UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE TECNOLOGIA E GEOCIÊNCIAS
DEPARTAMENTO DE ENGENHARIA DE PRODUÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

PLÍNIO MARCIO DA SILVA RAMOS

MULTIMODAL DATA-DRIVEN APPROACHES FOR INFERRING DROWSINESS FROM MACHINE LEARNING MODELS IN INDUSTRIAL ENVIRONMENTS WITH CRITICAL SAFETY SYSTEMS

Recife
2024

PLÍNIO MARCIO DA SILVA RAMOS

MULTIMODAL DATA-DRIVEN APPROACHES FOR INFERRING DROWSINESS FROM MACHINE LEARNING MODELS IN INDUSTRIAL ENVIRONMENTS WITH CRITICAL SAFETY SYSTEMS

Doctoral thesis presented to the Programa de Pós-Graduação em Engenharia de Produção to Universidade Federal de Pernambuco for the doctorate degree attainment as part of the requirements of the Engenharia de Produção.

Concentration area: Operations Research.

Advisor: Prof. Dr. Márcio José das Chagas Moura.

Recife

2024

.Catalogação de Publicação na Fonte. UFPE - Biblioteca Central

Ramos, Plínio Marcio da Silva.
Multimodal Data-Driven Approaches for Inferring Drowsiness from Machine Learning Models in Industrial Environments with Critical Safety Systems / Plinio Marcio da Silva Ramos. - Recife, 2024.
164f.: il.

Tese (Doutorado) - Universidade Federal de Pernambuco, Centro de Tecnologia e Geociências, Programa de Pós-Graduação em Engenharia De Produção, 2024.
Orientação: Márcio José das Chagas Moura.
Inclui referências e apêndices.

1. EEG; 2. O&G; 3. Fusão de Dados; 4. Visão Computacional; 5. Aprendizado de Máquina; 6. Aprendizado de Máquina Quântico. I. Moura, Márcio José das Chagas. II. Título.

UFPE-Biblioteca Central

PLÍNIO MARCIO DA SILA RAMOS

MULTIMODAL DATA-DRIVEN APPROACHES FOR INFERRING DROWSINESS FROM MACHINE LEARNING MODELS IN INDUSTRIAL ENVIRONMENTS WITH CRITICAL SAFETY SYSTEMS

Doctoral thesis presented to the Programa de Pós-Graduação em Engenharia de Produção to Universidade Federal de Pernambuco for the doctorate degree attainment as part of the requirements of the Engenharia de Produção.

Approved in: 18/11/2024.

EXAMINATION BOARD

_____
Prof. Dr. Márcio José das Chagas Moura (Advisor)
Universidade Federal de Pernambuco

_____
Prof. Dr. Marcelo Hazin Alencar (Internal Examiner)
Universidade Federal de Pernambuco

_____
Profª. Drª. Isis Didier Lins (Internal Examiner)
Universidade Federal de Pernambuco

_____
Prof. Dr. Paulo Fernando Ferreira Frutuoso e Melo
Universidade Federal do Rio de Janeiro (External Examiner)

_____
Prof. Dr. Marcelo Ramos Martins
Universidade de São Paulo (External Examiner)

ACKNOWLEDGEMENTS

I am grateful to God for the possibility of continuous learning in the various areas of life, for providing light, strength and wisdom in the most precise moments.

I would like to express my deep gratitude to my advisor Prof. Dr. Márcio Moura for his guidance and constructive advice.

To Prof. Dr. Caio Maior, who assisted me at different times from my first day at CEERMA until the completion of this thesis. Prof. Dr. Isis Didier for all the support during these doctoral years during various projects. To July and Lavínia, for all the support and moments of relaxation in the laboratory, and to all the colleagues who made me feel at home since my first day at CEERMA.

I thank my parents, Paulo and Maria, for the love that nurtured my development, for always believing in education and making every effort to prioritize it, in addition to being the solid foundation on which I built several steps in my life.

To my sister, Patricia Ramos, for being an example of dedication, perseverance and faith, which I mirror myself daily.

Finally, to all my friends and family who truly support me and supported me, even indirectly, in the development of this research.

Thanks to Programa de Recursos Humanos 38.1 (PRH 38.1 - Análise de Riscos e Modelagem Ambiental na Exploração, Desenvolvimento e Produção de Petróleo e Gás) of Agência Nacional de Petróleo, Gás Natural e Biocombustíveis (ANP) - Financiadora de Estudos e Projetos (Finep) management and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for the financial support through research scholarships and grants.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

“Learn the lessons. Give thanks for the blessings. Understand the purpose. Take your time.”

(----------)

ABSTRACT

Catastrophic accidents have been an issue in complex industries like oil and gas (O&G), chemical, and nuclear sectors, despite ongoing efforts to improve safety. While physical systems have advanced, human factors such as fatigue, drowsiness, and inattention remain significant risks, leading to reduced performance, errors in judgment, and an increased likelihood of accidents. Fatigue-related factors—poor rest, sleep deprivation, night shifts, stress, and prolonged monotony—are common in safety-critical environments and frequently result in drowsiness and lapses in attention. However, the subjective nature of self-reported drowsiness presents a challenge in detecting early signs to reduce potential risks and prevent accidents in organizations with high safety and environmental demands. Thus, this thesis presents an all-encompassing framework addressing operator performance and attention-related challenges in safety-critical industrial systems through several key contributions. First, it explores the application of machine learning (ML) and quantum machine learning (QML) for electroencephalogram (EEG) signal analysis, leveraging ensemble models and advanced neural network architectures to improve accuracy in detecting drowsiness. The introduction of variational quantum algorithms applied to EEG data analysis, which highlights quantum computing's potential to process large, complex datasets in industrial safety contexts, emerges as one of novel contribution of this work. Second, the thesis proposes a data fusion approach that combines physiological and visual (EEG and facial) data to enhance the robustness of drowsiness detection systems. This fusion is implemented at both the decision and feature levels, with experimental results showing significant improvements in recall and accuracy compared to single-modality approaches. Third, the development of a real-time web-based application, DrowsinessNET, integrates the detection model into a practical tool for monitoring drowsiness in high-risk environments. This application highlights the feasibility of applying advanced detection models in real-world scenarios. Finally, a simulator-based experiment was conducted to assess operator performance in automated O&G operations, particularly focusing on the impact of automation-related factors such as overconfidence, boredom, and inattention. The experiment reveals that automation can induce human errors and reduce attentiveness in monotonous tasks, further emphasizing the critical need for integrating human reliability technologies in safety-critical systems. Thus, this thesis pushes the boundaries of research field in human performance and operational safety by introducing multimodal data-driven models (ML/QML/DL), data fusion techniques, and practical applications to prevent accidents and enhance safety in high-risk industries.

Keywords: EEG, O&G, Data fusion, Computer Vision, Machine Learning, Deep Learning, Quantum Machine Learning.

RESUMO

Acidentes catastróficos têm sido um problema em indústrias complexas como petróleo e gás (O&G), química e nuclear, apesar dos esforços contínuos para melhorar a segurança. Embora os sistemas físicos tenham avançado, fatores humanos como fadiga, sonolência e desatenção continuam sendo riscos significativos, levando à redução do desempenho, erros de julgamento e maior probabilidade de acidentes. Fatores relacionados à fadiga — descanso insuficiente, privação de sono, turnos noturnos, estresse e monotonia prolongada — são comuns em ambientes críticos de segurança e frequentemente resultam em sonolência e lapsos de atenção. No entanto, a natureza subjetiva da sonolência autorrelatada apresenta um desafio na detecção de sinais precoces para reduzir riscos potenciais e prevenir acidentes em organizações com altas demandas de segurança e ambientais. Assim, esta tese apresenta uma estrutura abrangente que aborda o desempenho do operador e os desafios relacionados à atenção em sistemas industriais críticos de segurança por meio de várias contribuições. Primeiro, ela explora a aplicação de aprendizado de máquina (ML) e aprendizado de máquina quântica (QML) para análise de sinal de eletroencefalograma (EEG), alavancando modelos de conjunto e arquiteturas avançadas de rede neural para melhorar a precisão na detecção de sonolência. A introdução de algoritmos quânticos variacionais aplicados à análise de dados de EEG, que destaca o potencial da computação quântica para processar conjuntos de dados grandes e complexos em contextos de segurança industrial, surge como uma das novas contribuições deste trabalho. Segundo, a tese propõe uma abordagem de fusão de dados que combina dados fisiológicos e visuais (EEG e faciais) para aumentar a robustez dos sistemas de detecção de sonolência. Essa fusão é implementada nos níveis de decisão e de recurso, com resultados experimentais mostrando melhorias significativas na recuperação e precisão em comparação com abordagens de modalidade única. Terceiro, o desenvolvimento de um aplicativo baseado na web em tempo real, DrowsinessNET, integra o modelo de detecção em uma ferramenta prática para monitorar a sonolência em ambientes de alto risco. Este aplicativo destaca a viabilidade de aplicar modelos avançados de detecção em cenários do mundo real. Finalmente, um experimento baseado em simulador foi conduzido para avaliar o desempenho do operador em operações automatizadas de O&G, focando particularmente no impacto de fatores relacionados à automação, como excesso de confiança, tédio e desatenção. O experimento revela que a automação pode induzir erros humanos e reduzir a atenção em tarefas monótonas, enfatizando ainda mais a necessidade crítica de integrar tecnologias de confiabilidade humana em sistemas críticos de segurança. Assim, esta tese expande os limites do campo de pesquisa em desempenho humano e segurança operacional ao introduzir modelos multimodais orientados a dados (ML/QML/DL), técnicas de fusão de dados e aplicações práticas para prevenir acidentes e aumentar a segurança em indústrias de alto risco.

