysis. The same method was used for the remaining alternatives that were deemed to have poor sleep quality. As a result, the statistical hypotheses to be evaluated were defined 4.5, where $\pi_1$ represents the proportion of SSQ and $\pi_2$ represents the proportion of PSQI.

$$H_0 : \pi_1 = \pi_2 \text{ vs } H_1 : \pi_1 \neq \pi_2 \tag{4.5}$$

The McNemar [62] is the right method for answering the research question, which is utilized for two dependent categorical variables with two classes each ($2 \times 2$). The statistic test ($\chi^2$) resulted in a value of 408, and the p-value was less than 0.001. The proportion of SSQ and the PSQI varies when the criterion of significance is set at 5%. Following this conclusion, it is critical to determine whether the participants are aware of their sleep quality is good or terrible. Table 12 shows the contingency table between the SSQ and PSQI final results, revealing that 478 participants (25%) rated their sleep quality as good, although it is actually poor. As a result, some Portuguese drivers are unaware of their sleep quality. As a result, identifying techniques to manage this issue is critical so that people can find dependable answers to their

sleep problems.

Table 12: Contingency Table of Pittsburgh by SSQ for the 1st Research Question.

| | **Pittsburgh** | | |
|---|---|---|---|
| **SSQ** | Bad | Good | Total |
| Bad | 673 | 24 | 692 |
| Good | 478 | 736 | 1 214 |
| Total | 1 151 | 760 | 1 911 |

Another interesting remark, in the literature, is that excessive daytime sleepiness is a result of poor sleep quality [71]. Thus, **Is it true that people with excessive daytime sleepiness have bad sleep quality?** The ESS and PSQI final scores are the features to be investigated for this study question since they determine who has normal or excessive daytime sleepiness and good or bad sleep quality, respectively. To begin, all participants with excessive daytime sleepiness must be chosen, and the proportions of bad sleep quality must be calculated. One-sample proportion test was implemented to determine whether the proportion of participants with bad sleep quality is significantly greater than 0.5 among those with excessive daytime sleepiness. The null hypothesis states that the true proportion of participants with bad sleep quality is 0.5. The alternative hypothesis states that the true proportion is greater than 0.5. Thus, the test statistic $(z)$ is equal to 15.64, and the p-value was less than 0.001. This provides strong statistical evidence that the proportion of participants with bad sleep quality is significantly greater than 0.5 among those with excessive daytime sleepiness. In terms of proportional values, 70.31% of Portuguese drivers with excessive daytime sleepiness had poor sleep quality, while the remaining participants (29.69%) have good sleep quality. As a result, it was proven, that persons who are very sleepy throughout the day have, in fact, poor sleep quality. This is consistent with the literature.

According to the following study [175], sleep disturbances can also emerge when work schedules are not in sync with the circadian rhythm. As a result, the following question emerges: **The circadian rhythm is adjusted with the work schedule? If not, does it influence their sleep quality and daytime sleepiness?** The MESSi final result (morning, evening, or intermediate) and the work time (day, night, shift work, or none of the above) were taken into account for this research topic. Its purpose is to determine whether the circadian rhythm and work time aspects are independent of one another. The $\chi^2$ test is suited for this investigation since the null hypothesis is about the independence of the attributes under consideration. Therefore, the results achieved were $\chi^2(6) = 16.23$, and $p - value = 0.013$. As a result, there is an association between work time and circadian rhythm. Using the graphic representation (Figure 10), it is evident that persons who are classified as morning types (green color) are more likely to work during the day, while those classified as evening types (blue color) are more likely to work at night. The proportions for shift work are balanced between morning and evening types. As an outcome, the work schedule is tailored to the circadian rhythm. However, noteworthy observations emerge regarding individuals categorized as evening and morning types. Among those classified as evening types and

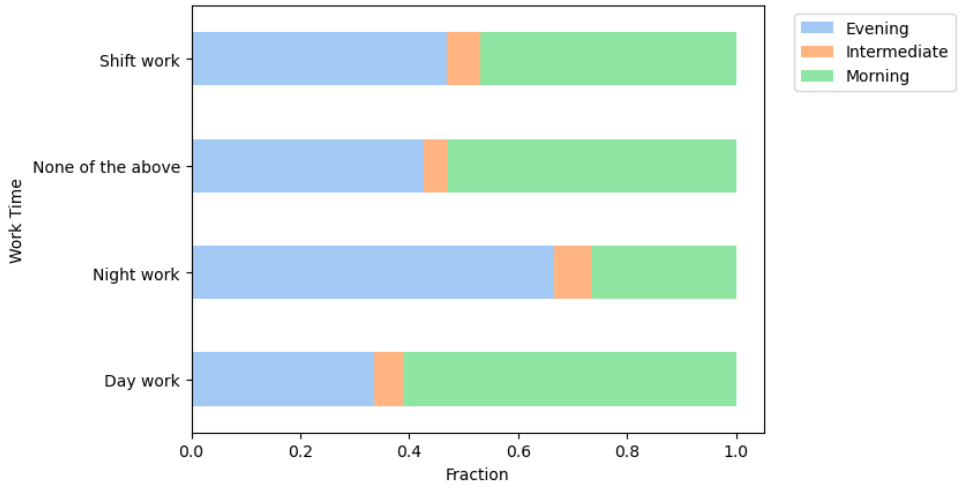



Figure 10: Circadian Rhythm by Work Time.

engaged in daytime work or shift schedules, 44.88% and 53.85%, respectively, exhibited symptoms of excessive daytime sleepiness, as indicated by survey responses. Conversely, for morning types working during nighttime hours or shifts, the prevalence of excessive daytime sleepiness was reported at 25% and 38.46% respectively. In a global context, 45.17% of participants categorized as evening types exhibit excessive daytime sleepiness, while only 34.47% of those classified as morning types display such sleep disorders. Another noteworthy observation regarding participants whose work schedules do not align with their circadian rhythms is that among those who habitually drive and had to stop due to drowsiness, 67.39% have excessive daytime sleepiness. Furthermore, among individuals who reported falling asleep while driving, 73.33% also have symptoms of excessive daytime sleepiness.

There is also an important investigation regarding excessive daytime sleepiness as a main symptom of a high risk of developing sleep apnea [3]. Therefore, it is intended to answer the following research question: **Is it true that people with a risk of developing sleep apnea also have excessive daytime sleepiness?** The major purpose of the analysis was to compare the final results of STOP-Bang and ESS in order to address the current topic. Aside from that, it was chosen to assess the odds ratio between the two characteristics to ascertain the strength of the link. The odds ratio obtained was 1.79, indicating that the potential of detecting a high risk of sleep apnea with excessive daytime sleepiness is, approximately, 1.8 times higher when compared to the group with normal daytime sleepiness. The 95% confidence interval was $]1.45; 2.24[$ and since the unit does not belong to the interval, there is a significant connection between the risk of developing sleep apnea and excessive daytime sleepiness.

## 4.4 Dashboard

The characterization of sleep disorders among Portuguese drivers garnered a significantly higher response rate than anticipated. Consequently, it is crucial to present the results in a concise manner to facilitate a comprehensive understanding of the research outcomes. Simplifying the presentation of findings will

enable anyone to delve deeper into the insights reached from the study.

Thus, data visualization is an approach for decision-making processes offering a powerful means to present the results achieved, facilitating faster and more informed decision-making. Besides that, visual tools, like dashboards, offer an easy-to-use way to explore and grasp data, making it understandable for people with different levels of expertise. Thereby, data visualization helps researchers improve their ability to analyze data and communicate their findings more effectively [69]. Moving forward, it is essential to explore the integration of data visualization within the framework of business intelligence in order to elucidate its significance in facilitating decision-making processes. A simple definition, for business intelligence, can be described as ***"[an] automatic system [that] is being developed to disseminate information to the various sections of any industrial, scientific or government organization"*** [69]. Basically, business intelligence system must provide timely information to the appropriate individuals in a suitable format and there are various tools and platforms to facilitate it. An illustrative instance in this context is Power BI, a data visualization and user-friendly interface solution engineered by Microsoft. This software facilitates the conversion of unprocessed data into visually engaging and interactive insights, thereby equipping users with the confidence to base decisions on data-driven analyses [59].

A dashboard (Figure 11) was created using the questionnaire data to share information about sleep problems with the public. It includes details like the number of responses from different regions (NUTS III and districts), sex distribution, driving license and experience, age categorized by sex, as well as results on daytime sleepiness, circadian rhythm, sleep quality, and the risk of developing obstructive sleep apnea.

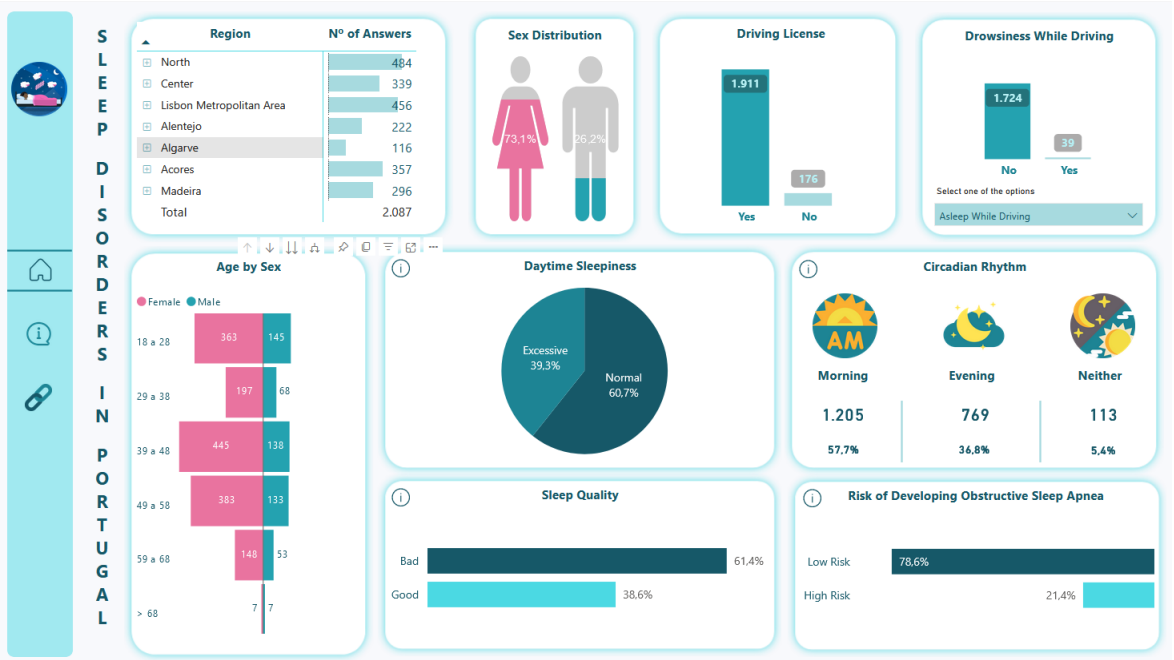
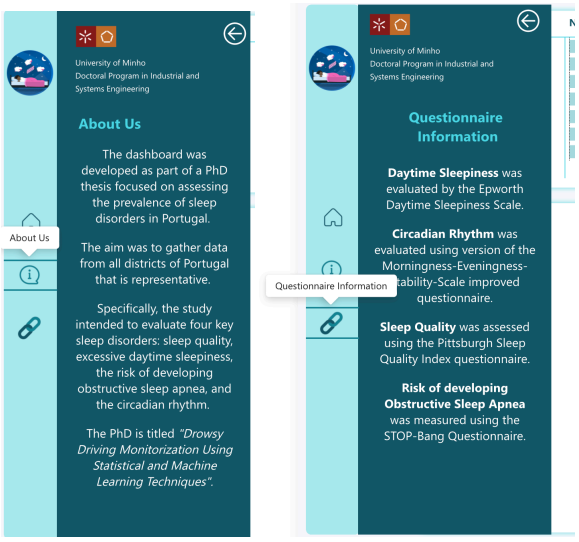



Figure 11: Dashboard for Sleep Disorders in Portugal.

A straightforward overview of the dashboard will be provided, particularly beneficial for individuals unfamiliar with Power BI. This description aims to facilitate comprehension of the visuals' functionality, aiding

in the extraction of valuable insights from the data. Therefore, the sidebar includes the dashboard's home, details about the study (under "About Us", Figure 12(a)), detailing the rationale behind dashboard development, the primary objective of the study, and the thesis title. Additionally, there is information regarding the names of the questionnaires employed for assessing sleep disorders (located under "Questionnaire Information", Figure 12(b)).




(a) About Us (b) Questionnaire Information

Figure 12: Side Bar Information.

One of the benefits of utilizing dashboards is the ability to create interactive visualizations. Simply put, when you select or click on one visualization, the others can also change based on what was selected. This feature enhances user engagement and allows for deeper exploration of the data, enabling users to gain insights from multiple perspectives with ease. Let's illustrate this with an example from the current report ( Figure 13): if the North region is selected or clicked, the other visuals will display only the values pertaining to participants from the North region. In this example, it's evident that the values for sex distribution, driving license status, and driving experience have changed. The same applies to the other visualizations as well. This interactive feature enables users to focus specifically on data relevant to their selected criteria, enhancing the precision and relevance of their analysis.

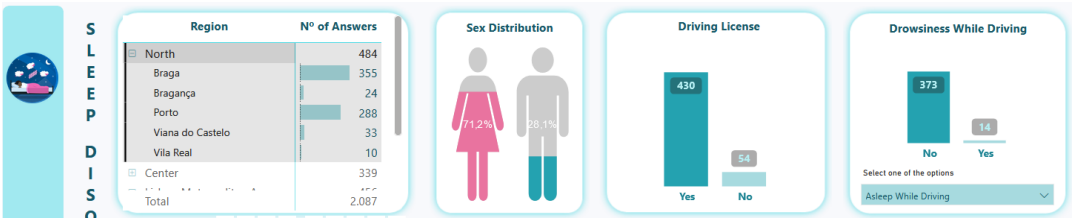



Figure 13: Interactive reports.

Additionally, dashboards offer further benefits, such as the utilization of tooltips and slicer visualizations. Tooltips provide supplementary details when hovering over data elements, enhancing user interaction and comprehension. Slicer visualizations enable users to filter data dynamically, allowing for customized views and deeper analysis [59]. Thereby, Figure 14(a) and Figure 14(b) display two distinct tooltips. The first tooltip pertains to participants with a driving license, providing how many of them habitually drives. The second tooltip presents participants' opinions on sleep quality, allowing for a comparison between questionnaire result and individual perceptions. In the context of slicers (Figure 14(c)), within the driving experience information, users can select from two options: "asleep during driving" or "stopped driving due to drowsiness."


(a) Habitually Drives (b) Sleep Quality (c) Driving Experience

Figure 14: Tooltips and Slicer Visualizations.

In conclusion, the dashboard report provides a comprehensive overview of sleep disorders among the Portuguese population, offering valuable insights about the sleep quality, daytime sleepiness, the risk of developing obstructive sleep apnea and the circadian rhythm. Through interactive visualizations, users can explore various facets of sleep-related issues, including regional differences, gender demographics, driving habits, perceptions of sleep quality, among others. The inclusion of tooltips and slicer visualizations enhances the user experience, allowing for deeper insights and customized analysis. Overall, the dashboard serves as a valuable tool for understanding the prevalence and impact of sleep disorders on driving behavior in Portugal, with implications for road safety and public health initiatives.

## 4.5 Summary

Sleep problems are a serious concern since they can have an impact on an individual's health, mood, and quality of life. As a result, multiple questionnaires are available to assess various sleep disorders, including daytime sleepiness, the risk of developing obstructive sleep apnea, sleep quality, and the type of circadian rhythm.

Portuguese drivers were examined for sleep-related issues, employing a stratified sampling approach to ensure representative data from every district of Portugal. After one year of disclosure, 1 911 complete responses were obtained for drivers, where the vast majority of participants have poor sleep quality

(60.23%), 22.14% have a high risk of developing sleep apnea, 38.78% have excessive daytime sleepiness, 60.02% are morning type, 34.75% are evening type, and the rest are intermediate. Comparing subjective sleep quality to that obtained from the questionnaire revealed that 25% of Portuguese drivers were unaware of their actual sleep quality. An intriguing finding emerged: 70.31% of drivers experiencing excessive daytime sleepiness also reported poor sleep quality. This indicates an association between excessive daytime sleepiness and poor sleep quality. The investigation also analyzed the alignment between the work schedule and the individual's circadian rhythm. It was established that the work schedule was appropriately adjusted with the majority of the individual's circadian rhythm. However, the analysis revealed significant differences in the prevalence of excessive daytime sleepiness between individuals categorized as evening and morning types whose work schedules do not match their circadian rhythm. Specifically, a higher proportion of evening types engaged in daytime work or shift schedules reported symptoms of excessive daytime sleepiness compared to morning types working during nighttime hours or shifts. These findings highlight the importance of considering individual circadian rhythms and work schedules in understanding and treating sleep disorders, particularly excessive daytime sleepiness. Finally, it was also demonstrated that the chance of developing sleep apnea is approximately 1.8 times higher in people who are very sleepy throughout the day. The results obtained are consistent with what is found in the literature regarding sleep disorders.

Therefore, this study makes several significant contributions to the existing literature. Firstly, it integrates four questionnaires to assess the risk of developing major sleep disorders such as sleep apnea, excessive daytime sleepiness, sleep quality, and circadian rhythm disruptions. The methodology is transparently outlined, including the determination of the required number of responses to ensure a representative sample of Portuguese drivers across all districts. Besides that, what sets this study apart is the substantial number of responses gathered, which fills a gap in the literature pertaining to Portuguese drivers and provides valuable insights into their sleep disorders. Another significant contribution was the development of a dashboard that not only disseminates the achieved results but also enhances the understanding of the gathered data by presenting the most important and valuable information to the Portuguese society.

# 5

# Drowsy Experiment

In this chapter, it was first described (Section 5.1) the driving simulation procedure with the identification of the questionnaires filled out, and all the devices used. Thereafter, Section 5.2 presents the analysis procedure with the description of steps developed to answer all the research questions related to drowsiness detection and prediction. Lastly, Section 5.3 addresses the description of the participants and contains the analysis of personal information, experience during the driving simulation and drowsiness stage.

## 5.1 Driving Simulation Procedure

Driving simulations were conducted to gather physiological data for detecting drowsiness. Prior to participation, all individuals were asked to sign a consent form detailing the study's purpose, the equipment involved, permission to collect biometric data, and assurance of anonymity (Appendix A details all the information). Four surveys regarding sleep disorders were filled out, assessing sleep quality (PSQI), excessive daytime sleepiness (ESS), circadian rhythm (ME), and risk of obstructive sleep apnea development (Stop-Bang) [45, 84, 31, 75]. The aim of collecting this data was to comprehensively understand each participant's sleep pattern, encompassing various aspects beyond just quality.

In the simulated driving scenario, the steering wheel, acceleration, and brake pedals of the Logitech® G27 driving system were employed. The choice fell on the commercial American Truck Simulator to replicate truck driving conditions due to its offering of extended, monotonous highway routes, encompassing varying speeds and day-night cycles, with speed limitations typically capped at 90 km/h. Participants were allotted a 10-minute adaptation period to familiarize themselves with the controls, preceding a 1-hour examination session. Prior to the session, participants were outfitted with the Empatica E4 wearable device to collect physiological data. The session followed a consistent route layout for all participants, involving initial and final city driving segments, with the majority of the simulation set on the highway. Instructions were provided to adhere to speed limits and traffic regulations throughout. Additionally, participants were exposed to monotonous music, such as white noise, through headphones during the examination to induce drowsiness.

Subsequently, the participants were required to complete a new questionnaire to collect personal information (gender, age, body mass index, sports practice), substances ingested (coffee, medicine, alcohol, cigarette), the presence of stress in the previous 24 hours, participant experience (symptoms, difficulty controlling the vehicle and recognizing obstacles, accidents), and the classification of the drowsiness level felt during the simulation though the KSS questionnaire [5].

Figure 15 summarizes the approach for the driving simulations, distinguishing what was developed in the sleep disorders questionnaires, driving simulation, and the final report.

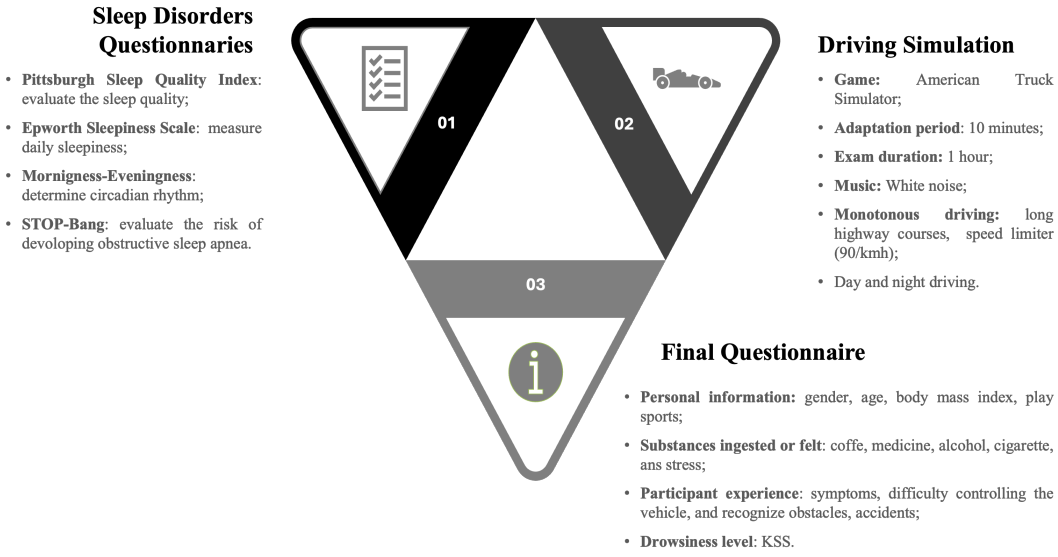



Figure 15: Procedure of the driving simulation.