Palavras-chave: EEG, O&G, Fusão de Dados, Visão Computacional, Aprendizado de Máquina; Aprendizado de Máquina Profundo, Aprendizado de Máquina Quântico.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

ANP	Agência Nacional de Petróleo, Gás Natural e Biocombustíveis
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CNN	Convolutional Neural Network
CV	Computer Vision
DL	Deep Learning
EEG	Electroencephalogram
ECG	Electrocardiogram
EOG	Electrooculogram
GTG	Gas Turbine Generator
HFD	Higuchi Fractal Dimension
HM	Hjorth Mobility
HC	Hjorth Complexity
HRSG	Heat Recovery Steam Generators
LSTM	Long Short-Term Memory
ML	Machine Learning
NISQ	Noisy Intermediate Scale Quantum
OPC	Open Platform Communications
O&G	Oil and Gas
QML	Quantum Machine Learning
RESNET	Residual Neural Network
SMS	Safety Management System
TPOT	Tree-based Pipeline Optimization Tool
VQE	Variational Quantum Eigensolver
VQA	Variational Quantum Algorithms

CONTENTS

1. INTRODUCTION

1.1. Initial Remarks

Catastrophic accidents have occurred over the years in complex and critical systems, such as the oil and gas (O&G), chemical, and nuclear industries (Sadeghi & Goerlandt, 2023). It is known that despite the numerous interventions made by governments and industries around the world to improve safety (including the existence of highly reliable equipment), human operators continue to play an important role in safety-critical operations (Zarei et al., 2023).

In the O&G industry, disasters such as Piper Alpha in 1988, the P-36 platform in the Campos Basin (Brazil) in 2001 and the Deepwater Horizon platform in the Gulf of Mexico in 2010, resulted in dozens of deaths, infrastructure and environmental disasters, with reports pointing out that human factors were elements, directly or indirectly, contributing to the occurrence of these accidents (Figueiredo et al., 2018). Other events, where human errors in emergency response are considered relevant factors, are discussed by (Almeida & Vinnem, 2020).

In these high-risk environments, human errors can grow into process inefficiency and, ultimately, major consequences such as life and property losses and environmental impact (Ferraro & Mouloua, 2021; Selvik & Bellamy, 2020; Theophilus et al., 2017; Zhou et al., 2017). It is already understood that some factors (e.g., poor resting habits, drug reactions, nutritional imbalances and sleep deprivation) or task-related conditions (e.g., night shift, stress and boredom) are associated with fatigue, which degrades the employee's overall performance in work environments and induces drowsiness (B & Chinara, 2021; Camden et al., 2019). In fact, drowsiness is linked to several types of accidents with drivers (Bekhouche et al., 2022; Bhandari, 2014; Cui et al., 2022; Hajinoroozi et al., 2016; P. Wang et al., 2018), but it is also investigated in the daily lives of industrial professionals. Researchers have already evaluated occupational accidents associated with drowsiness in systems such as railway or aviation operations, estimating the proportion of accidents involving operator fatigue, with values ranging from 3.8 to 21% for aviation, for example (Gander et al., 2011). In the road context, it is estimated that 10–30% of deaths are related to drowsiness (Vicente et al., 2016). (Forsman et al., 2014) indicate that 28% of private drivers and 47% of long-distance truck drivers admitted to dozing off at the wheel at least once. There are some drowsiness studies in

the field of civil construction, chemical/petrochemical industries and metallurgy (Lavie et al., 1982; Waage et al., 2012). Particularly in O&G industries, operators report drowsiness in the first days of night shifts (Parkes, 2012; Waage et al., 2012). However, these studies have a more subjective character, without dealing with the challenge of detecting early signs and, then, reduce potential risks and prevent accidents in organizations with operations critical to safety and the environment.

Traditionally, detection of drowsiness and inattention has relied on two major approaches: computer vision - CV (e.g., images/videos) or biological (e.g., electroencephalogram, EEG) information (Bekhouche et al., 2022; Belakhdar et al., 2018; Cui et al., 2022; Murthy et al., 2022). Indeed, the biological approach provides a detailed characterization that associates, for example, EEG patterns with the subject's drowsiness level (Bajaj et al., 2020; Belakhdar et al., 2018; Lotte et al., 2018). The distinction between states of alertness and drowsiness can then be perceived through neurological changes in the subject's state of attention (G. Li et al., 2015) or through body language and facial expressions (You et al., 2019).

Brain waves have already been used to investigate drowsiness through sleep stage analysis, providing basic insights into EEG patterns to assess and categorize different sleep states (Hong & Baek, 2021). The brain produces electrical impulses (i.e., brain waves) that take different forms when a change in psychosomatic states occurs (i.e., from alertness to drowsiness). Transitions between different phases can be observed and recorded by multiple channels using strategically placed electrodes to obtain balanced measurements of all brain areas. Typically, researchers studying drowsiness use different EEG channels as well as various feature extraction methods to detect alertness and drowsiness. However, based on investigations, there is no strict consensus on which EEG channels are most appropriate for differentiating states of alertness and drowsiness (Belakhdar et al., 2016; Garcés Correa et al., 2014; Greenblatt et al., 2023; W. Li et al., 2012; Vanhollebeke et al., 2022; Yu et al., 2013). In spite of this, there seems to be a high correlation in the detection of drowsiness with the central and posterior regions of the cerebral area (Lin et al., 2005; Picot et al., 2008; Vanhollebeke et al., 2022; C. Zhao et al., 2017).

On the other hand, applying CV techniques proved valuable in capturing visual symptoms associated with drowsiness. By analyzing facial expressions, eye movements and other visual features, CV methods can capture additional information about an individual's level of alertness (C. B. S. Maior et al., 2020; Pandey & Muppalaneni, 2023).

Eye movements, yawning, and slow or dull facial movements are visual signs that can be tracked and analyzed using CV techniques. By monitoring these visual indicators, it becomes possible to infer the level of drowsiness of individuals (You et al., 2019). Indeed, researchers employ various approaches for detection, with a predominant focus on utilizing eye-related resources (Nayak & Kim, 2021; Paneru & Jeelani, 2021). Nevertheless, other facial components, such as the mouth or the entire face, have also been adopted, although it is important to note that certain facial expressions may provide misleading information (Ngai et al., 2022).

However, despite the potential of these two data sources (EEG and computer vision), there remains a notable gap in the literature regarding the combination of multiple data modalities and the use of more advanced methods. For instance, EEG, as an independent data source, has been extensively analyzed with various machine learning (ML) techniques to detect patterns associated with drowsiness. However, the combination of different signals in complex architectures, especially in critical industrial environments, remains an area that requires further exploration. The integration of multiple signals holds great promise for improving the accuracy and robustness of drowsiness and inattention detection systems.

Additionally, recent advancements in quantum computing have opened new avenues for solving computational problems over classical methods, such as in the prime factorization of a number (Shor, 1997) or simulation of quantum systems (Georgescu et al., 2014). Variational Quantum Algorithms (VQAs) (Cerezo et al., 2021) are considered one of the most promising through Quantum Machine Learning (QML) approaches for achieving heuristic quantum advantages with Noisy Intermediate Scale Quantum (NISQ) devices, enables the processing of high-dimensional EEG data, leading to more accurate detection and cognitive states.