During the simulation, a facial recognition system, as outlined in [34], was implemented to automatically capture blink duration. Despite efforts to collect blink duration data from all participants, it became evident that the results were largely unreliable. The algorithm developed proved sensitive to facial and bodily movements, as well as external factors such as wearing glasses and variations in lighting conditions, resulting in inaccuracies in blink detection timing. As a result, it was decided not to use such information. A potential solution to this challenge could involve adopting a Smart Eye Pro [55] camera system, renowned for its advanced remote eye-tracking capabilities. This device offers a comprehensive range of metrics including head orientation, eye position, gaze direction, pupil diameter, fixations, blinks, and eyelid movements. Notably, it operates effectively under various lighting conditions and accommodates the presence of glasses or non-infrared sunglasses.

## 5.2 Analysis Procedure

During the driving simulation, various data points were extracted for analysis to address the research questions and achieve the study's objectives. Specifically, from the recorded videos, the drowsiness level was determined utilizing Wierwille and Ellsworth's drowsiness scale [159]. To construct the drowsiness

stage dataset, the software `Shotcut` was utilized due to its functionality allowing for the visualization of recorded videos, event definition at specific timestamps, and exportation to a text file (.txt). Events were established every two minutes to facilitate classification based on participant behavior. Consequently, the dataset comprises timestamps corresponding to each event alongside the respective drowsiness stage (ranging from 1 to 5).

Moreover, for the physiological information, it was necessary to perform a more complex analysis. Empatica E4 device has a PPG sensor that measures the blood volume pulse that was used to derive the heart rate variability. It also contains a 3-axis accelerometer sensor to extract the motion-based activity [108]. For the R-R intervals, it was necessary to remove outliers in order to eliminate abnormal values, which were defined as the limits in Equation 5.1, where $\bar{x}_{R-RIntervals}$ is the sample mean and $s^2_{\text{R-R Intervals}}$ is the sample variance. In addition, ectopic beats were also treated since they are caused by irregularities in cardiac conduction and might result in a beating interval that is too short or too long [42]. Then it was computed cubic interpolation to replace the values removed. The signals were pre-processed based on the work developed by [104] for the assessment of the heart rate variability.

$$\bar{x}_{\text{R-R Intervals}} + s^2_{\text{R-R Intervals}} < \text{R-R Intervals} < 350ms \tag{5.1}$$

All the pre-processing was developed using the library `hrv_analysis` [35] that takes into consideration the `remove_outliers`, `remove_ectopic_beats` and `interpolate_nan_values` functions. These functions were implemented to remove the outliers (limits defined in Equation 5.1) and ectopic beats values and implement the interpolation defining the method as cubic. After cleaning the signal, the function `get_time_domain_features` was employed to identify the heart rate variability time domain features, `get_frequency_domain_features` the frequency domain features, `get_csi_cvi_features` and `get_sampen` the non-linear domain features, every two minutes, respectively. In terms of the accelerometer, there is information on the X, Y, and Z values every second and, in order to resample at two minutes, it was needed to identify the coefficient of variation (Coefficient of Variation (CV)), Equation 5.2, to decide which metric to use for this aggregation, where $s$ denotes the standard deviation and $\bar{x}$ signifies the mean of a given sample.

$$CV = \frac{s}{\bar{x}} \times 100 \tag{5.2}$$

This dispersion metric is crucial for understanding the variance of percentage in the mean value, particularly when extreme values are present. Therefore, if the coefficient was greater than 33, the median value was used. Otherwise, it was considered the mean value. This calculation was developed for each axis, individually. Besides that, this analysis started with the importation of the CSV file, from the Empatica E4, applying the `pandas` library. The mean and median values were extracted using the `df.mean()` and `df.median()` functions.

With all the information clean and in agreement, the implementation of the multivariate process control analysis was the next step. Firstly, it was needed to normalize the data, in order to achieve better results, applying the `df.mean()`, and `df.sd()` functions of `pandas` library. Then, the PCA was performed

using the `pca` library and function (has the same name), where the number of components was specified as one. In the `fit_transform` function was defined the features to use for the determination of the principal component. From this function, it was possible to identify the loadings and score values. This approach was replicated for time, frequency, non-linear heart rate variability features, and accelerometer. Finally, anomaly detection was developed through the identification of the Hotelling $T^2$ and SPE statistics and the respective upper limits controls. Different functions were manually developed to identify the statistics and the respective limit control (the code is presented in the Appendix B). Therefore, all of these steps (Figure 16) were developed to identify drowsiness transitions and signal noise, through the accelerometer information.

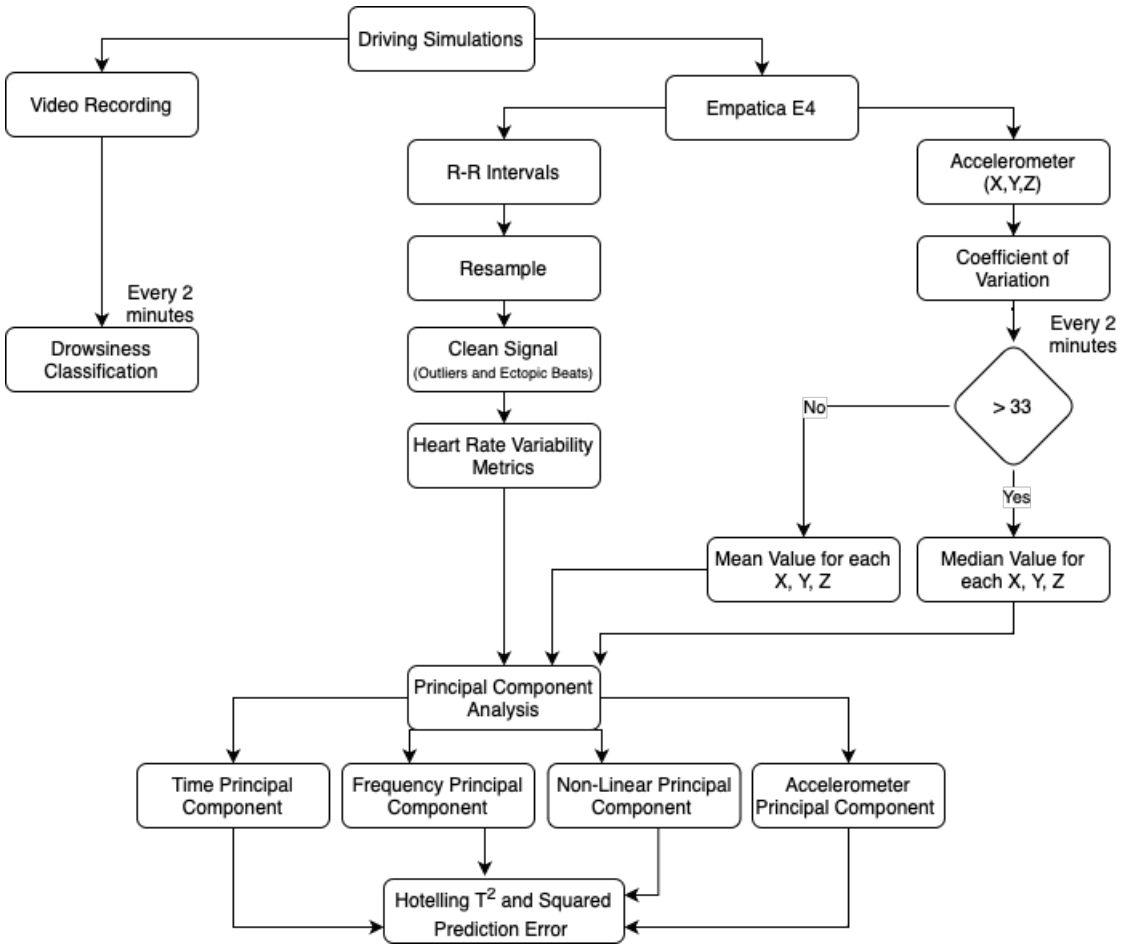



Figure 16: Drowsy detection analysis flowchart.

At present, the dataset has been prepared for the implementation of machine learning algorithms and forecasting methods. Feature selection optimization was carried out to determine the optimal subset of features to be considered, employing eight different models. Subsequently, the data was partitioned into training and testing sets using a stratified approach, allocating 30% of the data for testing purposes. Moreover, to maintain the proportion of awake and drowsy data, this partitioning was performed with consideration of class balance. The division process was facilitated using the `train_test_split` function from

the `sklearn.model_selection` library. On the other hand, multi-objective optimization was implemented utilizing the `pymoo` function, which employs a genetic algorithm. This application aims to identify the optimal features for drowsiness classification. Two objective functions were defined to minimize both Type I and Type II Errors, leveraging the `confusion_matrix` function from the `sklearn.metrics` library. The optimization process was configured with a maximum of 1500 generations and a population size of 100. Binary random sampling, two-point crossover, and bit flip mutation techniques were employed. The machine learning models considered for this optimization included logistic regression, ET, XGB classifier, AdaBoost, kNN, NB, SVM, and LDA. The XGB classifier model from the `xgboost` library was utilized, while the remaining models were implemented using the `sklearn` library. Additionally, a grid search was conducted for all models to determine the optimal parameters for drowsiness classification based on balanced accuracy. The Pareto front results were displayed for all optimizations. Subsequently, the evaluation focused on identifying the best subsets of features, labeled as $F_1$ and $F_2$, along with the models that yielded superior results for further analysis. This selection process involved considering the quartiles of Type I and Type II errors, along with calculated weights to identify only two subsets. Following this, evaluation metrics including accuracy, precision, recall, and F1 score were computed for the best two subsets as well as for the entire feature set. Additionally, a comparative analysis was performed utilizing AUC values.

Finally, drowsiness forecasting was implemented using the XGB regression model. This process was carried out individually for each participant, focusing on predicting each feature for the last two minutes of the simulation. The `xgboost` Python package was utilized, invoking the `XGBRegressor` function to learn and predict each feature for the specified time frame. Parameters for the XGB regression model were configured as follows: a base score of 0.5, a booster using a tree-based model (gbtree), 1000 estimators, early stopping set to 50, the objective set to regression with squared loss, a maximum tree depth of 3, and a learning rate of 0.01. To assess the quality of the forecasting, the predicted values for all features were incorporated into the classification models. Type I and Type II errors, accuracy, precision, recall, and F1 score were re-evaluated. Figure 17 illustrates a flowchart detailing all the steps involved in the drowsiness prediction process.

## 5.3 Participants Description

Professors and research fellows from the Polytechnic Institutes of Cávado and Ave were invited to participate in driving simulations for an event. Fifty-seven individuals agreed to join the study, primarily comprising men with a normal body mass index. The distribution of participants by body mass index (Body Mass Index (BMI)) revealed that overweight, underweight, and Class II and Class I obesity categories followed. The age range of participants varied from 18 to 56 years, with a predominant representation of younger individuals. Additionally, approximately half of the participants (50.88%) engaged in sports activities, with running, gym workouts, tennis, swimming, and football being the most common.

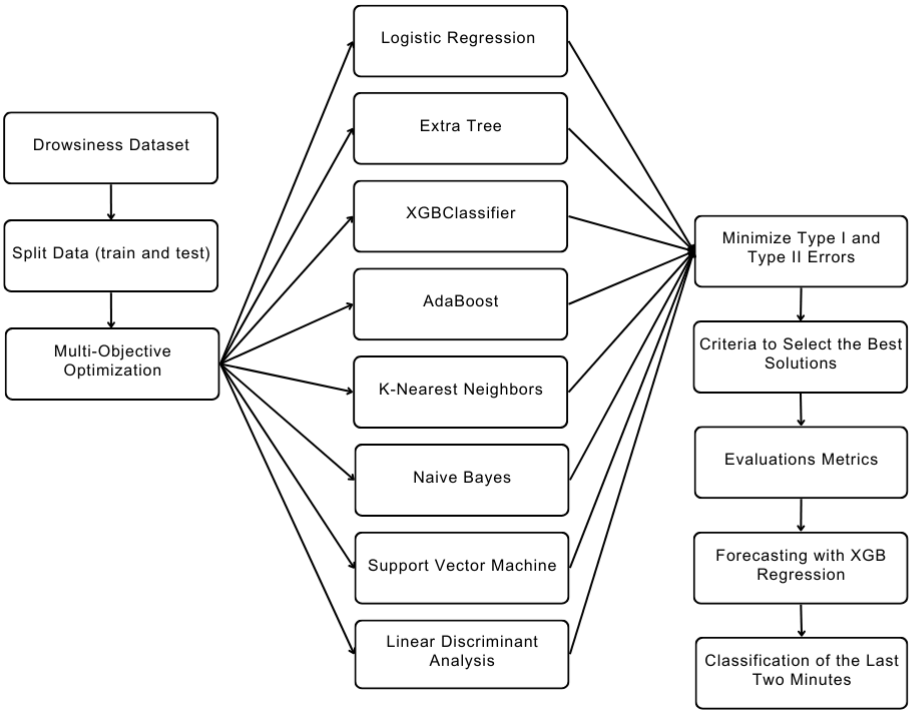



Figure 17: Drowsy prediction analysis flowchart.

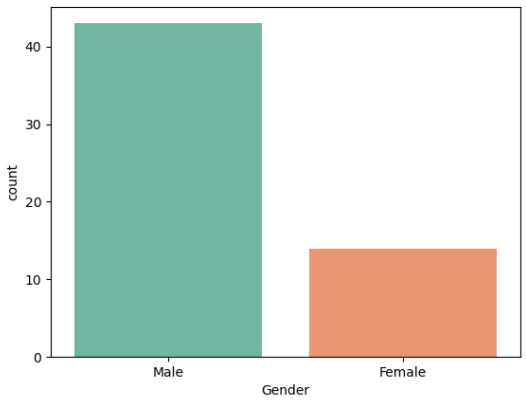
Only 14% of individuals reported having a health issue, with asthma (37.5%) and hypertension (25%) being the most prevalent, followed by type I diabetes (12.5%), myasthenia gravis (12.5%), and Gilbert syndrome (12.5%). A minority of participants were identified as having a high likelihood of developing obstructive sleep apnea (10.91%), experiencing excessive daytime sleepiness (40%), and reporting poor sleep quality (25.45%). Moreover, the majority of participants (70.91%) identified as neither morning nor evening types, followed by morning (16.37%), evening (10.91%), and definitively evening types (1.82%).

Additionally, participants were asked about their consumption of substances and experiences of stress in the 24 hours leading up to the simulation. According to the data presented in Table 13, the majority of participants reported consuming coffee but did not take medicine, alcohol, or smoke, and did not experience stress.

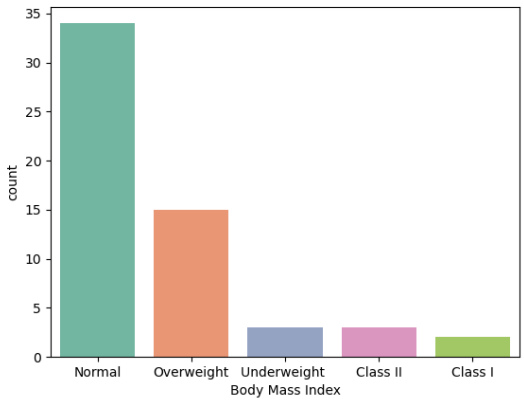
Table 13: Substances ingested or felt, in the 24 hours, before the simulation.

| | **Drank Coffee** | **Had Medicine** | **Drank Alcohol** | **Smoke Cigarette** | **Felt Stress** |
|---|---|---|---|---|---|
| No | 24 | 45 | 52 | 53 | 41 |
| Yes | 33 | 12 | 5 | 4 | 16 |

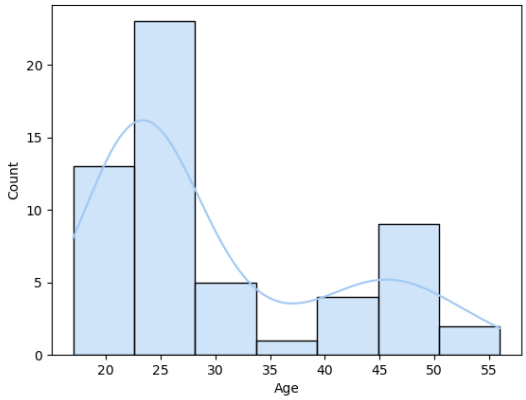
The participants were also asked to share their experiences during the simulation, specifically regarding symptoms they may have encountered such as sickness, vision problems, headache, weariness, itching eyes, concentration problems, and anxiety. They were instructed to rate the intensity of these



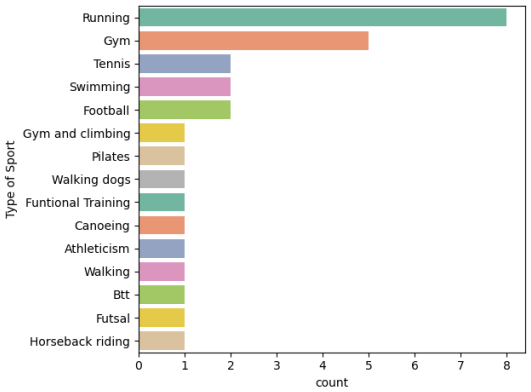

(a) Gender




(b) BMI




(c) Age Distribution




(d) Type of Sports

Figure 18: Personal information of the participants.

symptoms on a scale from 0 to 5, ranging from none at all to very much. The findings, detailed in Table 14, indicate that vision issues, weariness, itching eyes, and concentration problems were the most commonly reported complaints during the simulation.

Table 14: Symptoms during the simulation.

| **Symptom** | **0** | **1** | **2** | **3** | **4** | **5** |
|---|---|---|---|---|---|---|
| Sicken | 45 | 9 | 0 | 1 | 2 | 0 |
| Vision Problems | 21 | 19 | 11 | 3 | 3 | 0 |
| Headache | 33 | 10 | 7 | 7 | 0 | 0 |
| Fatigue | 15 | 7 | 13 | 12 | 7 | 3 |
| Itching Eyes | 22 | 12 | 8 | 6 | 6 | 3 |
| Concentrate Problems | 12 | 16 | 11 | 7 | 9 | 2 |
| Anxiety | 38 | 8 | 5 | 3 | 2 | 1 |

Another important aspect to consider is the participants' experience with the vehicle. They were asked

whether they encountered any issues operating the car, experienced accidents, or faced difficulties in identifying potential obstacles and responding promptly. Participants rated their experience on a scale of 0 to 5, with 5 indicating significant difficulty. While a majority of participants reported no challenges in controlling the vehicle, 40.37% experienced issues, and 56.14% were involved in accidents. Notably, 66.67% of participants had difficulty recognizing obstacles, with varying degrees of severity among individuals.

The final inquiries focused on participants' level of drowsiness, assessed using the KSS questionnaire. They were asked to indicate when they felt most drowsy (at the start, middle, or end of the simulation), whether they experienced increased drowsiness during nighttime, and to rate the difficulty of staying awake on a scale from 0 to 5 (0 representing none and 5 representing a lot). It's important to note that subsequent questions were only directed to individuals who exhibited evidence of drowsiness. The range of drowsiness levels considered spanned from "some signs of sleepiness" to "very sleepy, great effort to stay awake, fighting sleep". The results tabulated in Table 15 illustrate findings from the KSS, revealing that 36 subjects exhibited signs of tiredness ranging from 6 to 9. Consequently, it is evident that a majority of participants experienced drowsiness during the simulation.

Table 15: KSS results.

| Level | Description | Count |
|---|---|---|
| 1 | Extremely alert | 1 |
| 2 | Very alert | 1 |
| 3 | Alert | 5 |
| 4 | Rather alert | 9 |
| 5 | Neither alert nor sleepy | 5 |
| 6 | Some signs of sleepiness | 9 |
| 7 | Sleepy, but no effort to keep awake | 12 |
| 8 | Sleepy, some effort to keep awake | 11 |
| 9 | Very sleepy, great effort keeping awake, fighting sleep | 4 |

Regarding when participants experienced the most drowsiness, the majority (80.56%) reported feeling it during the middle of the simulation, with fewer experiencing it at the end (13.89%), and only a small percentage (5.55%) at the beginning. Additionally, it was observed that the primary indicators of drowsiness were most notable during nighttime (72.22%). Notably, none of the participants rated the difficulty of staying awake as a 5. Only 19.44% assigned a score of 4, while 22.22% attributed scores of 2 and 3 to the difficulty of staying awake. It's important to note that these responses are self-reported and may not always align with video evidence. For instance, a couple of individuals appeared to almost fall asleep during the simulation, indicating significant difficulty in keeping their eyes open and staying awake. Consequently, a higher score (such as 5) for the difficulty of staying awake would have been expected, rather than the observed score of 3.

Following the classification of drowsy stages through videotaping participants' faces, the percentages of each phase were computed using Wierwille and Ellsworth's drowsiness scale. This scale categorizes drowsiness into five stages, from not drowsy (S1) to excessively drowsy (S5). However, it was noted that out of the total participants, 12 individuals were not suitable for classification due to recording delays compared to real-time. To enhance clarity, the simulations were conducted for one hour, while the accompanying video lasted one and a half hours in total. Consequently, it was decided to exclude individuals affected by recording delays, totaling 12 participants, from further analysis. This decision resulted in a sample size of 45 for subsequent analyses. Notably, Figure 19 illustrates the overall percentage for each drowsiness level. It is evident that most participants did not reach level S5, which exhibits lower variability (the gap between the third and first quartiles, known as the interquartile range) compared to the other levels. Conversely, levels S1, S2, S3, and S4 display greater variability, with interquartile ranges of 23.23, 27.86, 19.92, and 36.33, respectively.

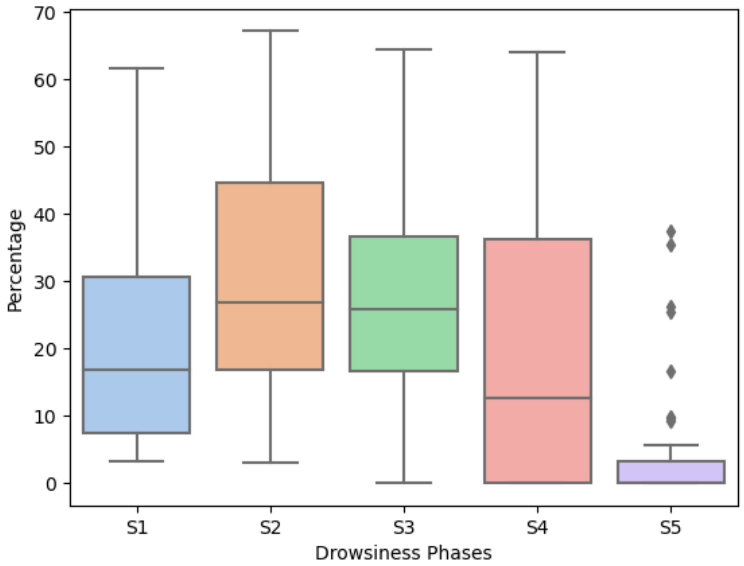



Figure 19: Percentage of each drowsiness stage.