Research has demonstrated the potential of QML for various EEG-related tasks. For instance, Y. Li et al. (2022) used QML to extract features from EEG data and classify them using a Quantum Support Vector Machine, which outperformed existing linear methods. Other studies have utilized techniques such as the Quantum Radial Wavelet Neural Network (M.R.Taha & K. Nawar, 2014) and the Quantum Recurrent Neural Network for the classification of epileptic EEG signals (Karayiannis et al., n.d.), and Quantum Neural Network for uncertainty measurement in neonatal seizures. However, in the applicability of drowsiness detection with EEG, QML approaches remain unexplored.

Beyond the application of ML and QML to EEG data, there is still a significant gap in the literature regarding the fusion of EEG and facial data for drowsiness detection, especially in high-stakes industrial operations. While several studies have examined these data sources individually, few have explored the fusion of multiple modalities. Generally, data fusion occurs at three distinct processing levels based on the degree of convergence according Cai et al. (2020) and Meng et al. (2020): pixel level, feature level, and decision level. Pixel-level fusion involves combining the original data layer, encompassing a thorough analysis of information before preprocessing. Feature-level fusion integrates various features linearly or nonlinearly after feature extraction to generate new fused features. Decision-level fusion entails making individual decisions based on different models and subsequently coordinating or merging them into a global decision.

The study by Mehmood et al. (2023) addressed mental fatigue in construction equipment operations, incorporating EEG, electrodermal activity, and video signals. However, their performance analysis was confined to the use of ML models (Artificial Neural Network, k-Nearest Neighbors, and Decision Tree). Other works have explored different fusion approaches when addressing drowsiness, such as Shahbakhti et al. (2023) study, which employed EEG data and eye blink analysis extracted from the EEG data itself for driver fatigue detection, using machine learning models (SVM, LDA, KNN, AdaBoost, ANN, and RF). L. Zhao et al. (2022) work adopted a multimodal fusion approach, combining ECG (electrocardiogram) data and video signals. It is noteworthy that other studies have focused on data fusion beyond physiological information, integrating and combining information from multiple sources or modalities to enhance the overall understanding or performance in drowsiness detection. For instance, Cheng et al. (2012) and B.-G. Lee & Chung (2012) considered participant-oriented and vehicle-oriented data, such as abnormal eye behavior, steering wheel activity, and vehicle trajectory.

Thus, these studies underscore the need to explore more comprehensive strategies for effective drowsiness detection in critical environments, such as in a monitoring control room in an oil and gas industry.

1.2. Motivation

Identifying fatigue in O&G operators is of current interest to global petroleum industries (National Transportation Safety Board, 2023). In fact, accident reports from these industries have highlighted that, directly or indirectly, human factors are the main

causes of accidents on platforms (Zarei et al., 2023). Based on knowledge in other fields of research, drowsiness is understood as one of the consequences of human fatigue and one of the precursor to inattention, and can be identified through physical aspects as well as through neurological changes (Bajaj et al., 2020) and also, through changes in heart patterns and muscle activity (Qi et al., 2018; Shiwu et al., 2011).

It is known that the advancement of current technology and lower costs allow people to use wearable sensory devices on a daily basis in order to capture information that may be useful for a certain purpose. Drowsiness detection mechanisms have been extensively studied in recent years.

1.2.1. Gaps in drowsiness detection approaches

Several studies (Bekhouche et al., 2022; Bhandari, 2014; da Silveira et al., 2016; Guarda et al., 2022; Murthy et al., 2022; Phan et al., 2023) have attempted to assess drowsiness levels using behavioral analyses based on camera's eye tracking systems and physiological markers found in EEG signals with individual approaches and classical models. Given that drowsiness patterns can vary across different populations, detection systems often rely on specific indications from a single method, increasing the possibility of missed detection (Kartsch et al., 2018). Thus, systems that integrate multi-source data or incorporate recent data processing techniques are suitable. Hence, one may highlight current gaps in the literature:

- Limited exploration of drowsiness detection methodologies, primarily relying on classical machine learning models for analyzing EEG data. However, there has been a notable absence in the investigation of more advanced approaches, such as quantum machine learning.
- Limited exploration of the data fusion approach involving the analysis of video and EEG data in the context of drowsiness, with most instances found in studies related to emotions.
- The predominant trend in data fusion studies leans towards evaluation and classification models, using mainly traditional machine learning methods. There is just a couple of studies exploring the application of more advanced models, such as deep learning or quantum-based methods, in the domain of drowsiness detection through the data fusion approach.

1.2.2. Experimental Setup for Assessing Operator behavior in a Refinery Control Room

Control rooms are ubiquitous in industries such as O&G (e.g. oil refineries, offshore and onshore production platforms, storage and distribution terminals). They serve as central hubs, where critical decisions are made to ensure efficient and safe operations. However, at the best of authors' knowledge, previous research has had substantial focus on driver-related drowsiness studies (Cheng et al., 2012; Hajinoroozi et al., 2016; Katsis et al., 2008; B.-G. Lee & Chung, 2012; Phan et al., 2023; Shahbakhti et al., 2023; P. Wang et al., 2018), neglecting investigations into attention-related tasks as expected in an O&G industry control room environment.

On the other hand, studies that focused on experimental set-up in control rooms have not addressed specific investigations into activities related to the O&G industry, and monitoring tasks that could induce operators to report changes in alertness. Instead, these works have focused primarily on operator performance at different levels of automation and task complexity (Bye, 2023; Jang et al., 2021; Omidi et al., 2018; M. A. Ramos & Mosleh, 2021).

This study utilizes an experimental setup to evaluate the impact of automation on operator performance in a simulated refinery control room. The simulation focuses on monitoring tasks, where operators must maintain acceptable water levels in boilers and ensure safe steam pressure. The simulation replicates refinery some challenges, such as low water levels affecting steam circulation or high-water levels risking equipment damage. This experimental phase is crucial for understanding how automation affects operator attention, leading to screen fatigue and lapses in decision-making in critical environments.

By studying how automation influences inattention in operators, this research aims to contribute new insights and in the creation of an EEG and video data repository, which will serve as a foundation for future research in human reliability. This database provides high-quality data that reflects the variability of human responses in real-world conditions, particularly in automated environments where monotony and reduced attention pose significant risks.

1.3. Objectives

This section describes the general and specific objectives of this thesis.

1.3.1. General Objective

The main objective of this thesis is to develop data-driven solutions for automatic drowsiness detection, integrating EEG and facial data, aiming to increase human reliability in safety-critical systems that require constant attention and rapid response. In addition, to propose an experimental setup within a simulated control room environment adapted to the oil and gas industry to evaluate operator performance..

1.3.2. Specific Objectives

Given the general objective of this research, some specific objectives can be listed as follows:

- Implement and evaluate ensemble machine learning models to analyze EEG data for detecting drowsiness, focusing on improving accuracy and reliability.
- Explore the applicability of variational quantum algorithms (VQAs) for drowsiness detection by utilizing VQC to classify patterns in EEG signals, offering a novel approach for high-dimensional data analysis.
- Develop a data fusion methodology that integrates EEG signals and computer vision techniques, creating a multi-modal system for enhanced drowsiness detection.
- Develop DrowsinessNET, a web-based interface for real-time drowsiness detection, incorporating the models developed in this research, to provide an accessible tool for monitoring operator fatigue in high-risk environments.
- Design and conduct a simulated experiment to evaluate the effects of automation and monotonous tasks on operator performance in a simulated oil and gas refinery control room.
- Critically analyze the performance of the developed models, highlighting their innovative contributions to the field and evaluating their impact on improving human reliability in safety-critical environments.
- Develop a systematic modeling approach for data extraction, feature engineering, and classification, ensuring a robust pipeline for drowsiness detection.
- Build final classifiers using advanced machine learning models to perform supervised tasks, evaluating the results against the baseline and determining their impact on model performance.

- Utilize the resulting drowsiness data from the methodology to perform practical classification tasks, demonstrating the practical application of the models for decision-making in safety-critical environments.

1.4. Outline of the Thesis

Besides this introduction chapter, this thesis has five additional chapters briefly described as follows:

- Chapter 2 provides a theoretical background on operator fatigue, drowsiness, and machine learning and deep learning techniques, setting up the stage for understanding their application in drowsiness detection; it also presents works that applied machine learning and deep learning to the drowsiness context.
- Chapter 3 introduces a comprehensive visual summary outlining the diverse approaches explored in the thesis in the upcoming chapters. Before delving into specific applications, the chapter provides a general overview of the DROZY dataset. This dataset, a curated collection of multimodal data capturing various aspects of drowsiness, stands as a foundational element for the applications.
- Chapter 4 dedicated to improving the identification and prediction of drowsiness from EEG data. The main objective is to contribute to the prevention of accidents through the development of robust data-driven models of multiple EEG sources, capable of effectively recognizing various characteristics associated with drowsiness. The models are applied to a real public database and evaluated regarding their accuracy in detecting drowsiness.
- Chapter 5 delves into the exploration of quantum machine learning techniques for the detection and analysis of drowsiness. Building upon the theoretical foundation established in earlier chapters, this chapter introduces the novel approach of utilizing quantum computing principles in the context of drowsiness detection. It describes the methodology in detail, including the preprocessing steps for quantum data encoding and the implementation of quantum model for drowsiness prediction.
- Chapter 6 presents the proposed framework that integrates machine and deep learning methods based on EEG signals and CV techniques to detect and analyze drowsiness. This chapter describes the methodology and framework architecture in detail. It then applies the proposed methodology to a public available database

of drowsiness. The analysis includes training the models and evaluating their performance in drowsiness prediction. Additionally, an application of the resulting models is demonstrated, showcasing their ability to accurately predict drowsiness based on the information extracted from EEG signals and facial images.

- Chapter 7 presents an experimental setup addressing the impacts of automation on human factors and human performance, highlighting potential challenges such as overconfidence, boredom, and automated systems transparency issues. The proposed experimental setup aims to assess human performance in control room operations of automated O&G systems, utilizing a low-fidelity simulation of a refinery process. This setup considers factors like task complexity, execution time, screen fatigue, and automation failures, ultimately recommending the application of a machine learning model for non-intrusively detecting drowsiness in monitoring tasks. Additionally, the chapter introduces the development of a web application interface for data analysis that allows early detection of signs of drowsiness, providing operators with the opportunity to act proactively to prevent accidents and ensure safe operations.
- Chapter 8: concludes remarks. This chapter consolidates the findings from previous ones, drawing overarching remarks and insights. It emphasizes the practical implications of the developed methodologies for enhancing safety in critical industries. Furthermore, the chapter highlights the limitations of the current research and suggests potential directions for future studies.

2. THEORETICAL BACKGROUND

This chapter provides key concepts and definitions about human factors and drowsiness and describes classic and advanced machine learning techniques that have been used to detect drowsiness, in addition to providing a literature review on relevant topics. The application discussed in Chapters 4, 5, 6 and 6.4 draws upon these concepts and insights from the literature review.

2.1. Human Reliability in the Context of Risk Management in Critical Safety Industries

There is an enormous effort and interest in different industries and societies to manage risks. The industrial landscape has traditionally focused primarily on equipment reliability, neglecting the significant impact of human reliability (Abílio Ramos et al., 2020). However, there has been a growing recognition of the importance of human factors in risk analysis, especially when considering accident reports in the critical safety industry.

For example, the O&G industry has suffered numerous catastrophic accidents, most attributed to organizational and operational human errors. The BP Texas City refinery explosion (2005) resulted in approximately 15 deaths, and the oil industry offshore disaster, Piper Alpha (1988), claimed the lives of 167 individuals and left dozens seriously injured. Investigations revealed that some of the tools and measures adopted were not robust enough to prevent such accidents (Almeida & Vinnem, 2020).

To address these concerns, the international standard ISO 14224 was established to provide essential guidelines for obtaining information about failures in the O&G sector, including specific guidance on data collection concepts and recording and categorizing failure causes (Morag et al., 2018). Indeed, fault and reliability performance information often receive significant attention in the design, operation and maintenance of equipment in the O&G industry, particularly for safety-critical systems. Nevertheless, the third edition, released in September 2016, focuses primarily on equipment failures, but places greater emphasis on human errors compared to previous editions (Selvik & Bellamy, 2020). For example, the definition used in ISO 14224 (Technical Committee ISO/TC 67. Materials petrochemical and natural gas industries, 2016) is as follows:

human error: "discrepancy between the human action taken or omitted and that intended".

Other definitions used are:

error: "discrepancy between a computed, observed or measured value or condition and the true, specified or theoretically correct value or condition"

Thus, while the primary focus is on understanding equipment behavior in terms of reliability performance, there is an inherent relationship with human factors. Even though this relationship may seem familiar to those involved in risk assessment of hazardous systems (nuclear, oil and gas, chemicals), it was largely ignored in previous editions of the standard (Selvik & Bellamy, 2020). Finally, in the third edition, a step is taken towards linking equipment reliability performance to human factors.

Furthermore, it is important to consider initiatives undertaken by regulatory bodies, such as the National Petroleum, Natural Gas and Biofuels Agency (ANP), which plays a fundamental role in Safety Management System (SMS) in the oil and gas industry in Brazil(ANP, 2023). The ANP, through its SMS, highlights the importance of human factors in risk analysis.

The latest version of the ANP's SMS explicitly recognizes human factors as an integral part of risk management. By mentioning the dissemination of the sector's best practices in Human Factors, the SMS reinforces the importance of considering organizational, environmental, technological and individual elements that influence behavior in the workplace and can impact health and safety.

Within the scope of management practices, the ANP SMS clearly establishes, in section 12.3.e, that the risk identification and analysis methodology must consider layout, human factors and external causes, as applicable(ANP, 2023). This reflects a deep understanding of the need to incorporate human factors considerations from the earliest stages of risk analysis. Furthermore, human reliability analysis is highlighted as the most appropriate methodology for evaluating human factors in tasks, especially those involving human-machine interactions.

This demonstrates that developments such as these have sparked discussion in the fields of human reliability analysis and human factors about how current systems, even the most automated ones, affect human performance and human error (Bye, 2023; Theophilus et al., 2017; Tinga et al., 2023). The implementation of an effective SMS is crucial in addressing the impact of human factors on risk management. An SMS is a comprehensive framework that integrates various elements, including organizational policies, procedures, and practices, to ensure the safety and security of personnel and assets within an industry (Pilanawithana et al., 2023). Within the O&G sector, the

adoption of a robust SMS can help mitigate the risks associated with human reliability. By incorporating human factors considerations into the system, organizations can identify and address potential sources of error among operators, thereby reducing the likelihood of accidents (Uflaz et al., 2022).

2.2. Drowsiness

Drowsiness is a state of reduced alertness that can result in decreased concentration, impaired decision-making, and fatigue (Sharma et al., 2021). It can be caused by a variety of factors, including poor sleep hygiene, age, and monotonous work. In industries demanding shift work, such as the O&G, nuclear, or in the transportation sector, including air, sea, and road transport, drowsiness plays a key role in workplace accidents and mishaps (Ramos et al., 2022). Indeed, environmental variables (e.g., noise, vibrations, temperature, poor lighting conditions) and lack of adequate rest and recovery due to stress contribute to increased drowsiness and fatigue (Sadeghi & Goerlandt, 2023). Organizations can identify and address potential sources of error among operators, thereby reducing the likelihood of accidents (Uflaz et al., 2022).

Shift work is prevalent in many industries (e.g., O&G), and sleep disturbance (i.e., poor sleep quality, insomnia, and daytime drowsiness) often results in accidents and errors in the workplace (Iridiastadi, 2021; Sadeghniiat-Haghighi et al., 2020). In fact, it has been reported that, compared to 8 hours of shift work, the risk of an incident or accident after 10 hours of shift work nearly doubles (Folkard, 2003).