Participant characterization and the implementation of the Multivariate Statistical Process Control method using PCA, led to the development of the following articles [11, 12].

# 6

# Modeling of Drowsy Driving

This chapter is dedicated to the modeling of drowsy driving, starting with drowsiness detection through multivariate statistical process control (refer to Section 6.1), aimed to enhance drowsiness classification and mitigate subjectivity in the assessment process. Following this, Section 6.2 undertakes an optimization of feature selection using machine learning techniques, intended to identify the optimal classification model and the subset of features for effectively distinguishing between awake and drowsy states. Moreover, Section 6.3 evaluates heart rate variability forecasting, which involves predicting drowsiness in advance and assessing the achieved results. Lastly, a summary of the present chapter is described in Section 6.4.

## 6.1 Drowsiness Detection

Detecting drowsiness at the wheel has been a challenge and, according to the study [1], multivariate statistical process control, along with PCA, can be implemented for drowsiness detection. In that study, both the time and frequency domains of the heart rate variability were taken into account, and Hotelling $T^2$ and SPE were computed based on just one principal component. Applying a similar methodology to the collected data, the time, frequency, and non-linear domains into one principal component. However, only the results of one participant will be reported, in this section, as a representation of the results achieved for all the participants. According to their assessment, the participant felt drowsy and struggled to remain alert. Therefore, the Hotelling $T^2$ and SPE statistics, along with their respective upper limit control were evaluated at a 95% confidence level. Figure 20 presents achieved results. For the SPE, two points are out of control, while only one point is out of control for the Hotelling statistic. However, three points are out-of-control using one principal component at the instances 9, 11, and 30. This means that instance 9 represents the 18th to 20th minutes of driving simulation, instance 11 the 22nd and 24th minutes, and so forth.

Following the identification process, the recorded video was analyzed to discern instances of being out-of-control and detect signs of drowsiness in participants. Notably, signals of drowsiness were predominantly observed at the final out-of-control point. Conversely, for the initial point, signs of drowsiness were

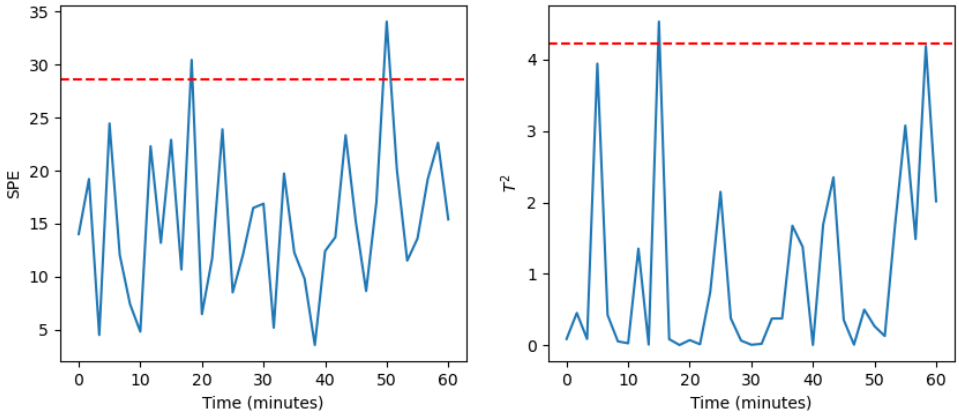



Figure 20: Hotelling $T^2$ and SPE using one principal component.

evident through slower blinks and head movements exhibited by the participant. Regarding the subsequent point, the participant exhibited even slower blinks, occasionally closing their eyes briefly. At the final point detected, the participant displayed unnecessary movements, a forward tilt of the head, yawning, and increased blinking to prevent complete closure of their eyes and avoid falling asleep. It was evident that the participant was exerting effort to remain awake. A similar methodology was applied to other participants, identifying points at the beginning and end of the driving simulation that could have been influenced by the placement of the wearable device. Additionally, signs of drowsiness, such as slower blinks, reduced eye movement, arm and head movements, yawning, and nearly closed eyelids, were also observed.

However, there are instances of drowsiness peaks that are currently not being detected, indicating a need for improvements to achieve more accurate results. One potential enhancement involves adjusting the number of principal components considered. Given that heart rate variability is segmented into three domains, a similar division was applied to the principal components analysis to improved the drowsiness detection. As a result, three principal components were considered, encompassing time, frequency, and non-linear features. The outcomes of this analysis are depicted in Figure 21. In the time domain, three out-of-control points were identified for both statistics, while the frequency and non-linear domains revealed one and three out-of-control points for Hotelling's $T^2$ and SPE statistics, respectively. Consequently, the heart rate variability at times 3, 9, 11, 13, 14, 20, 26, 30, 33, 34, 35, and 36 were flagged as out-of-control. It is essential to note that the points identified using one principal component also manifested when utilizing three principal components. Upon video inspection, the participant exhibited slower blinks, endeavored to keep their eyelids from closing, experienced brief moments of falling asleep, displayed head movements, increased blinking, yawning, unnecessary movements, and demonstrated a struggle to remain awake.

Therefore, it becomes evident that by employing the suggested methodology, which involves three principal components, a greater number of drowsiness indicators can be identified. In this specific instance, the identification increased from three points to twelve, with the implementation of three principal

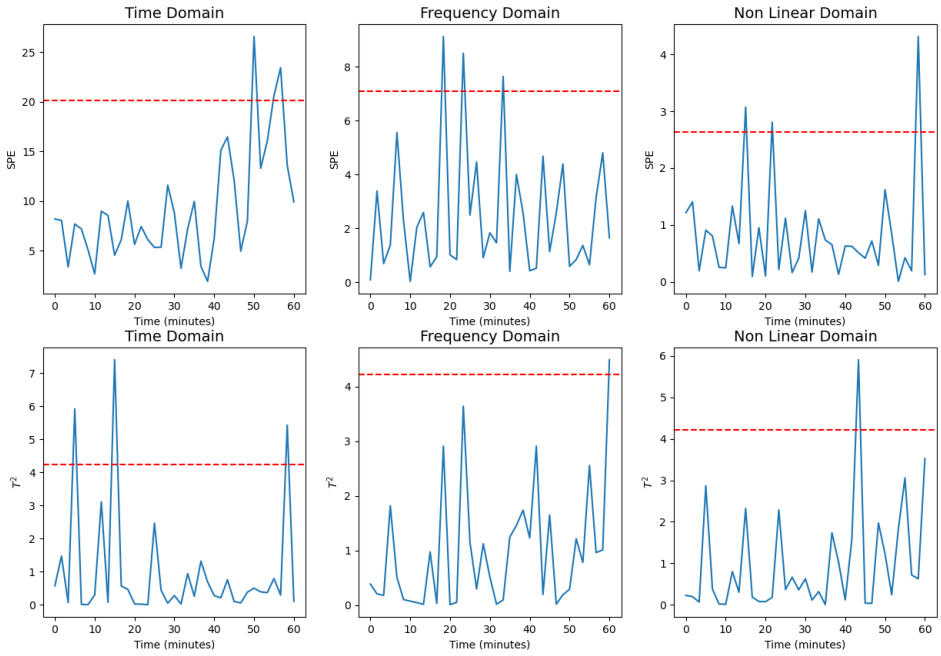



Figure 21: Out-of-Control points using the proposed methodology.

components encompassing the points identified using only one principal component. It is crucial to emphasize that this methodology was consistently applied to all participants, yielding the same conclusion: a more comprehensive detection of drowsiness signs.

Despite the promising outcomes achieved by the proposed methodology in detecting additional signs of drowsiness, it remains imperative to validate these results by comparing them with manually classified drowsiness. Furthermore, Figure 22 illustrates the drowsiness periods (blue line) representing the out-of-control points (value 1) identified using the proposed methodology alongside the drowsiness levels (orange line) determined using the KSS scale. This visualization indicates that certain points align with transitions in drowsiness levels, a trend observed consistently across other participants as well.

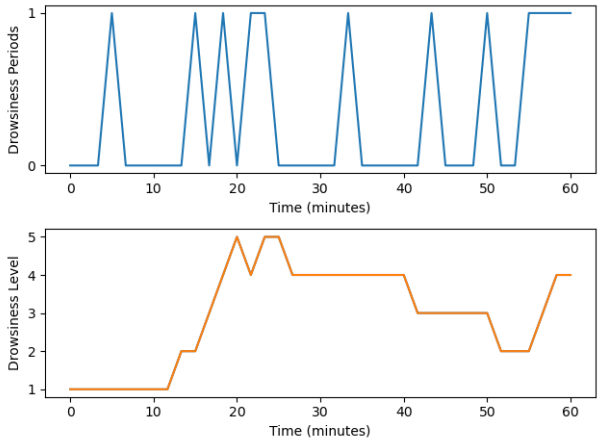
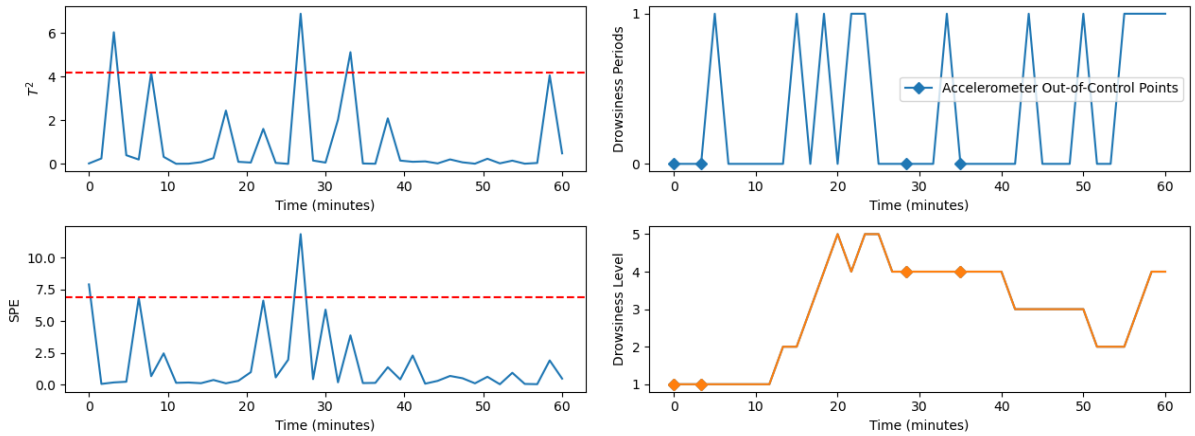



Figure 22: Identification of drowsiness transitions.

Nevertheless, a concern arises regarding whether the out-of-control points are influenced by signal

noise, possibly stemming from abrupt movements. At times, participants may become distracted or not fully alert, leading to rapid movements to regain control of the vehicle. It is crucial to ascertain whether such noise is affecting the out-of-control points within the heart rate variability domains, thereby potentially restricting the identification of genuine drowsiness transitions. For this analysis, a new principal component was computed utilizing the accelerometer data. The Empatica E4 accelerometer continuously records gravitational force along three axes (X, Y, and Z parameters) every second [134]. To ensure consistency in comparison between datasets, the collected accelerometer data must be organized in accordance with heart rate variability measures. Every two minutes, the CV was computed and if the value is less than 33, indicating consistency in the sample, the mean value is analyzed. However, if the CV exceeds 33, it is recommended to use the median value, which is less influenced by extreme values [23]. Following this recommendation, in our analysis, the mean value for X, Y, and Z values was computed every two minutes for consistent data. Conversely, in cases of inconsistency, the median value was utilized. The subsequent step involved implementing multivariate statistical process control with accelerometer data as a principal component. Figure 23(a) illustrates the outcomes obtained for both Hotelling's $T^2$ and SPE regarding the accelerometer data. Both statistics identified instances 0 and 1 as out-of-control points, with instance 2 detected solely by the Hotelling's $T^2$ statistic. Instance 0 corresponds to the initial two minutes following the replacement of the wearable device, potentially attributed to significant movements. When comparing these points with those identified using heart rate variability, as depicted in Figure 23(b), no common points are observed in this specific instance. This suggests that signal noise resulting from sudden movements may not influence the out-of-control points identified through heart rate variability.

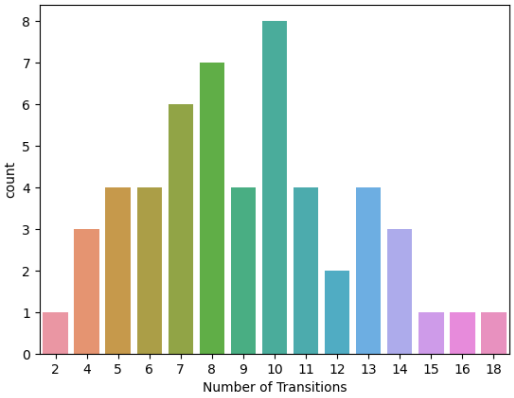



(a) Multivariate Statistical Process Control for Accelerometer. (b) Accelerometer and Heart Rate Variablity Out-f-Control Points

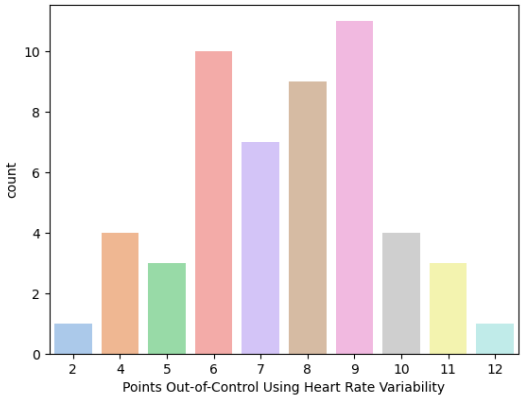
Figure 23: Accelerometer and heart rate variability analysis

The out-of-control points for both heart rate variability and accelerometer must be carefully scrutinized for all the participants. This analysis aimed to determine whether the out-of-control points from the heart rate variability were really unaffected by noise. To gain deeper insights, draw conclusions, and devise

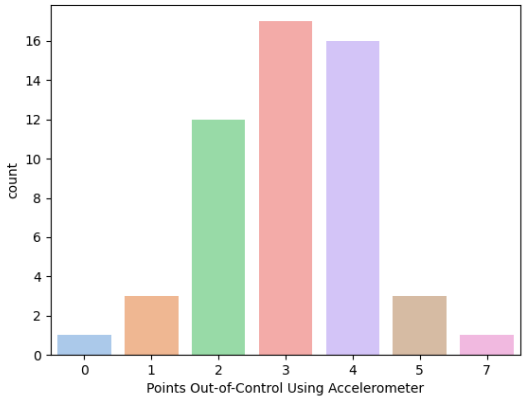
improvements, an evaluation was conducted on the number of transitions in drowsiness levels, the out-of-control points identified using both heart rate variability and accelerometer data, and the commonality of out-of-control points. This analysis is depicted in Figure 24. Considering the data from all participants, there were observed 2 to 18 transitions in drowsiness levels, with 7, 8, and 10 being the most frequent occurrences (see Figure 24(a)). However, the number of out-of-control points ranged from 2 to 12 for heart rate variability (refer to Figure 24(b)) and from 0 to 7 for the accelerometer (refer to Figure 24(c)). The highest counts were recorded between 6 to 9 for heart rate variability and 2 to 4 for accelerometer data. When comparing the out-of-control points identified through both heart rate variability and accelerometer data (see Figure 24(d)), it was found that 39.62% had one point in common, 11.32% had two points in common, and 7.55% had more than two points in common. Notably, 41.51% of cases did not share any out-of-control points.



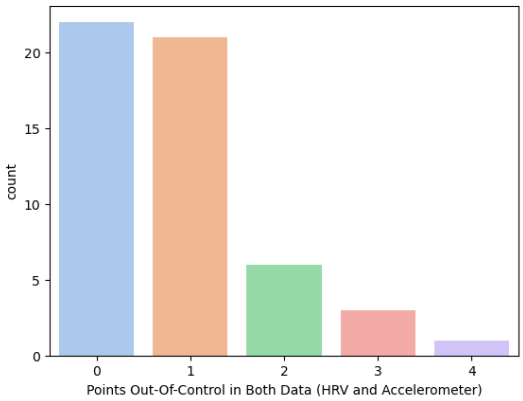

(a) Number of drowsiness transitions




(b) Heart rate variability out-of-control points




(c) Accelerometer out-of-control points




(d) Out-of-control points in both heart rate variability and accelerometer.

Figure 24: Analysis of transitions and out-of-control points.

Another crucial analysis involves evaluating transitions when the out-of-control point is attributed to

both heart rate variability and accelerometer data. According to the findings, 14.29% of the out-of-control points shared in common corresponded to a transition in drowsiness levels. Furthermore, it was observed that the majority of out-of-control points (66%) identified through the accelerometer occurred at the beginning of the simulation. Given that the replacement of the wearable device took place before the simulation, it is plausible that these points are associated with the placement of the device.

Following this analysis, it becomes apparent that adjustments are necessary. The classification of drowsiness is subjective, and it is conceivable that the classification may not always be entirely accurate. Therefore, there is a need to enhance the classification process by taking into account the out-of-control points identified through heart rate variability. As a result, the recorded video was utilized to determine whether it is feasible to anticipate or delay a drowsiness transition attributed to a misclassification. Following this analysis, it was necessary to adjust the drowsiness classification for most participants, as their observed behavioral signs indicated a different level of drowsiness. This adjustment enabled a substantial reduction in the subjectivity of the ratings. Additionally, the new classification for the participant under analysis is depicted in Figure 25, where the blue line represents the old classification, and the orange line reflects the improved classification derived from the proposed methodology. It is important to note that the first out-of-control point identified through heart rate variability occurred at the beginning of the simulation, and the recorded video commenced after this event. Consequently, it was not feasible to discern the reasons behind these occurrences.

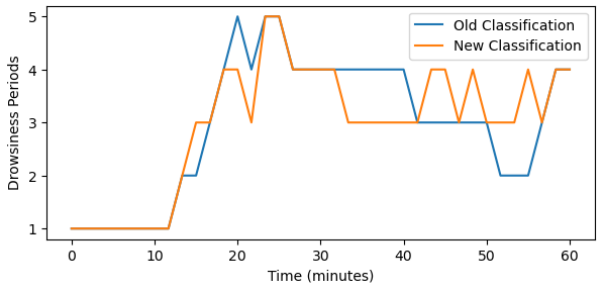



Figure 25: Example of a drowsiness classification improvement.

The precision and recall metrics were generated to assess the performance of the suggested technique. It is meant to test the out-of-control points' capacity to accurately reflect a drowsiness transition, which can be computed using accuracy. However, it is also meant to test the ability to recognize true drowsiness transitions, represented by the recall measure. Table 16 presents the results for both accuracy and recall comparing the methodology from the first drowsiness classification and the improved considering the multivariate statistical process control. The precision and recall values for the first drowsiness classification were 0.37 and 0.21, respectively. With the improved classification, precision was equivalent to 0.83, and recall was equal to 0.47. Using the improved classification, 83% of the out-of-control points indicate a drowsiness transition (precision value), while only 47% of the drowsiness transitions are accurately detected (recall value).

Table 16: Precision and recall metrics.

| Drowsiness Classification | Precision | Recall |
|---|---|---|
| First Classification [13] | 0.37 | 0.21 |
| Improved Classification | 0.83 | 0.47 |

It is evident that the proposed methodology has significantly enhanced the drowsiness classification and yielded improved outcomes, despite its limited ability to accurately identify true drowsiness transitions. However, it is essential to evaluate each out-of-control point to determine whether a drowsy transition has occurred. The proposed method effectively addresses several issues, such as the subjectivity of classification, failure to detect drowsiness peaks, and susceptibility to signal noise. Although this approach is not flawless and may not identify all drowsiness transitions, it is noteworthy that the majority of out-of-control points indeed signify a transition in drowsiness. This represents a significant achievement.

Hence, out of the 1672 records of heart rate variability along with their respective drowsiness classes, it is worth noting that approximately 68% of the records belong to participants classified as drowsy. As depicted in Figure 26, a comparison between the drowsy and awake states, considering both the original and improved data, is illustrated. It is evident that as the number of records classified as awake decreases, the number of drowsy records increases.

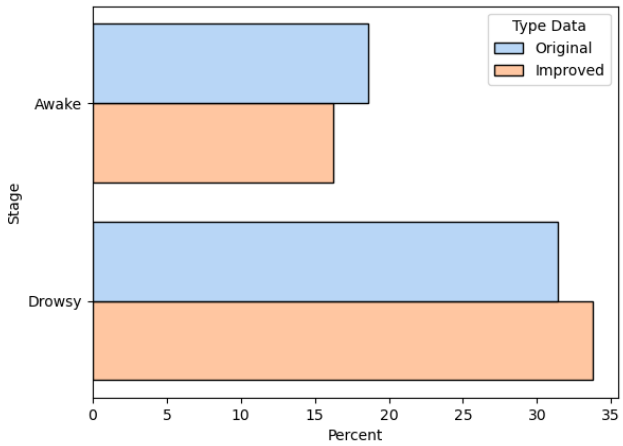



Figure 26: Original and improved data.

## 6.2 Feature Selection Optimization using Machine Learning

Highlighting the importance of identifying influential features in drowsiness detection, the thesis explores the critical process of optimizing feature selection, using the genetic algorithm and implementing the wrapper method. Besides that, this investigation extends to the implementation of diverse machine learning models, namely LR, ET, XGB, AdaBoost, kNN, NB, SVM, and LDA to classify the drowsiness stages. It also includes the integration of multi-objective optimization, strategically designed to minimize both Type I and

Type II errors. Therefore, Type I errors manifest when the model wrongly predicts drowsiness in an awake individual, while Type II errors occur when the model fails to identify drowsiness in an individual who is actually drowsy. The primary goal is to achieve a balance between preventing unnecessary alarms (Type I errors) and ensuring increased correct classification by reducing instances of persons being incorrectly recognized as awake (Type II errors).