Similarly, fatigue is one of the leading factors contributing to accidents in the transport sector, such as aviation, road, public transport (e.g., buses, trains), large vehicle industries (e.g., bulldozers and excavators) as well as in several other sectors, such as construction (Gu & Guo, 2022; J. Li et al., 2020), mining (Gruenhagen et al., 2021), maritime (Andrei et al., 2020; van Leeuwen et al., 2021). For example, 4–8% of civil aviation mishaps are caused by drowsy pilots (Caldwell, 2005). In 2017, the US National Highway Traffic Safety Administration (NHTSA) estimated that fatigue caused 91,000 accidents, 50,000 injuries, and 800 fatalities in the US alone. Besides, concerns regarding the risk of accidents due to fatigue have been identified by the mining sector, where operators often work 12-hour shifts (Kartsch et al., 2018), and for maritime transport, for which it is reported the highest number of incidents between 3:00 am and 7:00 am, portraying a period when humans are not at their maximum alert level (van Leeuwen et al., 2021).

In this context, safety remains a major challenge for almost all critical operations activities. At the same time, human reliability plays an important role not only in service industries but also in manufacturing companies. Indeed, operators with a decreased level of vigilance reduce their skills of perception, recognition, and control of the workplace and, therefore, may pose a danger to their own lives and, in most cases, the lives of other people (Ji et al., 2004; Ramos et al., 2022).

2.3. Drowsiness experimental set-up

Despite the evident challenges posed by drowsiness in the O&G industry, there remains a notable lack of dedicated datasets and experiments addressing this critical issue within this sector, particularly in environments such as refinery control rooms. The demanding nature of offshore work, characterized by harsh conditions and stressful environments, has long been acknowledged (Parkes, 2012). The physical and psychosocial stressors, coupled with intensive work patterns, make the O&G industry particularly susceptible to fatigue-induced lapses in performance and attention.

Indeed, the existing body of literature on the offshore environment highlights the complexities of shift schedules, stressors, and their impact on the health and safety of personnel. Various studies have delved into the effects of extended work schedules, day/night shift rotations, and the implications of intensive rosters on offshore installations (Waage et al., 2012). Research conducted in other sectors, such as the automotive industry, has extensively examined driver drowsiness (Cheng et al., 2012; Hajinoroozi et al., 2016; Katsis et al., 2008; B.-G. Lee & Chung, 2012; Phan et al., 2023; Shahbakhti et al., 2023; P. Wang et al., 2018). However, when it comes to the O&G industry, there's an evident gap in dedicated experiments and datasets directly addressing the issue of drowsiness for control operations centers such as a refinery.

For example, in the Jacobé de Naurois et al. (2018) study, the driving protocol lasted 100 to 110 minutes in a static driving simulator, installed in a temperature-controlled room at 24 °Celsius. Participants drove right after lunch, a time recognized as risky for inducing drowsiness. The simulated environment, generated with SCANeR Studio, featured a drowsiness-inducing monotonous stretch of highway, with a sudden introduction of traffic intended to alter the driver's level of drowsiness. Data including driving performance, eyelid and head movements, and physiological indicators were recorded. Physiological, behavioral and car-related variables were collected, along with

participant information and driving time, and were used in models specifically developed for this study.

The study of Arefnezhad et al. (2020) utilized a fixed-base driving simulator (Automated Driving Simulator of Graz, ADSG) at the Graz University of Technology. The simulator incorporated visual cues through eight large LCD displays, a sound system generating engine sounds and vibrations, and a full car setup with an automatic gearbox. Participants controlled the vehicle using a force feedback steering wheel, electronic gas pedal, and force feedback brake pedal. Facial-based and physiological measures, including eye movement data (SmartEye® eye-tracking system), heart rate variability signals (mobile ECG and wearable biosignal recorder), were recorded. The driving tests involved 93 sessions with 47 participants, all licensed drivers covering different age and gender groups. Tests were conducted in rested and tired states, with participants driving in a monotonous simulated environment for 30 minutes, covering a closed path resembling a three-lane motorway. Ground truth for drowsiness levels (alert, moderately drowsy, extremely drowsy) was established through video observations. The methodology involved features extraction from vehicle-based measures and ECG signals.

In another context (Eskandarian & Mortazavi, 2007), experiments were conducted at the Center for Intelligent Systems Research (CISR) Truck Driving Simulator Laboratory using a fixed-base driving simulator with a full-size real truck cab. The simulator included advanced features such as sophisticated vehicle dynamics, superior graphics, a five-channel projection system, and an instrumented truck cab with original displays and controls. The scenario simulated a 52-mile stretch of Interstate 70 to induce monotony and drowsiness. Thirteen truck drivers participated, aged between 23 and 60, each carrying out morning and evening driving sessions. The evening session aimed to induce drowsiness due to sleep deprivation. Drowsiness was assessed subjectively using a five-level rating scale based on driver behavior, performance, and eye closure. Severity of Drowsiness (SEVD) was introduced as a variable measured by the total time with drowsiness level divided by driving time.

Developments in refinery simulations can be seen in (Andrade, 2018) that is concerned with human behavior and performance in economic activities, especially in industrial environments such as refineries. The study proposes an alternative approach to collecting data for Human Reliability Assessment from simulation sessions. Using Game Engines to create 3D Virtual Environments (VE), the study explores a more budget-

friendly approach. An experiment was conducted to validate this approach by creating a VE to simulate the evacuation of a refinery plant in the event of a toxic cloud release.

In turn, (Soares, 2019)highlights the importance of training emergency teams in O&G refineries, given the dangerous characteristics of the environment and the risks associated with accidents. The study proposes the creation of a Virtual Reality (VR) scenario to simulate emergencies in a refinery, aiming to improve team training. Through techniques such as Preliminary Risk Analysis and the use of ALOHA software, potential dangers were identified and critical regions for LPG leaks were defined. The virtual scenario was developed based on an emergency simulation, using the Unreal Engine 4 game engine, and tested by volunteers who work at the refinery. The results indicated that VR can be an effective tool for training firefighters, bringing them close to real emergency scenarios. Although the studies discussed advances in refinery simulations, focusing on improving the training of emergency responders and collecting data on human behavior, it is important to highlight that they do not directly address drowsiness, but rather the use of simulators for different purposes.

2.4. Exploring Drowsiness via Data

Over the years, technology improvements have enabled several methodologies to address fatigue monitoring and drowsiness detection. Models using CV approaches (from camera sensors) as well as biological approaches (from heart rate sensors or electrooculogram signals) have already been used to identify drowsiness (Kim et al., 2017; C. B. S. Maior et al., 2020; Wei et al., 2013). Among these approaches, EEG signal analysis is one of the most used method to detect drowsiness (Hong & Baek, 2021; Kartsch et al., 2018; Ramos et al., 2022).

Drowsiness, presenting a multifaceted nature, requires deeper exploration from multiple data sources. Thus, exploring drowsiness through two dimensions (EEG and facial data) aims to discover insights into and between physiological and visual signals for the development of robust detection systems.

2.4.1. Electroencephalogram - EEG

Brain waves have been used to investigate drowsiness through sleep stage analysis based on EEG patterns (Hong & Baek, 2021). The brain produces electrical impulses (i.e., brain waves) that take different forms when a change in psychosomatic states happens (e.g., from alertness to drowsiness). These brain waves may be recorded over time, and

distinct bandwidths (i.e., delta (δ), theta (θ), alpha (α), beta (β) and gamma (γ)) can be determined (Okello et al., 2016). Table 1 summarizes the brainwave frequencies and functions in normal activities.

Table 1 - Different EEG bandwidths and frequencies.

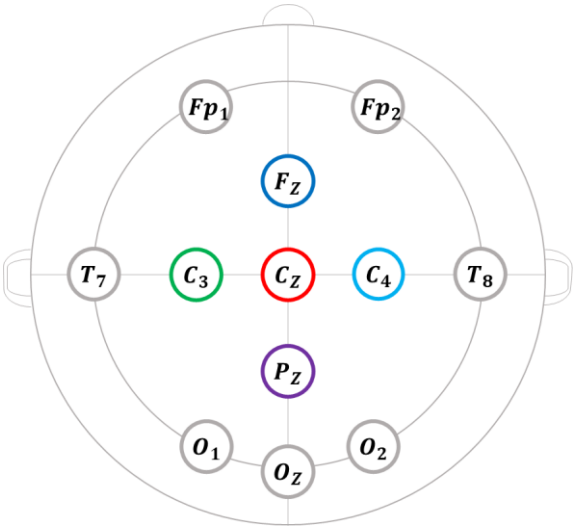
Bandwidth	Frequency	Normal Activities
Delta (δ)	0.1–4 Hz	artifacts, sleep, hyperventilation
Theta (θ)	4–8 Hz	drowsiness, idling
Alpha (α)	8–12 Hz	closing the eyes, inhibitory control
Beta (β)	12–30 Hz	alertness, stress, active thinking, focus
Gamma (γ)	30–70 Hz	voluntary motor movement, learning, and memory

Source: Adapted from (Birjandtalab et al., 2017).