Henceforth, Figure 27 showcases the non-dominated solutions reached for each model, amounting to 100, 19, 27, 100, 37, 96, 53, and 100 solutions for the distinct models, each with unique subsets of features. Primarily, the evaluation metrics exhibit identical values across models but with different sets of selected features. The following objective was to discern the most pertinent solutions for further investigation. To achieve this, quartile analysis was conducted for both Type I and Type II errors. The horizontal line represents the third quartile ($Q_3$) for Type II error, signifying that 75% of the values are equal to or less than 10.16%. Conversely, the vertical lines represent the first ($Q_1$) and third quartile ($Q_3$) for Type I Error. Here, 25% of the values are equal to or less than 18.33%, while 75% are equal to or less than 24.90%. By segregating the solutions based on this quartile division, the optimal solutions reside within the range delineated by the Type I error's first quartile and the third quartile for Type II Error. Consequently, there exist 6 solutions each for the ET and XGB models, and 2 solutions for the kNN, NB, and SVM models.

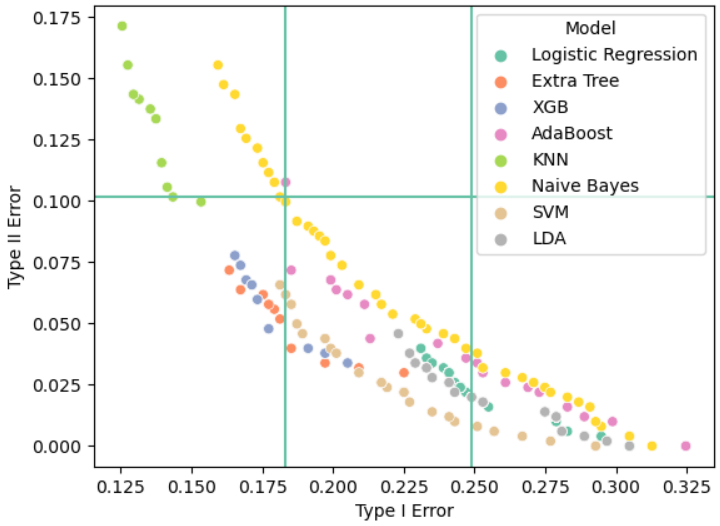



Figure 27: Pareto front: Type I Error against Type II error.

This analysis refines the focus, identifying specific solutions for in-depth consideration across various models. Subsequently, a depth analysis of the selected features was imperative to identify the optimal subset for drowsiness classification. In this attempt, Figure 28 presents a boxplot for the five best models that produce the best results, emphasizing the number of features used. Particularly, the NB model has the fewest number of features, whereas the kNN model has the most features to examine. Meanwhile, the XGB, SVM, and ET models remain within the previously described range of results, with minor differences.

It is worth noting an outlier in the ET model, indicative of a distinctive case within the dataset. This assessment of the number of features provides insights into the variance in feature selection methods among models, providing information on the varied degrees of complexity inherent in their decision-making processes.

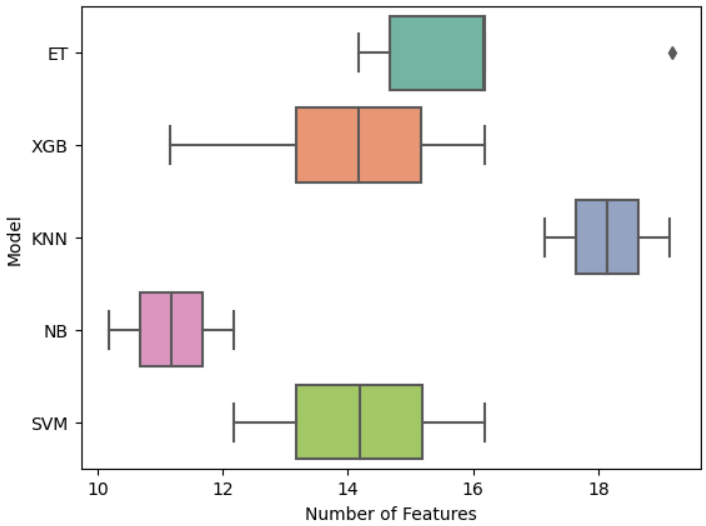



Figure 28: Number of features selectef for ET, XGB, KNN, NB, and SVM models.

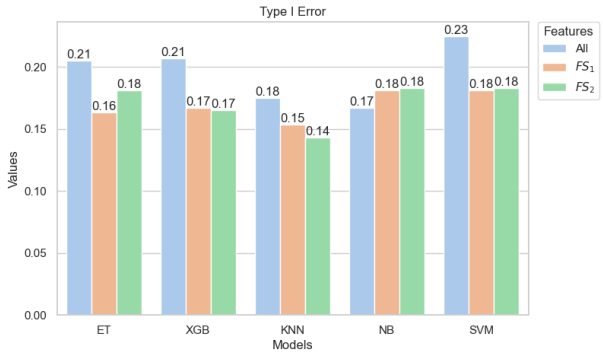
An essential analysis entails examining whether metrics improve with a decreased amount of features versus using the whole set. However, instead of considering all alternative solutions, a criterion was created to select only two distinct feature subsets for each model. Therefore, Equation 6.1 defines this criterion, which integrates values acquired for both Type I and Type II errors, as well as the count of selected features. The application of this criterion is specifically adapted to the ET and XGB models, given the number of solutions. The application of this criterion increases the selection process, enabling a focused examination of feature subsets that optimize error metrics. This strategic approach ensures that the selected feature subsets find a balance between improved performance and reduced complexity, which is critical in the search of efficient and interpretable drowsiness classification models.

$$\text{weight} = 0.35 \times (\text{Type I Error} + \text{Type II Error}) + 0.30 \times \text{Number of Features} \tag{6.1}$$

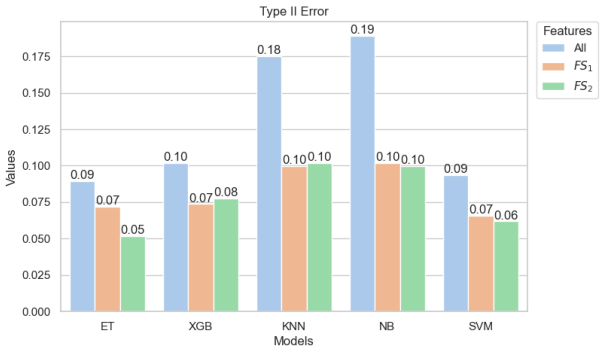
After evaluating the weight value, a selection process enabled to identify two distinctive feature subsets, denoted as $FS_1$ and $FS_2$, for each model. These subsets serve as the basis for computing essential evaluation metrics such as accuracy, precision, recall, and F1 score. It is imperative to note that the selected features exhibit variability across models, the designation of $FS_1$ features does not remain constant across different models. This highlights that the ideal feature set for each model is dependent on its unique characteristics and complexities. Moreover, for comprehensive analysis, evaluation metrics were also extracted considering all the features. This method provides for a comprehensive comparison, comparing the performance of models based on the selected feature subsets ($FS_1$ and $FS_2$) against those

based on all available features. Such comparison analysis allows for discovering the impact of feature selection on each model's overall performance, providing useful insights in searching for the ideal balance between feature efficiency and the model performance.

Examining the outcomes presented in Figure 29 concerning Type I and Type II errors, a visible pattern emerges. It becomes apparent that the $F_1$ and $F_2$ feature subsets outperform the utilization of the entire set of heart rate variability features, except for the model NB. In terms of Type I error (Figure 29(a)), $F_1$ exhibits superior performance for the ET model, $F_2$ performs better for kNN, and for the remaining models, identical values are observed for both $F_1$ and $F_2$. It is worth noting that kNN had the lowest Type I error among the models considered. Shifting the focus to Type II error (Figure 29(b)), $F_2$ demonstrates better results for ET and SVM, while $F_1$ outperforms for XGB. The remaining models exhibit equivalent values for both $F_1$ and $F_2$. Remarkably, the ET model emerges as the best performer in Type II error among the models compared.




(a) Type I Error




(b) Type II Error

Figure 29: Type I and II errors for drowsiness classification.

Consequently, the outcomes for accuracy, precision, recall, and $F_1$ metrics, depending on the subset of features chosen, are presented in Figure 30. When using a more simplified number of features, a distinct trend appears, indicating notable increases in accuracy, precision, recall, and F1 score. In terms of accuracy (Figure 30(a)), it is evident that the subset $FS_2$ yields superior results for the ET, kNN, and SVM models. The rest of the models exhibit comparable values for both $FS_1$ and $FS_2$, with the ET model standing out with higher accuracy. Considering precision (Figure 30(b)), $FS_1$ presents superior

performance for the ET model, whereas $FS_2$ outperforms for the kNN model, culminating in its optimal precision. Other models have almost comparable precision values for $FS_1$ and $FS_2$. Moving on to recall (Figure 30(c)), $FS_2$ delivers superior results for the ET and SVM models, with the ET model exhibiting the highest recall value. In terms of F1 score, models achieve comparable values for both $FS_1$ and $FS_2$ feature subsets, with the ET model demonstrating the highest F1 score. Upon comprehensive analysis, the ET achieved the best accuracy, recall, and F1 score when considering the $FS_2$ feature subset. This demonstrates the usefulness of feature selection in improving the model's predictive performance and highlights the potential of ET as the preferable model for drowsiness classification.


(a) Accuracy

(b) Precision

(c) Recall

(d) F1

Figure 30: Evaluation metrics for drowsiness classification.

In the evaluation of drowsiness detection models, the AUC serves as a crucial metric to assess their performance. The areas under the curve were generated (Figure 31) for the ET, XGB, kNN, NB, and SVM models and feature subsets ($F_1$ and $F_2$). For the ET model, both the $F_1$ and $F_2$ feature subsets demonstrated high AUC scores of 0.74, indicative of their robust discriminatory capabilities. Similarly, the SVM model exhibited promising results, with AUC values of 0.72 for $F_1$ and 0.74 for $F_2$. The XGB model had consistent performance, with AUC scores of 0.73 for both $F_1$ and $F_2$ feature subsets. However, the NB and kNN models achieved slightly lower AUC values of 0.67 for both $F_1$ and $F_2$, indicating a comparatively moderate discriminative ability. These findings underscore the effectiveness of the ET and SVM models

in accurately classifying drowsiness.

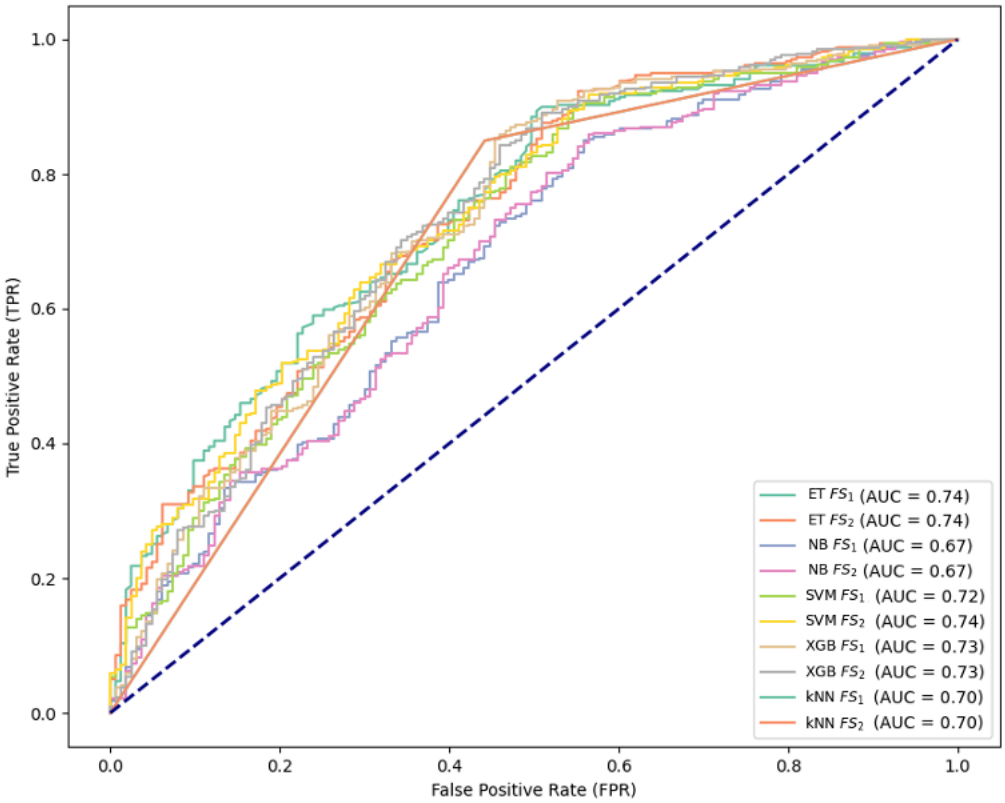



Figure 31: AUC curve for ET, XGB, kNN, NB, and SVM models.

Following the comprehensive analysis of evaluation metrics across all models, it becomes essential to highlight the selected features. Table 17 provides a detailed overview of the chosen features for each model, excluding $max\_hr$, $nni\_50$, $pnni\_50$, and $Sport$ that were not selected by any model. Notably, features such as $median\_nni$, $pnni\_20$, and $lf\_hf\_ratio$ appear as recurrent selections across various models, signifying their importance in the classification of drowsiness. The first two select features, $median\_nni$ and $pnni\_20$, are derived from the time domain of heart rate variability, while the $lf\_hf\_ratio$ is associated with the frequency domain. The prevalence of these features underscores their potential as key contributors to the models' performance, requiring further investigation into their individual impact on drowsiness detection. On the other hand, the features $mean\_nni$, $mean\_hr$, $nni\_20$, $sex$, $BMI$, and $Epworth$ were selected by 6 models, with the first three belonging to the time domain of heart rate variability, followed by personal information and sleep disorders.

Following the acquisition of these results, it is imperative to conduct a more specific analysis to distinguish between the drowsy and awake stages for the features that were identified most frequently. This study employed hypothesis testing for continuous independent samples, considering the 95% confidence level, starting with the assessment of normality in both awake and drowsy samples. Subsequently, the t-test was applied when both samples exhibited a normal distribution, while the Mann-Whitney test was utilized when at least one of them did not follow a normal distribution [112]. This analysis was conducted

Table 17: Features selected for each model.

| | | ET | | XGB | | kNN | | NB | | SVM | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Domain | Features | $FS_1$ | $FS_2$ | $FS_1$ | $FS_2$ | $FS_1$ | $FS_2$ | $FS_1$ | $FS_2$ | $FS_1$ | $FS_2$ | N |
| Time | mean_nni | x | | | | x | x | x | x | x | | 6 |
| | median_nni | x | x | x | x | x | x | | | x | x | 8 |
| | range_nni | | | | | x | x | | | x | x | 4 |
| | mean_hr | | | x | x | x | x | | | x | x | 6 |
| | min_hr | | | | | | | | | x | | 1 |
| | std_hr | | | | | | x | | | x | x | 3 |
| | sdnn | | | x | x | x | | x | x | | | 5 |
| | rmssd | | | | | x | x | | | | | 2 |
| | nni_20 | x | x | | x | | | x | x | | x | 6 |
| | pnni_20 | | x | x | | x | x | x | x | x | x | 8 |
| | cvsd | x | | x | x | x | x | | | | | 5 |
| | cvnni | | | | | | x | | | x | x | 3 |
| | sdsd | | x | | | | x | | | | | 2 |
| Frequency | power_vlf | x | | x | x | | | x | | | | 4 |
| | power_lf | | x | | | | | | | | | 1 |
| | power_hr | | | | | x | x | | | | | 2 |
| | total_power | | | x | x | x | x | | | | | 4 |
| | lf_hf_ratio | x | x | | | x | x | x | x | x | x | 8 |
| Non-Linear | csi | x | x | | | x | x | | | | | 4 |
| | cvi | x | | | | x | x | | | | | 3 |
| | Modified_csi | x | x | | | | x | | | x | x | 5 |
| | sampen | | | x | x | | | x | x | x | | 5 |
| Personal Information | Sex | x | x | x | | x | x | | | x | | 6 |
| | Age | | x | | | | | x | x | x | x | 5 |
| | BMI | x | x | x | x | x | x | | | | | 6 |
| Sleep Disorders | Epworth | x | | x | x | x | x | | | x | | 6 |
| | Evening | x | x | x | | | | | | | | 3 |
| | Morning | | | x | | | | x | x | | | 3 |
| | Neither | x | x | x | | | | | | x | x | 4 |
| | STOP-Bang | | | | x | x | | x | x | | | 4 |
| | Pittsburgh | | x | | | | | | | x | x | 3 |
| | N | 12 | 12 | 11 | 11 | 17 | 19 | 9 | 8 | 15 | 11 | |

individually for each participant to comprehend how the machine learning models are acquiring knowledge. To perform this analysis, it is important to achieve a balanced representation of both the drowsy and awake states by randomly sampling from each state to match the count of the less frequent class. First, the number of observations for each state was calculated, and then, the minimum count between the two states was identified. Subsequently, sample data was randomly selected sample data from each

state, ensuring that the sampled counts were equal to the minimum value. Initiating the analysis with the variables identified most frequently, significant differences were observed for $median\_nni$ between the awake and drowsy states in 16 participants, and the mean values are presented in Table 18. It is worth noting that $median\_nni$ represents the median value of the time between heartbeats, and intuitively, this value should be higher in the drowsy condition than in the awake state. Upon examining the individual values, it is evident that for 15 participants, the $median\_nni$ values are indeed higher during the drowsy state. For instance, in participant 14, the mean value for $median\_nni$ is 727.66, whereas, for the drowsy state, it is 820.06, underscoring a consistent pattern across the participants. However, for the participant 24, the opposite trend was identified. Analyzing the responses of that participant to the questionnaires, the following was observed: the participant experienced stress 24 hours before the simulation, and during the simulation, it also reported feelings of anxiety, and difficulties in attention, and concentration. It is essential to recognize that these symptoms may influence heart rate values, and there might be underlying pathologies not accounted for in the analysis.

Table 18: Mean value for awake and drowsy state of $median\_nni$ feature.

| ID | Awake | Drowsy |
|---|---|---|
| 14 | 727.66 | 820.06 |
| 15 | 724.10 | 772.60 |
| 16 | 785.52 | 822.48 |
| 20 | 775.87 | 805.34 |
| 21 | 634.07 | 913.06 |
| 24 | 873.92 | 828.10 |
| 28 | 677.32 | 942.28 |
| 29 | 745.24 | 799.74 |
| 33 | 704.80 | 741.43 |
| 38 | 843.61 | 912.84 |
| 41 | 778.59 | 926.21 |
| 47 | 639.03 | 683.13 |
| 48 | 698.29 | 759.10 |
| 53 | 809.66 | 858.92 |
| 55 | 755.15 | 868.35 |
| 58 | 842.42 | 959.06 |

In relation to $pnni\_20$ feature, significant differences were observed in 11 participants between the awake and drowsy states. Table 19 shows the mean values for the awake and drowsy states of the $pnni\_20$ feature. In terms of numerical values, the awake state exhibited a higher $pnni\_20$ compared to the drowsy state. It is crucial to recall that $pnni\_20$ represents the proportion of $nni\_20$ divided by the total number of NN intervals, where $nni\_20$ signifies the count of instances where the time difference between heartbeats exceeds 20 milliseconds. A higher $pnni\_20$ value suggests increased physiological activity during wakefulness, indicating more instances of heartbeats with differences exceeding 20 milliseconds. This underscores the potential of $pnni\_20$ as a valuable metric for discerning autonomic activity variations

between awake and drowsy conditions. Furthermore, a similar conclusion was corroborated in the research conducted by [49].

Table 19: Mean value for awake and drowsy state of $pnni\_20$ feature.

| ID | Awake | Drowsy |
|---|---|---|
| 12 | 84.87 | 80.25 |
| 14 | 80.67 | 71.43 |
| 15 | 73.95 | 67.23 |
| 19 | 80.67 | 69.33 |
| 20 | 75.63 | 68.49 |
| 39 | 85.29 | 78.99 |
| 40 | 81.09 | 71.89 |
| 45 | 80.67 | 63.87 |
| 46 | 83.19 | 65.55 |
| 47 | 60.92 | 36.55 |
| 50 | 78.99 | 59.66 |

For the feature $lf\_hf\_ratio$, significant differences were observed between the awake and drowsy states for 6 participants. On average, values during the drowsy state were higher compared to the awake state, Table 20. However, it is important to note that $lf\_hf\_ratio$ represents the ratio between $LF$ (low frequency, associated with sympathetic activity dominance) and $HF$ (high frequency, associated with parasympathetic activity). It would be expected that the drowsy state would exhibit lower values than the awake state. To address this discrepancy, a reassessment of questionnaire responses revealed that three individuals consumed coffee before the simulation, with one of them also being a smoker. Additionally, two participants were taking medication, one for hypertension, and another participant had obesity. These conditions can impact heart rate, particularly the metrics related to frequency. These findings highlight the importance of considering external factors such as caffeine intake, smoking, medication use, and pre-existing medical conditions when interpreting metrics of heart rate variability. The same issue was identified in the study by [49], and they successfully linked it to stress.

Table 20: Mean value for awake and drowsy state of $lf\_hf\_ratio$ feature.

| ID | Awake | Drowsy |
|---|---|---|
| 14 | 0.89 | 1.94 |
| 15 | 0.80 | 1.49 |
| 30 | 0.89 | 1.76 |
| 46 | 1.89 | 2.65 |
| 47 | 0.93 | 4.26 |
| 55 | 1.52 | 3.17 |

## 6.3 Forecasting

In the pursuit of advancing the classification of drowsy states through machine learning models, the subsequent phase involved validating the potential to forecast drowsiness. This step encompassed the deployment of XGB and ET regression models to forecast the final two minutes of simulation, based on the subset of features identified as $FS_1$ and $FS_2$. Subsequently, the analysis extended to the classification of new data. Considering the forecasted values for each feature, the process involved utilizing the trained ET classification model, established earlier, to classify the drowsiness stage. Since the main goal of this PhD thesis is to predict drowsiness through the utilization of a wearable device, it is expected to achieve that goal by employing both forecasting and classification models.

Thus, forecast analysis was confined to the features of the ET classification model, chosen deliberately due to its superior overall performance, as elucidated above. In the context of forecasting the last two minutes of simulation, the MSE, RMSE, and MAE were employed as key metrics. The evaluation metrics for the XGB and ET regressor models, as presented in Table 21, offer valuable insights into the performance of these models in predicting the heart rate variability based on the feature subset $FS_1$. The results reveal notable variations in the performance of the two models across different features, and minimizing errors is crucial for optimal performance in heart rate variability prediction. For the forecasting of $FS_1$ features, the XGB regressor model consistently outperforms the ET regression model across all the metrics, showcasing lower MSE, RMSE, and MAE values. This suggests that the XGB regression model provides more accurate predictions for $mean\_nni$, $median\_nni$, $nni\_20$, $cvsd$, $power\_vlf$, $lf\_hf\_ratio$, $csi$, $cvi$, and $Modified\_csi$ features, as evidenced by reduced errors.

Table 21: Evaluation metrics for the XGB and ET regression models for the feature subset $FS_1$.

| $FS_1$ | | | | | | |
|---|---|---|---|---|---|---|
| | ET Forecasting | | | XGB Forecasting | | |
| Metrics | MSE | RMSE | MAE | MSE | RMSE | MAE |
| mean_nni | 1.06 | 1.03 | 0.80 | 0.86 | 0.93 | 0.57 |
| median_nni | 1.05 | 1.03 | 0.75 | 0.89 | 0.94 | 0.57 |
| nni_20 | 0.97 | 0.98 | 0.79 | 0.38 | 0.62 | 0.38 |
| cvsd | 0.99 | 0.99 | 0.76 | 0.31 | 0.56 | 0.32 |
| power_vlf | 2.38 | 1.54 | 1.17 | 0.88 | 0.94 | 0.65 |
| lf_hf_ratio | 1.40 | 1.18 | 0.93 | 0.73 | 0.85 | 0.56 |
| csi | 1.30 | 1.14 | 0.95 | 0.43 | 0.66 | 0.49 |
| cvi | 1.14 | 1.07 | 0.87 | 0.31 | 0.55 | 0.39 |
| Modified_csi | 1.44 | 1.20 | 0.96 | 0.56 | 0.75 | 0.51 |

Remarkably, a consistent pattern emerges in the $FS_2$ features (Table 22), revealing that the XGB regression model consistently outperforms the ET regression model by exhibiting lower errors across all features. Besides that, the lower MSE, RMSE, and MAE values associated with the XGB regression model signify its superior ability to capture underlying patterns in the data, contributing to enhanced best

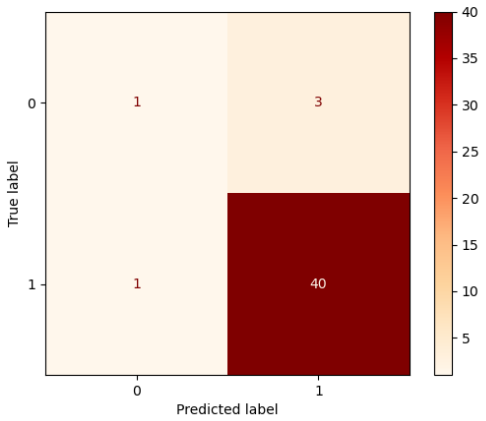
results in forecasting heart rate measures. These findings underscore the effectiveness of the XGB regression model in minimizing prediction errors for the $median\_nni$, $nni\_20$, $pnni\_20$, $sdsd$, $power\_lf$, $lf\_hf\_ratio$, $csi$, and $Modified\_csi$ features, emphasizing its potential as a robust tool for heart rate variability prediction in real-world applications.

Table 22: Evaluation metrics for the XGB and ET Regression models for the feature subset $FS_2$.

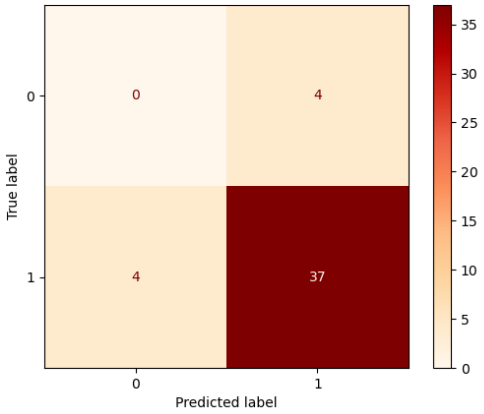
| $FS_2$ | | | | | | |
|---|---|---|---|---|---|---|
| | ET Forecasting | | | XGB Forecasting | | |
| Metrics | MSE | RMSE | MAE | MSE | RMSE | MAE |
| median_nni | 1.09 | 1.04 | 0.78 | 0.89 | 0.94 | 0.57 |
| nni_20 | 0.92 | 0.96 | 0.75 | 0.38 | 0.62 | 0.38 |
| pnni_20 | 0.94 | 0.97 | 0.76 | 0.38 | 0.62 | 0.38 |
| sdsd | 1.02 | 1.01 | 0.77 | 0.26 | 0.51 | 0.33 |
| power_lf | 1.66 | 1.29 | 0.95 | 0.78 | 0.88 | 0.57 |
| lf_hf_ratio | 1.37 | 1.17 | 0.92 | 0.73 | 0.85 | 0.56 |
| csi | 1.39 | 1.18 | 0.98 | 0.43 | 0.66 | 0.49 |
| Modified_csi | 1.44 | 1.20 | 0.96 | 0.56 | 0.75 | 0.51 |

Following the comparative analysis of the heart rate variability predictions, it becomes imperative to apply the ET classification model (the one chosen in Section 6.2) to classify the individual's drowsiness state. In order to comprehensively assess the efficacy of the classification model, four distinct confusion matrices were employed to evaluate the predictive performance based on both XGB and ET forecasting values. This meticulous analysis holds significant importance as it serves as a pivotal means to substantiate the model's capability in forecasting drowsiness, thereby enabling the determination of its predictive accuracy and potential for advanced drowsiness prediction. The evaluation of the classification models (Figure 32), employing both XGB and ET forecasting values across two distinct feature subsets ($FS_1$ and $FS_2$), reveals nuanced insights into their predictive performance for anticipating drowsiness. For the XGB model applied to the $FS_1$ feature set (Figure 32(a)) a commendable outcome is observed with 40 true positives and a mere 1 false negative, demonstrating its efficacy in correctly identifying instances of drowsiness. Nevertheless, the analysis reveals a presence of 3 false positives, instances where the participant is genuinely awake but erroneously classified as drowsy by the model. This discrepancy raises practical implications, as misidentifying wakefulness for drowsiness may lead to unnecessary interventions or alerts. Conversely, the model demonstrates a more favorable performance in terms of false negatives, with only 1 occurrence. This scenario, although less frequent, is critical as it represents instances where the participant is genuinely drowsy, yet the model incorrectly categorizes them as awake. Conversely, the XGB model with the $FS_2$ feature set (Figure 32(b)) displays slightly diminished performance, marked by 4 false negatives and 4 false positives alongside 37 true positives. In contrast, the ET model, within the $FS_1$ feature subset, achieves 39 true positives while incurring only 2 false negatives and is accompanied by the presence of 3 false positives. In the case of the $FS_2$ feature subset, the ET model achieves 36 true

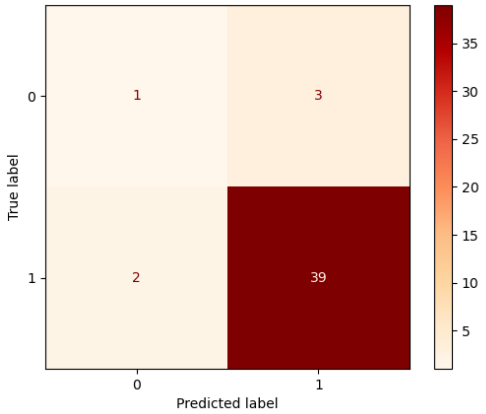
positives but at the cost of 5 false negatives and 3 false positives.



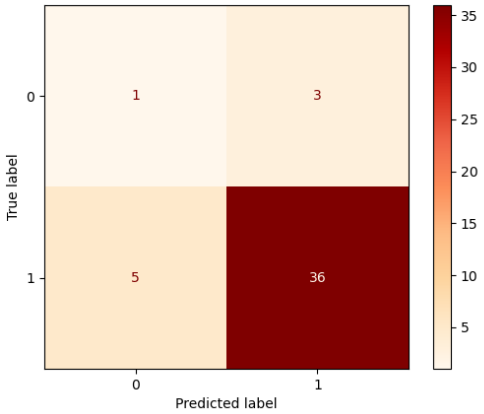

(a) Classification for XGB forecasting of $FS_1$.




(b) Classification for XGB forecasting of $FS_2$.




(c) Classification of ET forecasting of $FS_1$




(d) Classification of ET forecasting of $FS_2$

Figure 32: Confusion matrix using the forecasting results.

Following an extensive examination of individual identification regarding false positives and negatives, a noteworthy pattern emerged. The investigation revealed a consistent presence of three false positives attributed to the same individuals (15, 27, and 38) across all confusion matrices. Subsequently, a thorough review of the corresponding video was conducted, leading to the identification of slight signs of drowsiness exhibited by these individuals. This observation not only emphasizes the persistence of the false positives but also suggests a common underlying factor contributing to the misclassifications, namely, the manifestation of slight drowsiness cues in the observed behavior. In the context of false negatives, it was noted that individual 39 consistently appeared in both the XGB analysis for $FS_1$ and $FS_2$ subsets of features. Similarly, individual 25 was identified as a recurrent false negative in the results obtained from the ET analysis. This consistency across different feature subsets and algorithms suggests a distinct pattern associated with these individuals, underscoring the need for a focused investigation into the factors

contributing to their recurrent misclassification as false negatives. Upon a detailed video analysis, it was observed that both participants exhibited sight signs of drowsiness. This implies that the model is inaccurately categorizing the drowsiness state of these individuals who are showing signs of drowsiness. The model is erroneously considering them to be awake, despite clear indications of drowsiness. The disparity between the observed signs and the model predictions highlights a limitation or flaw in the model's ability to effectively recognize drowsiness in these specific cases.

After a thorough examination of the confusion matrix results, it is necessary to compute evaluation metrics (Table 23) for both forecasting models (XGB and ET) and their related feature subsets, notably $FS_1$ and $FS_2$. The evaluation of the XGB forecasting model reveals notable strengths across multiple metrics. In terms of the $F_1$ score evaluation metric score, it achieved values of 0.95 and 0.90 for the $FS_1$ and $FS_2$ feature subsets, respectively. This reflects a balanced performance in precision and recall, underscoring the model's efficacy in accurately identifying instances while minimizing false positives and false negatives. Additionally, the low Type I error rates (0.07) indicate a commendable ability to avoid false positive classifications. However, a slightly elevated Type II error rate for the $FS_2$ features (0.09) suggests a somewhat increased tendency to miss positive instances compared to the $FS_1$ features (0.02). Despite this, the overall accuracy remains high at 0.91, signifying the model's robust performance in classifying instances across both positive and negative classes. The precision values (0.93 for $FS_1$ and 0.90 for $FS_2$) further emphasize the model's ability to minimize false positives, while the impressive recall values (0.98 for $FS_1$ and 0.90 for $FS_2$) underscore its capability to effectively identify positive instances. The AUC is moderate at 0.61 for $FS_1$ and 0.45 for $FS_2$, indicating that improvements must be made. Turning to the ET forecasting model, it showcases competitive performance across various metrics. The $F1$ evaluation metric scores for both $FS_1$ and $FS_2$ subsets stand at 0.94 and 0.90, respectively, reflecting the model's effectiveness in maintaining a balance between precision and recall. Similar to XGB, the ET model demonstrates low Type I error rates (0.07), emphasizing its precision in avoiding false positives. However, a higher Type II error rate for the $FS_2$ features (0.11) suggests an increased likelihood of missing positive instances in comparison to the $FS_1$ features (0.07). Despite this, the overall accuracy is commendable at 0.89, indicating the model's proficiency in correctly classifying instances. Precision values (0.93 for $FS_1$ and 0.92 for $FS_2$) underscore its ability to minimize false positives, while impressive recall values (0.95 for $FS_1$ and 0.88 for $FS_2$) highlight its effectiveness in identifying positive instances. The AUC for the ET model is 0.60 for $FS_1$ and 0.56 for $FS_2$, suggesting a moderate discriminatory ability, with potential for improvements, particularly for the $FS_2$ subset.

In summary, both the XGB and ET forecasting models demonstrate strengths in accurately predicting drowsiness. Examining the individual strengths and weaknesses inherent in each model, it is evident that XGB, when utilizing the $FS_1$ subset of features, demonstrated superior performance across the majority of evaluation metrics for drowsiness classification. Nevertheless, it is noticeable that there are few participants in the awake state, and this outcome can be attributed to the consideration of the last two minutes of the driving simulation. Given the inherently monotonous nature of driving, it is common for individuals not to be fully awake during this period, leading to the manifestation of subtle signs of drowsiness. There

Table 23: Evaluation metrics of the forecasting results.

| Metric | XGB Forecasting | | ET Forecating | |
|---|---|---|---|---|
| | $FS_1$ | $FS_2$ | $FS_1$ | $FS_2$ |
| Type I Error | 0.07 | 0.09 | 0.07 | 0.07 |
| Type II Error | 0.02 | 0.09 | 0.04 | 0.11 |
| Accuracy | 0.91 | 0.82 | 0.89 | 0.82 |
| Precision | 0.93 | 0.90 | 0.93 | 0.92 |
| Recall | 0.98 | 0.90 | 0.95 | 0.88 |
| F1 | 0.95 | 0.90 | 0.94 | 0.90 |
| AUC | 0.61 | 0.45 | 0.60 | 0.56 |

might be a concern about the model's ability to accurately identify when an individual is awake. Hence, the decision was made to incorporate the first 2 minutes of the simulation for 10 participants who were initially excluded due to video issues. Two participants are not being considered as they wear glasses, and due to the reflection of the monitor's light, the eyes cannot be clearly visualized. This choice is grounded in the understanding that examining the initial moments provides insights into whether or not individuals exhibit signs of drowsiness. It is crucial to remember that participants underwent a 10-minute adaptation period, adding context to the decision to include the first 2 minutes of simulation data. Out of the 10 participants, it is known that 7 of them are in an awake state, while the remaining individuals exhibit slight signs of drowsiness. The information about these individuals has been incorporated into the forecasting obtained from the best model mentioned above, in this instance, XGB, considering the $FS_1$ subset of features. Thus, Figure 33 illustrates the confusion matrix, taking into account the 10 participants not included in the previous analyses, along with the forecast values for the XGB model considering the subset of features $FS_1$. Based on the obtained results, the model correctly identified the 7 awake individuals and the 3 drowsy ones.

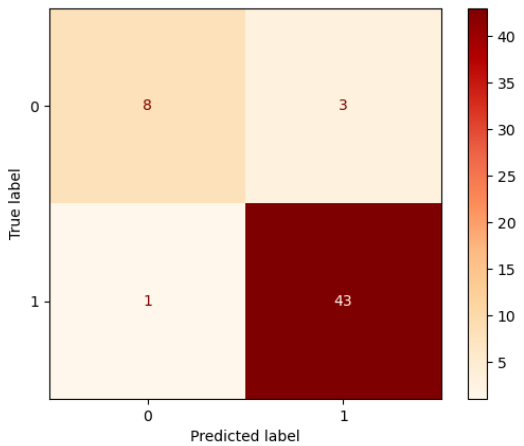



Figure 33: Confusion matrix using new information and forecasting results of XGB using $FS_1$.

The obtained results exhibit significant promise, as the model effectively discriminates between awake and drowsy states, despite the availability of more extensive information regarding the drowsy state. This holds profound implications, as the proposed methodology demonstrates the potential to predict drowsiness. Such predictive capabilities not only contribute to enhancing our understanding of wakefulness and drowsiness but also offer a practical avenue for accident prevention. Importantly, the implementation of this methodology can be achieved through a cost-effective solution, thereby amplifying its accessibility and applicability in real-world scenarios.

## 6.4 Summary

Predicting drowsiness at the wheel remains a significant challenge, necessitating effective solutions for defining preventive actions and reducing accidents. The subjective nature of awake and drowsy state classification poses an additional challenge, as it should not depend on individual opinions. In this context, this PhD thesis successfully tackles the drowsiness detection challenge through an innovative methodology employing multivariate statistical process control with PCA. Initial promising results using a single principal component derived from heart rate variability domains led to further refinement, considering three principal components corresponding to time, frequency, and non-linear domains of the heart rate variability. This not only validated the methodology's effectiveness but also enhanced the detection of subtle signs of drowsiness. The comparison with manual classifications and application of the KSS scale affirmed the relevance of the proposed approach. Integration of accelerometer data provided valuable insights into noise impact, suggesting that heart rate variability instances were less influenced by sudden movements. The analysis revealed that 14.29% of shared out-of-control points align with drowsiness transitions, and 66% of accelerometer points occur at the simulation's outset. Recognizing the subjective nature of drowsiness classification, adjustments were made by comparing it with heart rate variability out-of-control points, resulting in improved precision (0.83) and recall (0.47). Despite challenges in recognizing true drowsiness transitions, the methodology significantly reduces subjectivity and aligns with a majority of out-of-control points. Based on 1 672 heart rate variability records, the study highlights a notable shift from awake to drowsy states, effectively addressing issues like subjective classification and susceptibility to signal noise.

Furthermore, the optimization of feature selection in drowsiness detection is explored, emphasizing the importance of identifying impact features. The approach employs a genetic algorithm within the wrapper method and evaluates various machine learning models for drowsiness classification. A multi-objective optimization strategy is implemented to minimize both Type I and Type II errors. The study meticulously analyzes the obtained solutions for each model, leading to the identification of specific feature subsets. The evaluation metrics, including accuracy, precision, recall, and F1 score, demonstrate improved performance with reduced feature complexity. The study highlights the effectiveness of the ET model, especially when considering a specific feature subset ($FS_2$). The analysis extends to the assessment of models using the AUC metric, emphasizing the discrimination capabilities of the ET and SVM models. Finally, the

selected features, such as $median\_nni$, $pnni\_20$, and $lf\_hf\_ratio$, are identified as key contributors to drowsiness classification. The findings provide valuable insights for developing efficient and interpretative drowsiness detection models. Following the acquisition of results, a detailed analysis was conducted to differentiate between drowsy and awake stages, focusing on frequently identified features. Employing hypothesis testing and individual participant assessments, 16 participants exhibited significant differences in $median\_nni$, consistently higher during drowsy states, except for participant 24, potentially influenced by reported stress. Besides that, $pnni\_20$ differences in 11 participants indicated higher values during wakefulness, aligning with existing research. However, $lf\_hf\_ratio$ differences in 6 participants contradicted expectations, potentially influenced by external factors like caffeine intake, smoking, medication, and pre-existing conditions. This emphasizes the need to consider external influences when interpreting heart rate variability metrics, as highlighted in existing studies

In the pursuit of enhancing drowsiness classification, this study employs XGB and ET regression models to forecast the last two minutes of a driving simulation. The subsequent classification analysis utilizes the ET model to classify drowsiness based on forecasted values and evaluation metrics (MSE, RMSE, MAE) underscore XGB's superior performance in forecasting heart rate variability. The classification analysis, using confusion matrices, revealed nuanced insights into predictive performance, highlighting XGB's strengths in identifying instances of drowsiness accurately. Nonetheless, challenges persist, including recurrent misclassifications for specific individuals, prompting further investigation. Overcoming limitations in wakefulness identification by incorporating the initial two minutes of simulation for specific participants reinforces the model's efficacy. This methodology not only refines our comprehension of wakefulness and drowsiness but also showcases promising predictive capabilities with practical implications for accident prevention, making a valuable contribution to the field of drowsiness detection in real-world scenarios.

# 7

# Conclusions and Future Work

*"It is hard to fail, but it is worse never to have tried to succeed." By Theodore Roosevelt*

Sleep is fundamental for physical and psychological health, significantly influencing cognitive processes, emotional regulation, and sustained attention. Moreover, sleep is not only crucial for personal health but also has profound societal impacts, affecting productivity, safety, and quality of life. The importance of sleep extends to its role in learning, memory consolidation, and attention span, which are vital for day-to-day functions and long-term mental health. The adverse effects of insufficient sleep can be severe, ranging from an increased risk of serious health conditions such as hypertension, diabetes, and cardiovascular diseases to heightened instances of workplace accidents and diminished overall productivity. Studies consistently indicate a clear connection between poor sleep quality and significant public health challenges, the urgent need for a more comprehensive understanding and interventions aimed at improving sleep health.

Our research was motivated by the existing gap in the evaluation of various sleep disorders and the awareness of sleep quality among the Portuguese population, across all districts and the necessity for a comprehensive study that encompasses a representative sample. By focusing on this aspect, the research aims to address the lack of knowledge concerning sleep-related challenges faced by individuals across Portugal. Understanding the prevalence and patterns of sleep disorders is essential for developing effective interventions and improving overall public health outcomes. Therefore, the motivation lay in bridging this gap in research, ultimately contributing to a better understanding of sleep quality and related disorders among the diverse drivers of Portugal. Another motivation for this research was the need to assess the alignment between circadian rhythms and work schedules. This analysis holds significance due to its potential impact on productivity and the prevalence of excessive daytime sleepiness. When work schedules deviate from individuals' natural circadian rhythms, it can lead to decreased productivity and an increased risk of accidents in the workplace. Notably, there is also a gap in the literature regarding this aspect, particularly in the context of the Portuguese population. By addressing this gap, the research aimed to

provide valuable insights to the field, potentially increasing awareness of this topic among the population through the presentation of relevant statistics. This approach aims to foster a deeper understanding of the importance of aligning circadian rhythms with work schedules. The research was also motivated by the need to address the issue of drowsiness at the wheel. This aspect demands the identification of reliable, cost-effective solutions that seamlessly integrate with the driver's activity. Current methodologies have demonstrated effectiveness in recognizing a driver's drowsy state and sending out alerts on time. However, a significant drawback is that these methods are often expensive, making them inaccessible to a wider population. This economic barrier creates a division, wherein only individuals with greater financial means can benefit from such potentially life-saving technology. Therefore, this research aimed to contribute to the advancement of non-intrusive, cost-effective solutions for addressing drowsiness at the wheel. By detecting and predicting drowsiness in advance, while the driver is still operating at full capacity and shortly before their abilities begin to diminish these innovations aimed to significantly enhance driving safety for all.