During different stages of sleep, neuronal activities from different cerebral areas interact with each other and manifest as a distinct EEG activity, leading to the differentiation of the night sleep into stages (I, II, III, and IV followed by rapid eye movement (REM) stage) (Sriraam et al., 2016). In particular, stage I may be separated into two parts: (i) stage I immediately following wake and (ii) stage I immediately preceding stage II (Picchioni et al., 2008). The transition from awake to stage I represents the change from alertness to drowsiness (Shepovalnikov et al., 2012). Transitions between the different stages can be observed in a multi-channel recording using strategically positioned electrodes that aim to achieve balanced measurements from all cerebral areas. The locations of the electrodes are based on the international system of electrode placement (Acharya et al., 2016), as illustrated in Figure 1.

The abbreviations in Figure 1 represent the pre-frontal (Fp), temporal (T), parietal (P), and occipital (O) cerebral lobes, and the central (C) area. The ‘z’ symbolizes ‘zero’ in the midline. The subscribed numbers of the remaining electrodes indicate the left (odd) or right (even) hemisphere of the brain from a relative distance from the zero line. Thus, this EEG information can be used to create models able to identify the first signs of drowsiness.

Figure 1 - 12-electrode placement system



Source: Adapted from (Hong & Baek, 2021).

In recent years, advancements in EEG technology have opened new avenues for monitoring cognitive states such as drowsiness with greater precision and convenience. EEG devices offer lightweight and user-friendly solutions for real-time brainwave monitoring, suitable for both research and personal use applications. Moreover, emerging technologies such as dry electrode systems or helmet-mounted EEG devices present promising opportunities for operators in industries like oil and gas. These innovations enable non-invasive and comfortable monitoring of cognitive states, providing valuable insights into operators' alertness levels without disrupting their workflow.

2.4.2. Facial data

CV involves the use of algorithms and techniques to process and analyze visual data, allowing systems to understand and extract meaningful information from digital images or videos, replicating the functionality of human visual systems (Paneru & Jeelani, 2021). In detecting drowsiness, CV can play a key role in identifying patterns and facial features associated with drowsiness, thus providing an effective means of continuously monitoring operators' alertness (Bekhouche et al., 2022).

One of the most common methods of detecting drowsiness with CV is tracking the eyes of individuals (Arshad et al., 2023; Kolus, 2024; C. B. S. Maior et al., 2020; Nayak & Kim, 2021; Paneru & Jeelani, 2021; Rahman et al., 2022). This process involves capturing images of the operators' faces using cameras and applying algorithms to analyze the opening and movement of the eyes, as well as the frequency and duration of blinks.

Abnormal changes in these patterns can indicate drowsiness, signaling the need for immediate intervention.

It is important to emphasize that CV for drowsiness detection is not limited to analyzing just the eyes. Other facial features such as sleepy facial expressions, head tilt and involuntary movements, can also be considered for drowsiness identification (Bhandari, 2014; Murthy et al., 2022; You et al., 2019). Combining multiple facial indicators may provide a more comprehensive and accurate approach to drowsiness detection, despite involving a higher computational cost.

Furthermore, CV can be combined with deep learning techniques, such as convolutional neural networks (CNNs) and combination of CNN and Long Short-Term Memory (LSTM), to improve the accuracy and efficiency of drowsiness detection systems (Kolus, 2024; Murthy et al., 2022). The training of models with datasets of facial images of people who are drowsy and in alert states allows the system to recognize specific patterns associated with drowsiness, thus increasing its detection capacity.

However, it is important to mention some limitations of CV in detecting drowsiness. Environmental conditions, image quality and individual variations can affect the accuracy of the systems(Bekhouche et al., 2022). Therefore, it is essential to carry out adequate tests and validations to guarantee the reliability and effectiveness of these systems before their implementation in industrial environments. In contrast, CV offers a non-intrusive approach to monitoring individuals' alertness (Kolus, 2024; C. B. S. Maior et al., 2020). Thus, this advantage supports the scalability of the CV approach, helping to prevent health and safety risks in critical industries.

2.5. Related Works

Drowsiness detection is an area where ML and DL models have been successfully applied, especially when considering EEG data and CV techniques individually (B & Chinara, 2021; Bajaj et al., 2020; Belakhdar et al., 2018; Farhangi, 2022; Ramos et al., 2022; Sharma et al., 2021; Sors et al., 2018; P. Wang et al., 2018). In this section, we will explore examples of these models and address the application of feature extraction methods that contribute to enhancing drowsiness detection.

2.5.1. Feature extraction

Nonlinear analytic advancements have made possible quickly and accurately extracting information from signals captured during the physiological processes in the

human brain. However, currently, there is no standard feature extraction method for EEG (Bajaj et al., 2020).

The combination of internal and external factors that influence the human organism is the cause of the nonlinear behavior of biological (including the nervous) systems, which is characterized by a high degree of variability in the time domain (non-stationarity) and randomness (Kesić & Spasić, 2016). A nonlinear measure often used for signal analysis is the Higuchi fractal dimension (HFD) (Higuchi, 1988). Hjorth parameters (Hjorth, 1970) are indicators of statistical properties commonly used in the analysis of EEG signals for feature extraction, and which have been used in many studies to effectively describe the characteristics of EEG (B & Chinara, 2021; Bajaj et al., 2020; Kesić & Spasić, 2016).

2.5.1.1. HFD

Given the EEG signal is represented as X of N data points $x_1, x_2, \dots, x_N$, we first construct a new time series "X_d^m", where m and d indicate the initial time and the interval time, respectively (Higuchi, 1988). For this work, we consider the default parameters from the eeglib library in the Python language (Cabañero-Gomez et al., 2021). Then, the new time series is defined as in Equation (1):

$$X_d^m = x_m, x_{m+d}, x_{m+2d}, \dots, x_{m+\left[\frac{N-m}{d}\right]d} \tag{1}$$

For each X_d^m, the length $L_m(d)$ is computed as Equation (2):

$$L_m(d) = \frac{\left\{\left[\sum_{i=1}^{\left[\frac{N-m}{d}\right]} \left|X_{m+id} - x_{m+(i-1)d}\right|\right] \frac{N-1}{\left[\frac{N-m}{d}\right]d}\right\}}{d} \tag{2}$$

The mean of $L_m(d)$is computed to find the HFD as shown below in Equation (3):

$$L(d) = \frac{1}{d}\sum_{M=1}^{d} L_m(d) \tag{3}$$

Polynomial curve fit is computed on a logarithmic value of "$\log(L(d))$" and "$\log(d)$" with a degree one. Finally, Higuchi Fractal Dimension is the coefficient (p_1) of a polynomial curve $P(x)$ show in Equation (4):

$$P(x) = p_1(x) + c \tag{4}$$

2.5.1.2. Hjorth parameters

Hjorth Mobility (HM) and Hjorth Complexity (HC) were introduced to describe the general characteristics of an EEG trace.

- Complexity: measures the neurophysiological changes in terms of frequency (B & Chinara, 2021). It is expressed by the square root of differences of two ratios as shown in Equation (5).

$$Complexity = \sqrt{\left[\frac{rms\left(\frac{d(X')}{dt}\right)}{rms(X')}\right] - \left[\frac{rms(X')}{rms(X)}\right]} \quad (5)$$

where $X' = \frac{dX}{dt}$ is the rate of change of EEG signal (X) with respect to time (t) and rms is the root mean square.