In summary, this thesis proposes and contributes to the advancement of knowledge in the field of sleep through a multifaceted approach. Beginning with the development and validation of a comprehensive sleep disorders questionnaire tailored to the Portuguese drivers the study aimed to conduct a comprehensive investigation, employing rigorous statistical analyses to offer a detailed understanding of the prevalence and patterns of sleep disorders among this demographic. Moreover, the implementation of data analysis techniques facilitated the visualization of the results, enabling a clearer interpretation of the findings. The second contribution entailed using an affordable wearable device to track heart rate variability for classifying and predicting drowsiness in advance. Additionally, beyond identifying the optimal model for this classification and prediction, the study evaluated the most effective subset of features for distinguishing between awake and drowsy states.

For a better understanding of the conclusions achieved, this chapter is divided into two sections. The initial section, titled "Contributions and Critical Review" (Section 7.1), addresses the research questions and offers a critical review of the results obtained throughout the research process, aimed at fulfilling the objectives outlined in this thesis. The following section outlines potential future research directions (Section 7.2), briefly exploring promising opportunities and their potential contributions to enhancing and advancing the work presented in this thesis, along with associated research challenges.

## 7.1 Contributions and Critical Review

This thesis is divided into three main parts, as outlined in Chapter 1. The first objective focused on evaluating the sleep quality and prevalence of excessive daytime sleepiness among Portuguese drivers. The second objective aimed to investigate the impact of circadian rhythm on driver alertness and performance, taking into account work schedules. Lastly, the thesis aimed to identify specific physiological signals, collected through wearable devices, capable of detecting and predicting drowsiness in drivers. Through these

objectives, the research provided valuable insights into understanding and mitigating the risks associated with drowsiness in driving contexts.

Within this framework, the research questions outlined in Section 1.1 have served as guiding principles for the work presented and discussed in the various chapters. Now, each research question will be analyzed to provide a conclusive answer.

**Q1: What are the sleep quality and prevalence of excessive daytime sleepiness among Portuguese drivers?**

Sleep quality and the prevalence of excessive daytime sleepiness are two significant sleep-related concerns. Sleep quality refers to the overall subjective experience of sleep, incorporating multiple dimensions of sleep architecture and continuity. It encompasses various factors such as total sleep time (the duration of sleep obtained), sleep onset delay (the time taken to fall asleep after going to bed), sleep fragmentation (the extent to which sleep is interrupted), total wake time (the amount of time spent awake during the sleep period), sleep efficiency (the percentage of time spent asleep compared to the total time spent in bed), and disruptive events during sleep (such as awakenings, disturbances, or abnormal movements). Bad sleep quality can have serious effects on an individual's well-being and functioning. It may impact mood, leading to increased irritability and mood swings. Additionally, poor sleep quality can impair attention, concentration, and memory, making it difficult to focus and retain information. Feelings of tiredness are also common consequences of poor sleep quality, contributing to reduced energy levels and productivity throughout the day. On the other hand, daytime sleepiness refers to a persistent feeling of drowsiness and the tendency to fall asleep during the daytime, even in situations where it's inappropriate or dangerous, such as during work or while driving. It can result from various factors, including inadequate sleep duration or quality, untreated sleep disorders like sleep apnea or insomnia, irregular sleep schedules, or certain medications. Therefore, excessive daytime sleepiness indicates a reduction in cognitive abilities, which can impact an individual's capacity to concentrate and, consequently, their productivity.

Considering all of these factors, addressing this question necessitated the creation of a questionnaire to be disseminated across all districts of Portugal. The questionnaire was structured into three main sections aimed at gathering personal information, driving-related questions, and assessing sleep disorders using the components from established questionnaires such as ESS, STOP-Bang, PSQI, and MESSi. Rather than aiming for the highest possible number of responses, a stratified sampling approach was adopted to ensure a representative sample from each region of Portugal, based on the NUTS II classification system, with the aim of determining the optimal number of responses needed for robust analysis and insights. Considering this statistical approach, the minimum number of responses required was determined to be 1 149 as mentioned in the Section 4.1.

After one year of dissemination and based on the results described in Section 4.2, 2 087 valid

responses were obtained, surpassing the initial minimum requirement. These responses provide a robust dataset for analyzing the sleep quality and prevalence of excessive daytime sleepiness among Portuguese drivers. In comparison to the number of responses reported in existing studies within the literature (Section 1.1), it significantly surpasses typical sample sizes for studies involving the Portuguese population. This study not only contributes valuable insights through the number of responses attained but also through its clear specification of the study sample and the process used to determine the target values for each region of Portugal. Among the respondents, notable findings emerge regarding sleep disorders: 22.14% of individuals are identified as having a high risk of developing sleep apnea, indicating a concerning potential health issue. Furthermore, a significant portion, 38.8%, report experiencing excessive daytime sleepiness, highlighting a prevalent concern affecting driver alertness and safety. Additionally, a substantial majority, 60.2%, report poor sleep quality, suggesting a widespread challenge that may impact various aspects of well-being and performance. Although these numbers provide an answer to the research question, statistical methods were further applied to extract even more valuable insights. Consequently, it was revealed that 25.0% of Portuguese drivers are unaware of their actual sleep quality, as they perceive it to be good when it is, in fact, poor. Another noteworthy observation is that individuals who experience significant daytime sleepiness often have bad sleep quality. In this study, it was found that 70.3% of Portuguese drivers with excessive daytime sleepiness also have poor sleep quality. Conversely, it was also revealed that individuals who experience significant daytime sleepiness are approximately 1.8 times more likely to develop sleep apnea. To enhance decision-making and enable faster, more informed choices, a dashboard was developed to present crucial and valuable information to Portuguese society.

This research makes significant contributions to the field by comprehensively exploring sleep-related concerns, particularly focusing on sleep quality and the prevalence of excessive daytime sleepiness among Portuguese drivers. By surpassing the initial target of valid responses and utilizing a stratified sampling approach, the research ensures a representative dataset for in-depth analysis. Furthermore, the development of a dashboard to disseminate crucial information to Portuguese society demonstrates the practical implications of the research findings. Overall, the thesis advances understanding of sleep disorders in the context of driving and underscores the importance of promoting healthier sleep habits and mitigating risks for drivers in Portugal. Additionally, this analysis served as the foundation for the article titled "Sleep Disorders in Portugal Based on Questionnaires" [15].

**Q2: What is the influence of the circadian rhythm on drowsiness at the wheel?**

The circadian rhythm, as described in Section 1.1, is our body's internal clock, dictating when an individual feel awake or sleepy. It influences various aspects of our daily life, including emotions, body temperature, and hormone levels. However, when our work schedule is not align with this natural rhythm, it can disrupt our rest cycles, physiological functions, and even our behavior. This

lack of coordination can contribute to the development of sleep disorders. When individuals experience excessive daytime sleepiness due to disruptions in their circadian rhythm or inadequate sleep, their ability to drive safely is compromised.

Given the outlined consequences, addressing this research question can be facilitated through the utilization of the previously mentioned questionnaire, which includes the MESSi questionnaire along with other assessments of sleep disorders. Additionally, a question regarding work schedules was included, offering four options: day, night, shift work, or none of the above for participants who do not work. Numerically, 60.0% of participants exhibit characteristics associated with morning types, while 34.8% display elements typical of evening types, with the remaining falling into the intermediate category. Moreover, 25.0% of participants show a distinctness effect greater than 19, indicating substantial fluctuations in performance and mood throughout the day, as revealed by the questionnaire. Hence, the subsequent analysis focused on assessing the alignment between work schedules and the circadian rhythm, revealing a significant association between the two. Among individuals classified as morning types, the majority tend to work during the day, whereas those identified as evening types are more inclined to work at night. However, for individuals engaged in shift work, the proportions between morning and evening types are balanced. While these are indeed relevant information, the focus should be on individuals whose work schedules are not align with their circadian rhythm. Therefore, among individuals classified as evening types who work during the day, 44.9% exhibited symptoms of excessive daytime sleepiness, while for those engaged in shift work, the percentage rises to 53.9%. Conversely, among individuals identified as morning types who work during the night or shifts, the prevalence of excessive daytime sleepiness was 25.0% and 38.5%, respectively. Two additional important analyses are based on the driving questions. Among individuals whose work schedule does not align with their circadian rhythm, and who typically drive but have had to stop the car due to drowsiness, 67.4% exhibit symptoms of excessive daytime sleepiness. Furthermore, among those who reported falling asleep while driving, 73.3% also display symptoms of excessive daytime sleepiness.

The findings highlight how the circadian rhythm has impact on driver drowsiness. Individuals whose work schedules are in conflict with their natural circadian rhythms are more likely to have excessive daytime sleepiness, particularly when engaged in driving activities. This finding highlights the importance of aligning work schedules with individuals' circadian preferences to mitigate the risk of drowsiness-related incidents on the road. Furthermore, the high prevalence of excessive daytime sleepiness among individuals who have had to stop their cars due to drowsiness, as well as those who have actually fallen asleep while driving, emphasizes the critical role of addressing sleep-related issues in promoting road safety. Therefore, interventions aimed at optimizing sleep schedules and raising awareness about the impact of the circadian rhythm on drowsiness at the wheel are essential for enhancing driver alertness and reducing the incidence of drowsiness-related accidents. It's important to note that drowsiness may also be influenced by factors beyond just the

circadian rhythm.

**Q3: What is the best classifier to predict drowsiness at the wheel?**

The issue of drowsiness at the wheel presents a critical challenge in ensuring road safety, necessitating the exploration of effective and affordable solutions. Traditionally, methods to address this concern have involved the integration of movement sensors within the wheel and strategically positioned cameras in the rear-view mirror. These systems demonstrate proficiency in detecting a driver's drowsy state and issuing alerts to prevent potential accidents. Yet, their broad utilization faces a hurdle due to the high implementation costs, restricting access primarily to a privileged demographic. This economic inequality increases the gap by allowing those with higher incomes to make use of this technology. Moreover, the studies prevalent in the literature tend to prioritize identifying the best model based on a singular metric rather than comprehensively understanding the importance of features and behavior across multiple evaluation metrics. Additionally, the studies often focus on classifying drowsiness rather than predicting it in advance.

Hence, to address this research question, driving simulations were executed employing a wearable device (Empatic E4) to gather heart rate variability data. This biometric information was then utilized to classify and forecast drowsiness levels. Following the simulations, drowsiness states of the participants were manually classified based on recorded video footage of their faces. Recognizing the inherent subjectivity of manual classification, steps were taken to address this limitation. Multivariate statistical process control, specifically employing three principal components (time, frequency and non-linear domains), was implemented to enhance drowsiness classification and minimize subjectivity, as elaborated in Section 6.1. Subsequent to this enhancement, feature selection optimization was carried out to minimize both Type I and II errors. This involved evaluating eight machine learning algorithms (LR, ET, XGB, AdaBoost, kNN, NB, SVM and LDA) to identify the best subset of features and classify drowsiness effectively. A complete analysis was conducted (as demonstrated in Section 6.2) to choose the best model based on parameters such as accuracy, precision, recall, F1 score, and AUC. Hence, the drowsiness classification model that yielded the most favorable outcomes was the ET model, reaching an accuracy of 76.0%, precision of 79.0%, recall of 89.0%, F1 score of 84.0%, and AUC of 74.0%. Notably, this model only considered 12 features, a reduction from the original 31 features. Nevertheless, the primary objective of the current research question is to forecast drowsiness in advance. Therefore, ET and XGB regression models were employed, as described in Section 6.3, to predict the last two minutes of the simulation data for each variable. Significantly, the XGB regression model demonstrated superior performance compared to the ET regression model when utilizing the subset of features identified through the feature selection process. Consequently, the predicted values for each feature were incorporated into the ET classification model to discern whether the individual was awake or drowsy. The best model exhibited errors in only 4 individuals, with 3 of them being classified as awake by the model while actually being drowsy. Upon manual inspection of the recorded videos, it was determined that three individuals did indeed display signs of

drowsiness, indicating an incorrect manual classification. Given the dataset's imbalance, with more information available on drowsy events, the model may accurately classify these instances but struggle with awake events. Consequently, there was a concern regarding the model's proficiency in discerning when an individual is awake. To address this, the first 2 minutes of simulation data for 10 participants initially excluded due to video issues were incorporated. Among these participants, 7 were confirmed to be in an awake state, while the remaining individuals showed slight signs of drowsiness. Encouragingly, the model correctly identified all 7 awake individuals and accurately detected the 3 drowsy ones. To elucidate the essential steps in addressing the current research question, a framework was designed (see Figure 34). This framework comprises four primary stages: driving simulations, drowsiness classification, features forecasting, and drowsiness prediction.

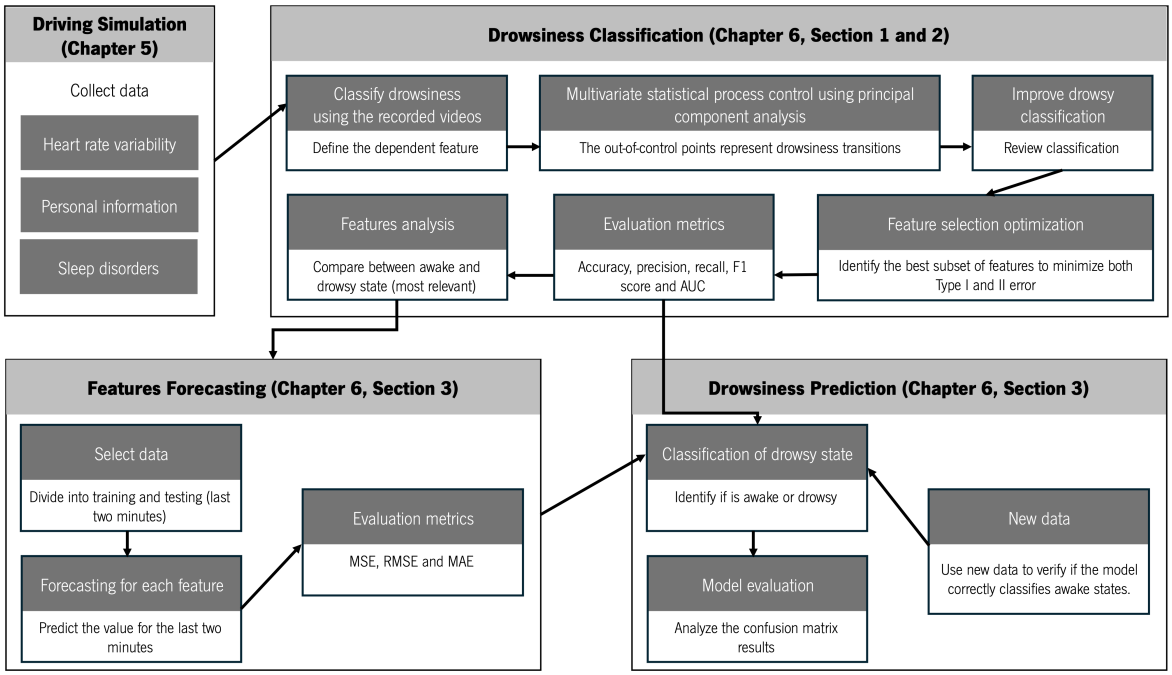



Figure 34: Predictive Framework for Drowsiness Detection.

This research question posed the most significant challenge within the scope of this thesis, demanding meticulous attention and thorough investigation. However, despite its complexity, promising outcomes were attained in predicting drowsiness in advance. Methodologically, every aspect of the study's implementation is meticulously detailed, providing transparency and reproducibility. An impressive contribution of this thesis is its ability to accurately predict drowsiness before it occurs, made achievable through the application of a cost-effective wearable device. This affordable solution ensures accessibility to a wider audience, allowing anyone to benefit from its capabilities. Besides that, throughout this study, five articles [13, 11, 12, 14, 109] were developed to provide some answers to the research question, and another one is currently being developed to present drowsiness predictions.

## 7.2 Future Research Directions

Awareness of several research pathways exists to enhance the findings presented in this thesis. Subsequently, the following highlights some intriguing prospects for future investigation.

**Dataset Expansion:** In both the study conducted on the Portuguese population and the driving simulations to predict drowsiness, expanding the sample size is essential. This expansion should aim to balance the representation of both male and female individuals and ensure a proportional representation of awake and drowsy states. Additionally, increasing the diversity of participants across different age groups, occupations, and driving experience levels can enhance the generalizability and robustness of the findings. Moreover, collecting data from a larger sample size enables researchers to conduct subgroup analyses, investigating potential differences in drowsiness patterns among various demographic groups. Furthermore, incorporating additional physiological metrics such as oxygen levels, eye movement patterns, and temperature into the drowsy dataset can provide valuable insights and improve the classifications and predictions.

**Improve Drowsiness Classification:** To refine drowsiness classification, it is imperative to transition to less subjective methods, leveraging insights from medical expertise. Collaborating with medical professionals specializing in sleep medicine can offer valuable insights into the physiological of drowsiness and inform the development of more robust classification models. Incorporating objective physiological markers, such as respiration rate and eye movement patterns, can provide a more accurate assessment of drowsiness levels. Additionally, integrating advanced signal processing techniques and machine learning algorithms trained on diverse datasets can enhance the classification accuracy. Moreover, exploring multimodal approaches that combine data from wearable devices with other physiological measures, such as brainwave activity, holds promise for improving drowsiness detection in real-world settings.

**Real-World Implementation:** The deployment of the proposed methodology for drowsiness prediction must transition from simulated environments to real-world contexts, where it can effectively alert drivers that are going to present signs of drowsiness. This involves integrating the predictive model into low-cost wearable devices equipped with physiological sensors, such as heart rate variability. These devices can continuously monitor the user's physiological signals and provide real-time alerts indicating the likelihood of drowsiness onset. By leveraging wearable technology, individuals can receive personalized alerts and take proactive measures to mitigate the risk of drowsiness-related incidents, such as taking breaks or adjusting their activities. Furthermore, collaborating with manufacturers to develop user-friendly interfaces and applications for these wearable devices can enhance usability and adoption among users. Field trials and longitudinal studies in various settings and populations are essential to validate the effectiveness and reliability of the predictive model in real-world scenarios, ensuring its practicality and impact on user safety and well-being.

**Literature Review and Research Field:** Continuous engagement with the literature and research field is paramount, especially as new studies and methodologies emerge. As an article incorporating

prediction information is being developed, it is essential to stay abreast of the latest advancements, insights, and trends in the field of drowsiness prediction and driver safety. This involves conducting regular literature reviews, attending relevant conferences and workshops, and engaging with fellow researchers and experts in the domain. By staying informed about current research directions, innovative approaches, and best practices, opportunities for collaboration, refinement of methodologies, and integration of novel insights into the predictive model can be identified. Additionally, contributing to the dissemination of findings through publications and presentations further enriches the collective knowledge base and fosters collaboration within the research community. Thus, ongoing engagement with the literature and research field ensures the relevance, rigor, and impact of the predictive model in addressing real-world challenges associated with drowsiness detection and driver safety.

# Bibliography


[1] E. Abe et al. “Development of drowsiness detection method by integrating heart rate variability analysis and multivariate statistical process control”. In: *SICE Journal of Control, Measurement, and System Integration* 9.1 (2016), pp. 10–17 (cit. on pp. 17, 21, 68).

[2] C. C. Aggarwal. *Data classification: Algorithms and applications.*, 2014. isbn: 9781466586758. doi: 10.1201/b17320 (cit. on p. 27).

[3] M. Aguiar et al. “Síndroma de apneia obstrutiva do sono como causa de acidentes de viação”. In: *Revista Portuguesa de Pneumologia* 15.3 (2009), pp. 419–431 (cit. on pp. 14, 54).

[4] M. Ahmed, A. N. Mahmood, and J. Hu. “A survey of network anomaly detection techniques”. In: *Journal of Network and Computer Applications* 60 (2016), pp. 19–31 (cit. on p. 43).

[5] T. Åkerstedt and M. Gillberg. “Subjective and objective sleepiness in the active individual”. In: *International journal of neuroscience* 52.1-2 (1990), pp. 29–37 (cit. on pp. 13, 60).

[6] C. F. Alcala and S. J. Qin. “Analysis and generalization of fault diagnosis methods for process monitoring”. In: *Journal of Process Control* 21.3 (2011), pp. 322–330 (cit. on p. 44).

[7] L. Ali et al. “Automated detection of Parkinson’s disease based on multiple types of sustained phonations using linear discriminant analysis and genetically optimized neural network”. In: *IEEE journal of translational engineering in health and medicine* 7 (2019), pp. 1–10 (cit. on p. 34).

[8] M. M. Ali et al. “Principles and recent advances in electronic nose for quality inspection of agricultural and food products”. In: *Trends in Food Science & Technology* 99 (2020), pp. 1–10 (cit. on p. 34).

[9] J. Allen. *Photoplethysmography and its application in clinical physiological measurement*. 2007. doi: 10.1088/0967-3334/28/3/R01 (cit. on p. 10).

[10] T. Almeida, C. Ramos, and J. Cardoso. “Sleep quality in the general Portuguese population”. In: *Annals of Medicine* 50 (2018), S144–S145 (cit. on pp. 2, 52).

[11] A. R. Antunes, A. C. Braga, and J. Gonçalves. “Drowsiness detection using multivariate statistical process control”. In: *International Conference on Computational Science and Its Applications*. Springer. 2022, pp. 571–585 (cit. on pp. 67, 96).

[12] A. R. Antunes, A. C. Braga, and J. Gonçalves. “Drowsiness transitions detection using a wearable device”. In: *Applied Sciences* 13.4 (2023), p. 2651 (cit. on pp. 67, 96).