- Mobility: is expressed as a ratio per unit of time and can also be conceived as a mean frequency (Hjorth, 1970). Mobility is measured as shown in Equation (6).

$$Mobility = \frac{rms(X')}{rms(X)} \quad (6)$$

2.5.1.3. Discrete Wavelet Transform

Another commonly used technique is the Discrete Wavelet Transform (DWT), which allows for the decomposition of the EEG signal into different frequency sub bands. The DWT decomposes a signal into approximation and detail signals, enhancing the reliability of estimations compared to the original signal (Mokarram et al., 2023). It can reliably localize non-stationary signals in the time and frequency domains and capture them. By applying the scale and shift technique, DWT breaks down the transient signal with short window scale for the high-frequency band and long window scale for the low-frequency band (Equation 7) (Tajani et al., 2023).

$$DWT(m,n) = \frac{1}{\sqrt{a^0{}_{\mathrm{m}}}} \sum_k x(k)\psi\left(\frac{(n-k)b^0{}_{\mathrm{m}}\, a^0{}_{\mathrm{m}}}{a^0{}_{\mathrm{m}}}\right) \quad (7)$$

where $x(k)$ is the input discrete signal, $\psi(n)$ is the mother wavelet (window function), m and n are time scale parameters, k is the number of coefficients, a_0^m is the variable for scale, $kb_0a_0^m$ is the variable for shift, and $1/(a_0^m)^{-1/2}$ is the energy normalization

component to guarantee that the scale is the same as the mother wavelet. After the signal is filtered, it is down-sampled with low pass and high pass filters. A signal processing technique called down-sampling reduces the information size by changing the sampling rate of a digital signal.

In this thesis, the Daubechies wavelets (Daubechies, 1992), specifically Daubechies 10 wavelets, were selected as the preferred mother wavelet among various other candidates. The choice was based on the tightly supported nature of Daubechies wavelets, which possess the maximum number of vanishing moments for a certain support length.

Thus, DWT operates by applying a set of wavelet functions, known as mother wavelets, to the signal, resulting in a decomposition into approximation and detail coefficients. Using DWT for preprocessing EEG data, we can effectively extract relevant features that contribute to accurate drowsiness detection and analysis. The resulting decomposition coefficients serve as input for subsequent classification or pattern recognition algorithms, improving the overall performance of the drowsiness detection system.

2.5.2. Machine and Deep Learning Models

EEG is a widely used technique for recording the electrical activity of the brain. As already mentioned, it provides information about brain wave patterns, which can vary according to the individual's state of alert. In the domain of sleep pattern identification, one popular approach is the utilization of Machine Learning algorithms, including K-Nearest Neighbor (KNN), Support Vector Machines (SVM), MLP, and Random Forests (RF). These models are trained with labeled datasets, where samples are classified into different sleep states (drowsy, alert, extreme drowsy, etc.).

In recent years, various disciplines and industries have successfully used (ML) to perform tasks that were intractable, tedious, or too expensive for humans (C. Maior et al., 2016). For example, (Farhangi, 2022) used six ML algorithms - decision trees, extra trees, K-Nearest Neighbor (KNN), Multilayer Perceptron (MLP), random forest (RF), and Support Vector Machine (SVM) -- to evaluate the effectiveness of EEG signal classification in detecting drowsiness while driving, and to understand how data pre-processing and hyperparameter tuning can improve drowsy EEG signal modeling. Based on features extracted from EEG data, these models can effectively learn to recognize patterns indicative of drowsiness. For instance,

- KNN assigns to an unmarked point the dominant class among its k closest neighbors within the training set. These closest neighbors are usually obtained using a metric of distance. KNN can approach any function that allows it to produce non-linear decision limits, depending on the k value and the size of training samples (Lotte et al., 2018; P. Wang et al., 2018).
- RF is defined as a group of classification trees trained on samples of training data using variables or resources selected at random in the tree generation process (P. Wang et al., 2018). Classification is performed based on the average prediction values of multiple decision trees. The number of trees and the depth of the tree are two parameters that can be adjusted.
- While classical approaches are designed to minimize errors in the training data set, SVM is based on the principle of structural risk minimization rooted in statistical learning theory (Widodo & Yang, 2007). SVM aims to build an optimal hyperplane that maximizes the margin between classes and handles large feature spaces (Lotte et al., 2018).
- MLP is one of the most popular networks that contain multiple successive layers, being the input layer, one or more hidden layers, and an output layer (X. Wang et al., 2020). Information is propagated from the input layer to the output layer through hidden layers and the weights of the network are updated during the training phase (Belakhdar et al., 2016).

For CV, as mentioned earlier, tracking the eyes, eyelids, and facial expressions can be used to identify visual indicators of drowsiness. The Deep Learning approach is widely adopted in this context (Kim et al., 2017; C. Lee & An, 2023; Liu et al., 2021). Since its initial success in many challenging image classification problems, it has rapidly developed and achieved success in many domains, such as speech recognition, sensor readings, spectrographs, electrocardiograms, and simulated data (Cui et al., 2022). These algorithms can be trained with datasets containing facial images of drowsy and alert people. This allows models to learn to recognize specific visual patterns associated with drowsiness, improving detection accuracy.

CNNs are architectures specifically designed for CV tasks. As one of the major deep learning models, CNN are capable of learning patterns and relevant features in images by using convolutional layers, which apply convolutional filters over the input

image, capturing local information and detecting features. A typical CNN structure consists of alternating convolutional and pooling layers, and then, plug fully connected layers at the end. Some extension layers, such as batch normalization (Ioffe & Szegedy, 2015) and dropout (Srivastava et al., 2014), have also been useful to accelerate convergence of the model and prevent overfitting (Cui et al., 2022).

In recent years, residual neural networks (ResNet) and their optimization are one of the hotspots in deep learning research that has emerged as a significant milestone in the field of CV (He et al., 2015; Xu et al., 2023). As the deep learning-based network evolves, its structure is deepening; while this helps the network to perform more complex feature pattern extraction, it may also introduce the problem of gradient disappearance or gradient explosion, which can lead to the following shortcomings: (1) Long training time but network convergence becomes very difficult or even non-convergent. (2) The network performance will gradually saturate and even begin to degrade, known as the degradation problem of deep networks (Feng et al., 2021). One of the key differentiators of ResNet compared to previous convolutional architectures is the utilization of residual connections, also known as "skip connections" or "identity shortcuts." These connections enable the direct transmission of the original input information to subsequent layers, bypassing certain intermediate layers. This helps address the problem of gradient degradation.

By incorporating residual connections, ResNets can be extremely deep, with hundreds of layers, while mitigating the performance degradation issue associated with deeper networks. Additionally, ResNets employ residual blocks, which consist of convolutional and activation layers, to learn feature representations at different levels of abstraction (Nijaguna et al., 2023).

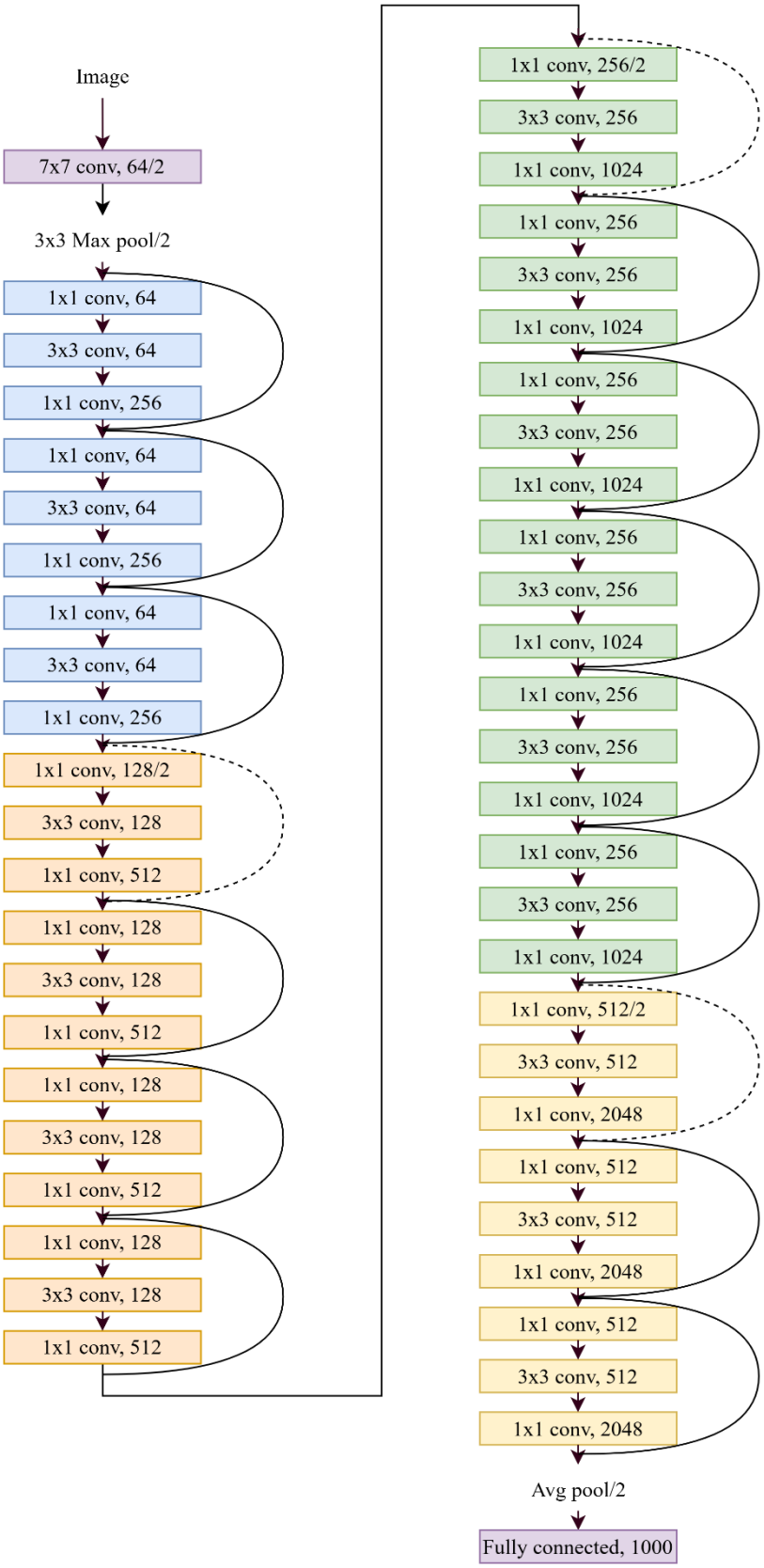
The ResNet architecture has been widely used and adapted for various CV tasks, including image classification, object detection, and semantic segmentation. Several ResNet variants, such as ResNet-18, ResNet-34, ResNet-50, ResNet-101, and ResNet-152, have been proposed, differing in terms of the number of layers and complexity(Du et al., 2022; He et al., 2015; Zhuang et al., 2022). These pre-trained variants, trained on large datasets like ImageNet (Deng et al., 2009), serve as a foundation for transfer learning in many applications. Specifically, ResNet50 consists of 50 layers in the architecture, as can be seen in Figure 2.