[13] A. R. Antunes et al. “An Intelligent System to Detect Drowsiness at the Wheel”. In: *2022 10th International Symposium on Digital Forensics and Security (ISDFS)*. IEEE. 2022, pp. 1–6 (cit. on pp. 74, 96).

[14] A. R. Antunes et al. “Feature Selection Optimization for Heart Rate Variability”. In: *2023 11th International Symposium on Digital Forensics and Security (ISDFS)*. IEEE. 2023, pp. 1–6 (cit. on p. 96).

[15] A. R. Antunes et al. “Sleep Disorders in Portugal Based on Questionnaires”. In: *International Conference on Computational Science and Its Applications*. Springer. 2023, pp. 97–113 (cit. on p. 93).

[16] M. Attaran and P. Deb. “Machine learning: the new’big thing’for competitive advantage”. In: *International Journal of Knowledge Engineering and Data Mining* 5.4 (2018), pp. 277–305 (cit. on p. 25).

[17] J. B. Awotunde et al. “Disease diagnosis system for IoT-based wearable body sensors with machine learning algorithm”. In: *Hybrid Artificial Intelligence and IoT in Healthcare* (2021), pp. 201–222 (cit. on p. 39).

[18] R. Bakeman and B. F. Robinson. *Understanding statistics in the behavioral sciences*. Psychology Press, 2005. isbn: 0805849440. doi: 10.4324/9781410612625 (cit. on p. 27).

[19] V. P. Balam, V. U. Sameer, and S. Chinara. “Automated classification system for drowsiness detection using convolutional neural network and electroencephalogram”. In: *IET Intelligent Transport Systems* 15.4 (2021), pp. 514–524 (cit. on p. 8).

[20] L. Barr, S. Popkin, H. Howarth, et al. *An evaluation of emerging driver fatigue detection measures and technologies*. Tech. rep. United States. Department of Transportation. Federal Motor Carrier Safety ..., 2009 (cit. on p. 9).

[21] M. S. Bennett. *Heart Rate Variability*. A B.I.G. Publishing Project, 2020 (cit. on pp. 10, 11).

[22] P. Bhowmik et al. “Cardiotocography data analysis to predict fetal health risks with tree-based ensemble learning”. In: *Inf. Technol. Comput. Sci* 5 (2021), pp. 30–40 (cit. on p. 39).

[23] K. H. Bindu et al. *Coefficient of variation and machine learning applications*. CRC Press, 2019 (cit. on p. 71).

[24] E. Y. Boateng and D. A. Abaye. “A review of the logistic regression model with emphasis on medical research”. In: *Journal of data analysis and information processing* 7.4 (2019), pp. 190–207 (cit. on p. 33).

[25] P. Bounkeomany. “Speech major depression detection based on Adaboost-ELM algorithm”. In: *2020 IEEE International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA)*. Vol. 1. IEEE. 2020, pp. 858–862 (cit. on p. 41).

[26] C. M. Bower and A. Gungor. “Pediatric obstructive sleep apnea syndrome”. In: *Otolaryngologic Clinics of North America* 33.1 (2000), pp. 49–75 (cit. on p. 14).

[27] M. A. Boyacioglu, Y. Kara, and Ö. K. Baykan. “Predicting bank financial failures using neural networks, support vector machines and multivariate statistical methods: A comparative analysis in the sample of savings deposit insurance fund (SDIF) transferred banks in Turkey”. In: *Expert Systems with Applications* 36.2 (2009), pp. 3355–3366 (cit. on p. 33).

[28] V. Brodbeck et al. “EEG microstates of wakefulness and NREM sleep”. In: *NeuroImage* (2012). issn: 10538119. doi: 10.1016/j.neuroimage.2012.05.060 (cit. on p. 8).

[29] A. Burkov. *The hundred-page machine learning book*. Vol. 1. Andriy Burkov Quebec City, QC, Canada, 2019 (cit. on pp. 26, 37).

[30] A. Burlacu et al. “Accurate and early detection of sleepiness, fatigue and stress levels in drivers through Heart Rate Variability parameters: a systematic review”. In: *Reviews in cardiovascular medicine* 22.3 (2021), pp. 845–852 (cit. on p. 20).

[31] D. J. Buysse et al. “The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research”. In: *Psychiatry research* 28.2 (1989), pp. 193–213 (cit. on pp. 15, 59).

[32] M. Cascalho et al. “Hábitos e problemas de sono na população adulta Portuguesa”. In: *Revista Psicologia, Saúde & Doenças* 21 (2020), pp. 110–111 (cit. on p. 2).

[33] R. Catarino et al. “Sleepiness and sleep-disordered breathing in truck drivers: Risk analysis of road accidents”. In: *Sleep and Breathing* 18.1 (2014). issn: 15221709 (cit. on pp. 2, 52).

[34] J. Cech and T. Soukupova. “Real-time eye blink detection using facial landmarks”. In: *Cent. Mach. Perception, Dep. Cybern. Fac. Electr. Eng. Czech Tech. Univ. Prague* (2016), pp. 1–8 (cit. on p. 60).

[35] R. Champseix, L. Ribiere, and C. Le Couedic. “A Python Package for Heart Rate Variability Analysis and Signal Preprocessing.” In: *Journal of Open Research Software* 9.1 (2021) (cit. on p. 61).

[36] V. Chandola, A. Banerjee, and V. Kumar. “Anomaly detection: A survey”. In: *ACM computing surveys (CSUR)* 41.3 (2009), pp. 1–58 (cit. on p. 43).

[37] I. Chelminski et al. “Psychometric properties of the reduced Horne and Ostberg questionnaire”. In: *Personality and Individual Differences* 29.3 (2000), pp. 469–478 (cit. on p. 15).

[38] T. Chen and C. Guestrin. “Xgboost: A scalable tree boosting system”. In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 2016, pp. 785–794 (cit. on pp. 41, 42).

[39] K.-Y. Chen et al. “Predictors of health-related quality of life and influencing factors for COVID-19 patients, a follow-up at one month”. In: *Frontiers in Psychiatry* 11 (2020), p. 668 (cit. on p. 33).

[40] Y. Chen, Q. Zhao, and L. Lu. “Combining the outputs of various k-nearest neighbor anomaly detectors to form a robust ensemble model for high-dimensional geochemical anomaly detection”. In: *Journal of Geochemical exploration* 231 (2021), p. 106875 (cit. on p. 38).

[41] D. Chicco, M. J. Warrens, and G. Jurman. “The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation”. In: *Peerj computer science* 7 (2021), e623 (cit. on p. 29).

[42] A. Choi and H. Shin. “Quantitative analysis of the effect of an ectopic beat on the heart rate variability in the resting condition”. In: *Frontiers in physiology* 9 (2018), p. 922 (cit. on p. 61).

[43] M. Choi et al. “Wearable device-based system to monitor a driver’s stress, fatigue, and drowsiness”. In: *IEEE Transactions on Instrumentation and Measurement* 67.3 (2017), pp. 634–645 (cit. on pp. 17, 18, 21).

[44] S. Chokroverty. “Overview of sleep & sleep disorders”. In: *Indian Journal of Medical Research* 131.2 (2010), pp. 126–140 (cit. on p. 14).

[45] F. Chung, H. R. Abdullah, and P. Liao. “STOP-Bang questionnaire: a practical approach to screen for obstructive sleep apnea”. In: *Chest* 149.3 (2016), pp. 631–638 (cit. on pp. 14, 59).

[46] H. R. Colten and B. M. Altevogt. *Sleep disorders and sleep deprivation: An unmet public health problem*. THE NATIONAL ACADEMIES PRESS, 2006. isbn: 0309101115 (cit. on p. 1).

[47] P. Dangeti. *Statistics for machine learning*. Packt Publishing Ltd, 2017 (cit. on p. 24).

[48] J. Daniel. *Sampling essentials: Practical guidelines for making sampling choices*. Sage Publications, 2011 (cit. on p. 47).

[49] V. Demareva et al. “Home-Based Dynamics of Sleepiness-Related Conditions Starting at Biological Evening and Later (Beyond Working)”. In: *International Journal of Environmental Research and Public Health* 20.17 (2023), p. 6641 (cit. on p. 82).

[50] J. M. Dos Santos. “Abordagem do doente com patologia do sono”. In: *Revista Portuguesa de Medicina Geral e Familiar* 22.5 (2006), pp. 599–610 (cit. on p. 14).

[51] P. Dubey et al. “Entrepreneurial marketing: an analytical viewpoint on perceived quality and customer delight”. In: *Journal of Research in Marketing and Entrepreneurship* 22.1 (2020), pp. 1–19 (cit. on p. 34).

[52] J. D. Edinger et al. “Derivation of research diagnostic criteria for insomnia: report of an American Academy of Sleep Medicine Work Group”. In: *Sleep* 27.8 (2004), pp. 1567–1596 (cit. on p. 15).

[53] I. El Naqa and M. J. Murphy. “What is machine learning?, in machine learning in radiation oncology”. In: *machine learning in radiation oncology, Cham: Springer* (2015), pp. 3–11 (cit. on p. 24).

[54] B. J. Erickson and F. Kitamura. *Magician’s corner: 9. Performance metrics for machine learning models*. 2021 (cit. on pp. 27, 28).

[55] S. Eye. *Smart Eye Pro – remote eye tracking system*. 2023-05. url: `https://smarteye.se/smart-eye-pro/` (cit. on p. 60).

[56] A. Falavigna et al. “Consistency and reliability of the Brazilian Portuguese version of the Mini-Sleep Questionnaire in undergraduate students”. In: *Sleep and Breathing* 15 (2011), pp. 351–355 (cit. on p. 15).

[57] X. Fan, B. C. Yin, and Y. F. Sun. “Yawning detection based on Gabor wavelets and LDA”. In: *Beijing Gongye Daxue Xuebao / Journal of Beijing University of Technology* (2009). issn: 02540037 (cit. on p. 9).

[58] Z.-g. Fang et al. “Application of a data-driven XGBoost model for the prediction of COVID-19 in the USA: a time-series study”. In: *BMJ open* 12.7 (2022), e056685 (cit. on p. 42).

[59] A. Ferrari and M. Russo. *Introducing Microsoft Power BI*. Microsoft Press, 2016 (cit. on pp. 55, 57).

[60] A. Ferrer. “Multivariate statistical process control based on principal component analysis (MSPC-PCA): some reflections and a case study in an autobody assembly process”. In: *Quality Engineering* 19.4 (2007), pp. 311–325 (cit. on pp. 43, 44).

[61] I. Filip et al. “Public health burden of sleep disorders: underreported problem”. In: *Journal of Public Health* 25 (2017), pp. 243–248 (cit. on p. 52).

[62] M. J. Fisher, A. P. Marshall, and M. Mitchell. “Testing differences in proportions”. In: *Australian Critical Care* 24.2 (2011), pp. 133–138 (cit. on p. 52).

[63] F. Forcolin et al. “Comparison of outlier heartbeat identification and spectral transformation strategies for deriving heart rate variability indices for drivers at different stages of sleepiness”. In: *Traffic injury prevention* 19.sup1 (2018), S112–S119 (cit. on p. 10).

[64] Y. Freund and R. E. Schapire. “A decision-theoretic generalization of on-line learning and an application to boosting”. In: *Journal of computer and system sciences* 55.1 (1997), pp. 119–139 (cit. on p. 40).

[65] G. D. Furman et al. “Early detection of falling asleep at the wheel: A heart rate variability approach”. In: *2008 Computers in Cardiology*. IEEE. 2008, pp. 1109–1112 (cit. on p. 10).

[66] P. Geurts, D. Ernst, and L. Wehenkel. “Extremely randomized trees”. In: *Machine learning* 63 (2006), pp. 3–42 (cit. on p. 39).

[67] M. Gonçalves et al. “Sleepiness and motor vehicle crashes in a representative sample of Portuguese drivers: the importance of epidemiological representative surveys”. In: *Traffic injury prevention* 16.7 (2015), pp. 677–683 (cit. on pp. 2, 52).

[68] M. Gonçalves et al. “Sleepiness at the wheel across Europe: a survey of 19 countries”. In: *Journal of sleep research* 24.3 (2015), pp. 242–253 (cit. on pp. 3, 52).

[69] W. Grossmann and S. Rinderle-Ma. “Fundamentals of business intelligence”. In: (2015) (cit. on p. 55).

[70] J. F. Hair et al. “Research methods for business”. In: *Education+ Training* 49.4 (2007), pp. 336–337 (cit. on p. 5).

[71] A. J. E. Hangouche et al. “Relationship between poor quality sleep, excessive daytime sleepiness and low academic performance in medical students”. In: *Advances in medical education and practice* (2018), pp. 631–638 (cit. on p. 53).

[72] T. Hastie et al. *The elements of statistical learning: data mining, inference, and prediction*. Vol. 2. Springer, 2009 (cit. on p. 23).

[73] J. Hatwell, M. M. Gaber, and R. M. Atif Azad. “Ada-WHIPS: explaining AdaBoost classification with applications in the health sciences”. In: *BMC Medical Informatics and Decision Making* 20.1 (2020), pp. 1–25 (cit. on p. 41).

[74] T. Hori et al. “Proposed supplements and amendments to ‘A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects’, the Rechtschaffen & Kales (1968) standard”. In: *Psychiatry and clinical neurosciences* 55.3 (2001), pp. 305–310 (cit. on p. 8).

[75] J. A. Horne and O. Östberg. “Individual differences in human circadian rhythms”. In: *Biological psychology* 5.3 (1977), pp. 179–190 (cit. on pp. 15, 59).

[76] D. W. Hosmer. “Applied logistic regression, 2”. In: (2000) (cit. on p. 33).

[77] S. Huang et al. “Applications of support vector machine (SVM) learning in cancer genomics”. In: *Cancer genomics & proteomics* 15.1 (2018), pp. 41–51 (cit. on p. 32).

[78] S. Hwang et al. “Feasibility analysis of heart rate monitoring of construction workers using a photoplethysmography (PPG) sensor embedded in a wristband-type activity tracker”. In: *Automation in construction* 71 (2016), pp. 372–381 (cit. on pp. 17, 21).

[79] M. Ingre et al. “Subjective sleepiness, simulated driving performance and blink duration: examining individual differences”. In: *Journal of sleep research* 15.1 (2006), pp. 47–53 (cit. on p. 9).

[80] J. E. Jackson and G. S. Mudholkar. “Control procedures for residuals associated with principal component analysis”. In: *Technometrics* 21.3 (1979), pp. 341–349 (cit. on p. 44).

[81] N. Japkowicz and M. Shah. *Evaluating learning algorithms: A classification perspective*. Cambridge University Press, 2011. isbn: 9780511921803. doi: 10.1017/CBO9780511921803 (cit. on p. 28).

[82] J. Jeppesen et al. “Using Lorenz plot and Cardiac Sympathetic Index of heart rate variability for detecting seizures for patients with epilepsy”. In: *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE. 2014, pp. 4563–4566 (cit. on p. 12).

[83] K. A. D. R. João et al. “Validation of the Portuguese version of the Pittsburgh sleep quality index (PSQI-PT)”. In: *Psychiatry research* 247 (2017), pp. 225–229 (cit. on p. 15).

[84] M. W. Johns. “A new method for measuring daytime sleepiness: the Epworth sleepiness scale”. In: *sleep* 14.6 (1991), pp. 540–545 (cit. on pp. 14, 59).

[85] N. Kalcheva, M. Todorova, and G. Marinova. “Naive Bayes Classifier, Decision Tree and AdaBoost Ensemble Algorithm–Advantages and Disadvantages”. In: *Proceedings of the 6th ERAZ Conference Proceedings (part of ERAZ conference collection), Online*. 2020, pp. 153–157 (cit. on p. 40).

[86] R. N. Khushaba et al. “Driver drowsiness classification using fuzzy wavelet-packet-based feature-extraction algorithm”. In: *IEEE Transactions on Biomedical Engineering* (2011). issn: 00189294. doi: 10.1109/TBME.2010.2077291 (cit. on p. 10).

[87] G. King, M. Tomz, and J. Wittenberg. “Making the most of statistical analyses: Improving interpretation and presentation”. In: *American journal of political science* (2000), pp. 347–361 (cit. on p. 33).

[88] R. E. Kleiger et al. “Time domain measurements of heart rate variability”. In: *Cardiology clinics* 10.3 (1992), pp. 487–498 (cit. on p. 10).

[89] J. M. Kortelainen et al. “Sleep staging based on signals acquired through bed sensor”. In: *IEEE Transactions on Information Technology in Biomedicine* 14.3 (2010). issn: 10897771 (cit. on p. 1).

[90] A. D. Krystal and J. D. Edinger. “Measuring sleep quality”. In: *Sleep medicine* 9 (2008), S10–S17 (cit. on p. 15).

[91] M. Kück and M. Freitag. “Forecasting of customer demands for production planning by local k-nearest neighbor models”. In: *International Journal of Production Economics* 231 (2021), p. 107837 (cit. on p. 38).

[92] T. Kundinger, N. Sofra, and A. Riener. “Assessment of the potential of wrist-worn wearable sensors for driver drowsiness detection”. In: *Sensors* 20.4 (2020), p. 1029 (cit. on pp. 19, 21).

[93] M. B. Kurt et al. “The ANN-based computing of drowsy level”. In: *Expert Systems with Applications* (2009). issn: 09574174. doi: 10.1016/j.eswa.2008.01.085 (cit. on p. 10).

[94] Y. Lacasse, M. Bureau, and F. Series. “A new standardised and self-administered quality of life questionnaire specific to obstructive sleep apnoea”. In: *Thorax* 59.6 (2004), pp. 494–499 (cit. on p. 15).

[95] S. Lakshmanaprabu et al. “Optimal deep learning model for classification of lung cancer on CT images”. In: *Future Generation Computer Systems* 92 (2019), pp. 374–382 (cit. on p. 34).

[96] B.-G. Lee, B.-L. Lee, and W.-Y. Chung. “Smartwatch-based driver alertness monitoring with wearable motion and physiological sensor”. In: *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2015, pp. 6126–6129 (cit. on pp. 16, 21).

[97] H. Lee, J. Lee, and M. Shin. “Using wearable ECG/PPG sensors for driver drowsiness detection based on distinguishable pattern of recurrence plots”. In: *Electronics* 8.2 (2019), p. 192 (cit. on pp. 18, 21).

[98] W. C. Liang et al. “Changes in physiological parameters induced by indoor simulated driving: Effect of lower body exercise at mid-term break”. In: *Sensors* 9.9 (2009), pp. 6913–6933 (cit. on p. 10).

[99] W. Lin et al. “Multiclass diagnosis of stages of Alzheimer’s disease using linear discriminant analysis scoring for multimodal data”. In: *Computers in Biology and Medicine* 134 (2021), p. 104478 (cit. on p. 34).

[100] C. C. Liu, S. G. Hosking, and M. G. Lenné. “Predicting driver drowsiness using vehicle measures: Recent insights and future challenges”. In: *Journal of Safety Research* (2009). issn: 00224375. doi: 10.1016/j.jsr.2009.04.005 (cit. on p. 9).

[101] L. López and J. Sánchez. “Discriminant methods for radar detection of hail”. In: *Atmospheric Research* 93.1-3 (2009), pp. 358–368 (cit. on p. 33).

[102] D. Lopez-Bernal et al. “Education 4.0: teaching the basics of KNN, LDA and simple perceptron algorithms for binary classification problems”. In: *Future Internet* 13.8 (2021), p. 193 (cit. on p. 38).

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

[104] M. Malik. “Geometrical methods for heart rate variability assessment”. In: *Heart rate variability* (1995) (cit. on p. 61).

[105] S. Malik, R. Harode, and A. Kunwar. “XGBoost: A deep dive into boosting”. In: *no. February* (2020) (cit. on p. 42).

[106] A. M. Martinez and A. C. Kak. “Pca versus lda”. In: *IEEE transactions on pattern analysis and machine intelligence* 23.2 (2001), pp. 228–233 (cit. on p. 34).

[107] J. Mateo et al. “Extreme gradient boosting machine learning method for predicting medical treatment in patients with acute bronchiolitis”. In: *Biocybernetics and Biomedical Engineering* 41.2 (2021), pp. 792–801 (cit. on p. 42).

[108] C. McCarthy et al. “Validation of the Empatica E4 wristband”. In: *2016 IEEE EMBS international student conference (ISC)*. IEEE. 2016, pp. 1–4 (cit. on p. 61).

[109] M. V. Meneses, A. R. Antunes, and J. Gonçalves. “Driver Drowsiness Classification using Machine Learning and Heart Rate Variability”. In: *2023 11th International Symposium on Digital Forensics and Security (ISDFS)*. IEEE. 2023, pp. 1–6 (cit. on p. 96).

[110] E. Michail et al. “EEG and HRV markers of sleepiness and loss of control during car driving”. In: *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE. 2008, pp. 2566–2569 (cit. on p. 10).

[111] A. Mittal et al. “Head movement-based driver drowsiness detection: A review of state-of-art techniques”. In: *2016 IEEE international conference on engineering and technology (ICETECH)*. IEEE. 2016, pp. 903–908 (cit. on p. 13).

[112] D. C. Montgomery and G. C. Runger. *Applied statistics and probability for engineers*. John wiley & sons, 2010 (cit. on pp. 24, 33, 35, 47, 79).

[113] C. L. Morais et al. “Tutorial: multivariate classification for vibrational spectroscopy in biological samples”. In: *Nature Protocols* 15.7 (2020), pp. 2143–2162 (cit. on p. 34).

[114] A. A. Mostafa et al. “GBO-kNN a new framework for enhancing the performance of ligand-based virtual screening for drug discovery”. In: *Expert Systems with Applications* 197 (2022), p. 116723 (cit. on p. 38).

[115] K. P. Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012 (cit. on p. 24).

[116] I. B. Mustapha et al. “Effective Email Spam Detection System using Extreme Gradient Boosting”. In: *arXiv preprint arXiv:2012.14430* (2020) (cit. on p. 42).

[117] H. Mydyti. “Data Mining Approach Improving Decision-Making Competency along the Business Digital Transformation Journey: A Case Study–Home Appliances after Sales Service”. In: *Seeu Review* 16.1 (2021), pp. 45–65 (cit. on p. 33).

[118] N. C. Netzer et al. “Using the Berlin Questionnaire to identify patients at risk for the sleep apnea syndrome”. In: *Annals of internal medicine* 131.7 (1999), pp. 485–491 (cit. on p. 14).

[119] M. T. Nwakuya and O. Mmaduka. “Ordered logistic regression on the mental health of undergraduate students”. In: *International Journal of Probability and Statistics* 8.1 (2019), pp. 14–18 (cit. on p. 33).

[120] J. F. O’Hanlon. *Heart rate variability: a new index of driver alertness fatigue*. Tech. rep. SAE Technical Paper, 1972 (cit. on p. 16).

[121] N. Orangi-Fard, A. Akhbardeh, and H. Sagreiya. “Predictive model for icu readmission based on discharge summaries using machine learning and natural language processing”. In: *Informatics*. Vol. 9. 1. MDPI. 2022, p. 10 (cit. on p. 41).

[122] S. P. et al. “Testing a system for predicting microsleep”. In: *Lekar a Technika* (2016). issn: 2336-5552 (cit. on p. 3).

[123] J. Pagel. “Excessive daytime sleepiness”. In: *American family physician* 79.5 (2009), pp. 391–396 (cit. on p. 14).

[124] S.-T. Pan et al. “A transition-constrained discrete hidden Markov model for automatic sleep staging”. In: *Biomedical engineering online* 11.1 (2012), pp. 1–19 (cit. on p. 10).

[125] A. Patel et al. “Extra-Tree Classification of Customer Review for Hotel Recommendation”. In: *2023 4th International Conference on Electronics and Sustainable Communication Systems (ICESC)*. IEEE. 2023, pp. 1035–1041 (cit. on p. 39).

[126] M. L. Patten. *Understanding research methods: An overview of the essentials*. Routledge, 2016 (cit. on pp. 4, 5).

[127] M. K. Pavlova and V. Latreille. “Sleep disorders”. In: *The American journal of medicine* 132.3 (2019), pp. 292–299 (cit. on p. 13).

[128] V. Phanikrishna B and S. chinara. “Time domain parameters as a feature for single-channel EEG-based drowsiness detection method”. In: *2020 IEEE International Students’ Conference on Electrical, Electronics and Computer Science (SCEECS)*. IEEE. 2020, pp. 1–5 (cit. on p. 8).

[129] D. A. Pisner and D. M. Schnyer. “Support vector machine”. In: *Machine learning*. Elsevier, 2020, pp. 101–121 (cit. on p. 30).

[130] S. R. Polamuri, K. Srinivas, and A. K. Mohan. “Stock market prices prediction using random forest and extra tree regression”. In: *Int. J. Recent Technol. Eng* 8.1 (2019), pp. 1224–1228 (cit. on p. 39).

[131] N. Rahmayanti et al. “Comparison of machine learning algorithms to classify fetal health using cardiotocogram data”. In: *Procedia Computer Science* 197 (2022), pp. 162–171 (cit. on p. 42).

[132] M. Ramzan et al. “A survey on state-of-the-art drowsiness detection techniques”. In: *IEEE Access* 7 (2019), pp. 61904–61919 (cit. on p. 9).

[133] C. Randler et al. “Morningness–eveningness and amplitude–development and validation of an improved composite scale to measure circadian preference and stability (MESSi)”. In: *Chronobiology international* 33.7 (2016), pp. 832–848 (cit. on p. 16).

[134] G. Regalia et al. “Multimodal wrist-worn devices for seizure detection and advancing research: focus on the Empatica wristbands”. In: *Epilepsy research* 153 (2019), pp. 79–82 (cit. on p. 71).

[135] R. Reis et al. “Validation of a Portuguese version of the STOP-Bang questionnaire as a screening tool for obstructive sleep apnea: Analysis in a sleep clinic”. In: *Revista Portuguesa de Pneumologia (English Edition)* 21.2 (2015), pp. 61–68 (cit. on p. 14).

[136] Y. Rejani and S. T. Selvi. “Early detection of breast cancer using SVM classifier technique”. In: *arXiv preprint arXiv:0912.2314* (2009) (cit. on p. 32).

[137] P. F. Rodrigues et al. “Initial psychometric characterization for the Portuguese version of the Morningness-Eveningness-Stability-Scale improved (MESSi)”. In: *Chronobiology international* 35.11 (2018), pp. 1608–1618 (cit. on p. 16).

[138] L. Rosenthal, T. A. Roehrs, and T. Roth. “The sleep-wake activity inventory: a self-report measure of daytime sleepiness”. In: *Biological Psychiatry* 34.11 (1993), pp. 810–820 (cit. on p. 14).

[139] J. V. Rundo and R. Downey. “Polysomnography”. In: *Handbook of Clinical Neurology*. Vol. 160. 2019. doi: 10.1016/B978-0-444-64032-1.00025-4 (cit. on p. 10).

[140] U. Saeed et al. “Fault diagnosis based on extremely randomized trees in wireless sensor networks”. In: *Reliability engineering & system safety* 205 (2021), p. 107284 (cit. on p. 39).

[141] A. Sahayadhas, K. Sundaraj, and M. Murugappan. *Detecting driver drowsiness based on sensors: A review*. 2012. doi: 10.3390/s121216937 (cit. on pp. 8, 9, 13).

[142] N. J. Salkind. *Statistics for people who (think they) hate statistics : Excel 2007 edition*. SAGE Publications, Inc., 2010 (cit. on pp. 27, 29).

[143] N. J. Salkind and B. B. Frey. *Statistics for people who (think they) hate statistics*. Sage Publications, 2019 (cit. on p. 28).

[144] A. F. Salles et al. “Detecção automática de sonolência em condutores de veículos utilizando redes neurais artificiais”. In: (2018) (cit. on p. 9).

[145] A. I. Samouco et al. *A Ciência do Sono - Da Fisiologia à (Psico)patologia*. 1st ed. PARSIFAL, 2020, p. 240. isbn: 978-989-8760-76-0 (cit. on pp. 1, 9).

[146] R. Sampaio, M. G. Pereira, and J. C. Winck. “Adaptação portuguesa do questionário de qualidade de vida (SAQLI) nos doentes com síndrome de apneia obstrutiva do sono”. In: *Revista Portuguesa de Pneumologia* 18.4 (2012), pp. 166–174 (cit. on p. 15).

[147] M. Saygın et al. “Investigation of sleep quality and sleep disorders in students of medicine”. In: *Turkish thoracic journal* 17.4 (2016), p. 132 (cit. on p. 13).

[148] R. E. Schapire. “Explaining adaboost”. In: *Empirical Inference: Festschrift in Honor of Vladimir N. Vapnik*. Springer, 2013, pp. 37–52 (cit. on p. 40).

[149] E. M. Senan and M. E. Jadhav. “Diagnosis of dermoscopy images for the detection of skin lesions using SVM and KNN”. In: *Proceedings of Third International Conference on Sustainable Computing: SUSCOM 2021*. Springer. 2022, pp. 125–134 (cit. on p. 38).

[150] F. Shaffer and J. Ginsberg. “An overview of heart rate variability metrics and norms”. In: *Frontiers in public health* 5 (2017), p. 258 (cit. on pp. 11, 12).

[151] A. Sharaff and H. Gupta. “Extra-tree classifier with metaheuristics approach for email classification”. In: *Advances in Computer Communication and Computational Sciences: Proceedings of IC4S 2018*. Springer. 2019, pp. 189–197 (cit. on p. 39).

[152] R. Sharda, D. Delen, and E. Turban. *Analytics, data science, & artificial intelligence: Systems for decision support*. Pearson Harlow, 2021 (cit. on pp. 26, 29–31, 35, 38).

[153] J. Shawe-Taylor and N. Cristianini. *Support vector machines*. Vol. 2. Cambridge university press Cambridge, 2000 (cit. on p. 32).

[154] Z. Shinar et al. “Autonomic changes during wake–sleep transition: A heart rate variability based approach”. In: *Autonomic Neuroscience* 130.1-2 (2006), pp. 17–27 (cit. on p. 11).

[155] C. F. Silva et al. “The Portuguese version of the Horne and Ostberg morningness-eveningness questionnaire: Its role in education and psychology”. In: *Revista Psicologia e Educação* (2002) (cit. on p. 16).

[156] N. Silver. *The signal and the noise: Why so many predictions fail-but some don't*. Penguin, 2012 (cit. on p. 24).

[157] R. H. Singh et al. “Movie recommendation system using cosine similarity and KNN”. In: *International Journal of Engineering and Advanced Technology* 9.5 (2020), pp. 556–559 (cit. on p. 38).

[158] M. Sokolova, N. Japkowicz, and S. Szpakowicz. “Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation”. In: *Australasian joint conference on artificial intelligence*. Springer. 2006, pp. 1015–1021 (cit. on p. 28).

[159] T. W. Strine and D. P. Chapman. “Associations of frequent sleep insufficiency with health-related quality of life and health behaviors”. In: *Sleep medicine* 6.1 (2005), pp. 23–27 (cit. on pp. 13, 14, 60).

[160] M. Sumathi and S. Raja. “Machine learning algorithm-based spam detection in social networks”. In: *Social Network Analysis and Mining* 13.1 (2023), p. 104 (cit. on p. 42).

[161] A. Tharwat et al. “Linear discriminant analysis: A detailed tutorial”. In: *AI communications* 30.2 (2017), pp. 169–190 (cit. on p. 34).

[162] M. Toğaçar, B. Ergen, and Z. Cömert. “Application of breast cancer diagnosis based on a combination of convolutional neural networks, ridge regression and linear discriminant analysis using invasive breast cancer images processed with autoencoders”. In: *Medical hypotheses* 135 (2020), p. 109503 (cit. on p. 34).

[163] M. Toichi et al. “A new method of assessing cardiac autonomic function and its comparison with spectral analysis and coefficient of variation of R–R interval”. In: *Journal of the autonomic nervous system* 62.1-2 (1997), pp. 79–84 (cit. on p. 12).

[164] S. Uddin et al. “Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction”. In: *Scientific Reports* 12.1 (2022), p. 6256 (cit. on p. 38).

[165] A. Vaz et al. “Tradução do Questionário de Berlim para língua Portuguesa e sua aplicação na identificação da SAOS numa consulta de patologia respiratória do sono”. In: *Revista portuguesa de pneumologia* 17.2 (2011), pp. 59–65 (cit. on p. 15).

[166] J. Vicente et al. “Drowsiness detection using heart rate variability”. In: *Medical & biological engineering & computing* 54 (2016), pp. 927–937 (cit. on p. 11).

[167] M. Walker. *Why we sleep: unlocking the power of sleep and dreams*. 2017 (cit. on pp. 2, 8, 9).

[168] Y. Wang and Y. Guo. “Forecasting method of stock market volatility in time series data based on mixed model of ARIMA and XGBoost”. In: *China Communications* 17.3 (2020), pp. 205–221 (cit. on p. 42).

[169] I. Wickramasinghe and H. Kalutarage. “Naive Bayes: applications, variations and vulnerabilities: a review of literature with code snippets for implementation”. In: *Soft Computing* 25.3 (2021), pp. 2277–2293 (cit. on p. 37).

[170] C. J. Willmott, K. Matsuura, and S. M. Robeson. “Ambiguities inherent in sums-of-squares-based error statistics”. In: *Atmospheric Environment* 43.3 (2009), pp. 749–752 (cit. on p. 29).

[171] Y. Wu et al. “Novel binary logistic regression model based on feature transformation of XGBoost for type 2 Diabetes Mellitus prediction in healthcare systems”. In: *Future Generation Computer Systems* 129 (2022), pp. 1–12 (cit. on p. 42).

[172] H. Yi, K. Shin, and C. Shin. “Development of the sleep quality scale”. In: *Journal of sleep research* 15.3 (2006), pp. 309–316 (cit. on p. 15).

[173] B. C. Yin, X. Fan, and Y. F. Sun. “Multiscale dynamic features based driver fatigue detection”. In: *International Journal of Pattern Recognition and Artificial Intelligence* (2009). issn: 02180014. doi: 10.1142/S021800140900720X (cit. on p. 9).

[174] I. N. Yulita et al. “Adaboost support vector machine method for human activity recognition”. In: *2021 International Conference on Artificial Intelligence and Big Data Analytics*. IEEE. 2021, pp. 1–4 (cit. on p. 41).

[175] P. C. Zee, H. Attarian, and A. Videnovic. “Circadian rhythm abnormalities”. In: *Continuum: Lifelong Learning in Neurology* 19.1 Sleep Disorders (2013), p. 132 (cit. on pp. 2, 53).

[176] L. Zhang et al. “Time series forecast of sales volume based on XGBoost”. In: *Journal of Physics: Conference Series*. Vol. 1873. 1. IOP Publishing. 2021, p. 012067 (cit. on p. 42).

[177] T. Zhang, S. Moro, and R. F. Ramos. “A data-driven approach to improve customer churn prediction based on telecom customer segmentation”. In: *Future Internet* 14.3 (2022), p. 94 (cit. on p. 34).

[178] A. Zheng and A. Casari. *Feature engineering for machine learning: principles and techniques for data scientists.* " O'Reilly Media, Inc.", 2018 (cit. on pp. 26, 27).

[179] F. Zhou et al. "Driver fatigue transition prediction in highly automated driving using physiological features". In: *Expert Systems with Applications* 147 (2020), p. 113204 (cit. on pp. 19–21).

[180] J. ZOMER. "Mini Sleep Questionnaire (MSQ) for screening large populations for EDS complaints". In: *Sleep'84* (1985) (cit. on p. 15).

# A

# Consent Form

## CONSENTIMENTO INFORMADO, LIVRE E ESCLARECIDO PARA PARTICIPAÇÃO EM INVESTIGAÇÃO

Conforme a lei 67/98 de 26 de Outubro e a "Declaração de Helsínquia" da Associação Médica Mundial (Helsínquia 1964; Tóquio 1975; Veneza 1983; Hong Kong 1989; Somerset West 1996, Edimburgo 2000; Washington 2002, Tóquio 2004, Seul 2008)

**Designação do Estudo**: Sono ao Volante 2.0 - Sistema de Informação para a previsão do sono ao volante e a deteção de distúrbio ou privação crónica do sono. Projeto nº 039720 (PT2020 Projetos de I&D em copromoção SI-47-2027-30)

**Consórcio**: Optimizer - Serviços e Consultoria Informática Lda, Laboratório de Inteligência Artificial e Ciência de Computadores (LIACC) da Faculdade de Engenharia da Universidade do Porto, Instituto Politécnico do Cávado e Ave (IPCA) e Instituto do Sono (ISCCI) - Centro Clínico e Investigação.

Fui informado(a) de que o Estudo de Investigação acima mencionado se destina a obter dados para previsão e deteção de distúrbios ou privação de sono, com o objetivo de avaliar a capacidade e agilidade na tomada de decisão e medir o impacto nas atividades diárias, pessoais e profissionais.

Para o efeito irá conduzir um veículo em simulação de computador, recorrendo a uma interface de volante e pedais, no decorrer de uma sessão com a duração de uma hora e dez. Enquanto realiza esta ação, serão recolhidos dados biométricos da atividade cerebral e cardíaca com os sensores Brainlink/NeuroSky, E4 Wristband e Microsoft Band 2, respetivamente.

Responderá ainda a cinco questionários que têm como objetivo avaliar a qualidade de sono de Pittsburgh, a escala de sonolência de Epworth, identificar o risco de síndrome de apneia obstrutiva de sono (STOP-Bang), o ritmo circadiano (morningness-eveningness) e, por último, referente às considerações finais sobre a simulação realizada.

Foi-me garantido que todos os dados relativos à identificação dos Participantes neste estudo são confidenciais e que será mantido o anonimato. Foi-me garantido que a captura de vídeo efetuada durante a experiência não será divulgada a terceiros e será destruída no final do estudo.

Fui ainda informado(a) que neste estudo está prevista a realização de testes num simulador e preenchimento de questionários tendo-me sido explicado em que consistem e quais os seus possíveis efeitos. Sei que posso recusar participar ou interromper a qualquer momento a participação no estudo, sem nenhum tipo de penalização.

Fui informado(a) de que não está contemplado qualquer ressarcimento ou remuneração para participação no estudo assim como não apresenta qualquer custo para os participantes.

Compreendi a informação que me foi dada, tive oportunidade de fazer perguntas e as minhas dúvidas foram esclarecidas.

# B

# Code for the Multivariate Statistical Process Control

The appendix presents the code used for multivariate statistical process control. It begins with the implementation of principal components in time, frequency, and non-linear domains. This is followed by the application of Hotelling's $T^2$ and SPE statistics, along with the corresponding control limits.

```
import numpy as np
import pandas as pd
from scipy.stats import chi2, f
from pca import pca
import matplotlib.pyplot as plt
import os

def principal_components(data):
    # normalize data
    mean = data.mean()
    std = data.std()
    data = (data-mean)/std

    # principal component for time features
    model_time = pca(n_components = 1)
    # fit transform
    scores_time = model_time.fit_transform(data.iloc[:,0:16])
    time_scores = scores_time['PC']
    loading_time = scores_time['loadings']

    # principal component for frequency features
```

```
    model_freq = pca(n_components = 1)
    # fit transform
    scores_freq = model_freq.fit_transform(data.iloc[:,16:21])
    freq_scores = scores_freq['PC']
    freq_scores = freq_scores.rename(columns={'PC1': 'PC2'})
    loading_freq = scores_freq['loadings']

    # principal component for  nonlinear features
    model_nonlinear = pca(n_components = 1)
    # fit transform
    scores_nonlinear = model_nonlinear.fit_transform(data.iloc[:,21:25])
    nonlinear_scores = scores_nonlinear['PC']
    nonlinear_scores = nonlinear_scores.rename(columns={'PC1': 'PC3'})
    loading_nonlinear = scores_nonlinear['loadings']

    # group all components
    component_scores = pd.concat([ time_scores,
     freq_scores, nonlinear_scores], axis =1)
    return
    data, component_scores, loading_time,
    loading_freq, loading_nonlinear

def hotelling_t2_individ(scores):
# compute the Hotelling T^2 statistic
    var = scores.std()
    hotelling = (scores**2)/(var**2 )
    return hotelling

def hotelling_limit_individ(scores,level_confidence):
# identify the limit control for the Hotelling T^2 statistic
    k = 1
    n = len(scores)
    limit_control_t2 = (k*(n+1)*(n-1))/(n*(n-k)) *
    f.ppf(level_confidence, k, n-k)
    return limit_control_t2

def q_statistic(input_features,loadings,scores):
# compute the SPE statistic
```

```
    estimation_x = np.dot(scores,loadings)
    error = input_features - estimation_x
    q_statistic = np.sum(error**2, axis=1)
    q_statistic = pd.DataFrame(q_statistic, columns = ['Q'])
    return q_statistic

def q_control_limits(df_statistic_q, level_confidance):
# identify the limit control for the SPE statistic
    df = (2*df_statistic_q.mean()[0]**2)/df_statistic_q.var()[0]
    limit_control_q2 =
    (df_statistic_q.var()[0]/(2*df_statistic_q.mean()[0]))*
    chi2.ppf(level_confidance,df)
    return limit_control_q2
```

# C

# Machine Learning Grid Search

This appendix includes the code used for the grid search applied to each implemented machine learning model.

```
from sklearn.model_selection import GridSearchCV
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import ExtraTreesClassifier, AdaBoostClassifier
from xgboost import XGBClassifier
from sklearn.neighbors import KNeighborsClassifier
from sklearn.svm import SVC
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis

# Combined parameter grid for selected classifiers
param_grid_combined = {
    'LogisticRegression': {
        'penalty': ['l2', 'l1', 'elasticnet'],
        'solver': ['lbfgs', 'liblinear', 'newton-cg', 'newton-cholesky', 'sag', 'saga
        'multi_class': ['auto', 'ovr', 'multinomial']
    },
    'ExtraTreesClassifier': {
        'n_estimators': [100, 200, 500],
        'min_samples_split': [2, 3, 4, 5],
        'max_features': [6, 10, 15, 30],
        'max_depth': [5, 10, 20]
    },
    'XGBClassifier': {
        'max_depth': range(2, 10, 1),
        'n_estimators': range(60, 220, 40),
        'learning_rate': [0.1, 0.01, 0.05],
```

```
        'eval_metric': ['rmse', 'logloss', 'error', 'auc'],
        'sampling_method': ['uniform', 'gradient_based']
    },
    'AdaBoostClassifier': {
        'n_estimators': [10, 20, 30, 40, 50, 60, 80, 90, 100],
        'learning_rate': [0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1],
        'algorithm': ['SAMME', 'SAMME.R']
    },
    'KNeighborsClassifier': {
        'n_neighbors': [1, 5, 10],
        'leaf_size': [5, 10, 20],
        'algorithm': ['ball_tree', 'brute', 'auto'],
        'weights': ['uniform', 'distance'],
        'p': [1, 2, 3]
    },
    'SVC': {
        'degree': [3, 5, 10],
        'gamma': ['scale', 'auto'],
        'max_iter': [100, 200, 500],
        'C': [0.01, 0.5, 1, 1.5, 2],
        'kernel': ['linear', 'poly', 'rbf', 'sigmoid']
    },
    'LinearDiscriminantAnalysis': {
        'solver': ['svd', 'lsqr', 'eigen'],
        'shrinkage': ['auto', 'float']
    }
}

# Initialize classifiers
classifiers = {
    'LogisticRegression': LogisticRegression(),
    'ExtraTreesClassifier': ExtraTreesClassifier(),
    'XGBClassifier': XGBClassifier(),
    'AdaBoostClassifier': AdaBoostClassifier(),
    'KNeighborsClassifier': KNeighborsClassifier(),
    'SVC': SVC(),
    'LinearDiscriminantAnalysis': LinearDiscriminantAnalysis()
}
```

```
# Loop through classifiers and perform grid search
for name, clf in classifiers.items():
    grid_search = GridSearchCV(clf, param_grid_combined[name],
     scoring = 'balanced_accuracy')
    grid_search.fit(X_train, y_train)
    print(f"Best parameters for {name}: {grid_search.best_params_}")
```