There are a series of convolutional (conv) layers in ResNet50 architecture. First, the conv layer is made of 7 × 7 kernel size and 64 different kernels with a stride size of

2. Then, 3 × 3 max pooling with the stride of size 2 is applied. In the next convolution, there are three conv layers (1 × 1, 64 kernels), (3 × 3, 64 kernels) and (1 × 1, 256 kernels), respectively, and these three layers are repeated 3 times. In the same process, three conv layers (1 × 1, 128 kernels), (3 × 3, 128 kernels) and (1 × 1, 512 kernels), respectively, are repeated four times; three conv layers (1 × 1, 256 kernels), (3 × 3, 256 kernels) and (1 × 1, 1024 kernels) respectively are repeated six times and other three conv layers (1 × 1, 512 kernels), (3 × 3, 512 kernels) and (1 × 1, 2048 kernels) respectively are repeated 3 times. Then, average polling (avg pool) is applied. Most of the hidden layers use Batch Normalization, and ReLU followed by a conv layer. The final layer of the original ResNet50 model is a fully connected (fc) layer with 1000 out-features (for 1000 class).

ResNet50 is much deeper and has fully connected layers. Without increasing the training error percentage, ResNet50 makes it straightforward to train networks with numerous layers. As the number of model layers increases, the consumption of computing resources becomes substantial, leading to the issue of the "disappearing gradient". Thus, ResNet provides a solution to the problem of gradient disappearance (Tian et al., 2022); this is one of the reasons why this model is used in this study. In addition, this study compared some algorithm models and selected ResNet-50 as the image recognition and feature extraction network as it achieved the best results in a series of experiments. Nevertheless, ResNet50 frequently needs longer training time as well (Nijaguna et al., 2023).

Figure 2 - Network architecture for ResNet50



Source: Adapted from (He et al., 2015) .

In addition, it is important to highlight that using ML/DL models for drowsiness detection requires a significant amount of labeled data and an adequate training process. Hence, it is necessary to collect a representative dataset containing EEG information and facial images from individuals in different alert states. Moreover, an integrated approach, combining EEG and CV data, allows the capture of information from different modalities, providing a more comprehensive and accurate view of the sleep state of individuals.

2.5.3. Automated Machine Learning

AutoML represents the automation of some or even all ML steps, such as preprocessing and cleaning the data, feature selection and feature engineering, model creation, model tuning, model ensembling, and model deployment, to cite a few steps (Goes et al., 2021). It is important to distinguish two AutoML approaches: Hyperparameter Optimization (HPO) and Neural Architecture Search (NAS). HPO automates finding a well-performing hyperparameter configuration of a given machine learning model on the dataset considered, including the kind of machine learning itself, its hyperparameters, and other data processing steps. Thus, HPO helps expert and non-expert personnel to avoid the laborious and vulnerable hyperparameter tuning process. For the HPO approach, we have used TPOT (Olson et al., 2016). On the other hand, NAS aims to find the architectural design of neural networks, optimizing the whole topology of the networks, for instance, the number of neurons, how to connect nodes, and which operators to choose. The user generally defines which metrics to optimize, the limiting model size regarding trainable parameters, or even the total time to search for an optimal architecture for specific applications. For the NAS approach, we used AutoKeras (Jin et al., 2023). Due to the extremely large search space, AutoML algorithms tend to be computationally expensive.

2.5.4. Quantum Machine Learning

QML is an emerging field at the intersection of quantum computing and machine learning. QML encompasses the application of quantum computing techniques to develop new algorithms and enhance existing ones for machine learning tasks (García et al., 2022). Understanding the basics of quantum computing is essential for comprehending the potential of QML. Quantum machine learning (QML) represents a novel intersection between quantum computing and machine learning, with the potential to overcome significant challenges faced by traditional methods. The fundamental properties of

quantum computing, such as superposition and entanglement, offer a robust approach for the analysis of complex signals, such as EEGs. Recent studies demonstrate that QML can improve signal classification and pattern detection in critical areas, such as the identification of epilepsy and other neurological disorders. In this context, this thesis proposes to explore the capabilities of QML for drowsiness detection, investigating whether this approach can offer advantages in terms of accuracy and speed in identifying states of reduced alertness in industrial workers.

The distinction between classical computing and quantum computing is crucial in addressing significant problem challenges. Quantum computing utilizes fundamental quantum properties like superposition, entanglement, and the measurement paradox to provide optimal solutions to complex problems. It also offers the advantage of computing much more powerful functions using number of quantum units, (i.e., qubits) and quantum gates.

In a Quantum Computer, the smallest unit of information is the qubit, which can exist in states of 0, 1, or a superposition of both (Y. Wang & Liu, 2022). Dirac's notation, known as bra/ket, is employed to represent quantum states and their transformations. The ket, represented by $|\psi\rangle$, has a dual called bra $\langle\psi|$ (Rieffel & Polak, 2011).

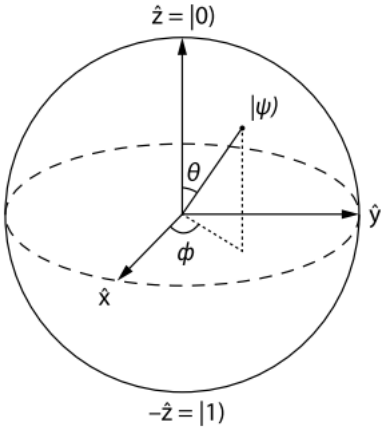
The state of a qubit can be expressed using $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, where α and β are complex numbers representing amplitudes, satisfying the relation $|\alpha|2 + |\beta|2 = 1$, and $|\alpha|2$ and $|\beta|2$ are the probabilities of a qubit collapsing to states "0" or "1" upon measurement. It can be rewritten as $|\psi\rangle = ei\gamma(cos\theta 2|0\rangle + ei\phi sin\theta\ 2|1\rangle)$, where θ, ϕ, and γ are real numbers (García et al., 2022).

The Bloch Sphere (Figure 3) serves as a geometric representation of the pure state space of a two-level quantum mechanical system, aiding in quantum mechanics and computation (Scherer, 2019). The angles θ and ϕ correspond to spherical coordinates representing a single qubit state. The vector $|\psi\rangle$ is defined as a Hilbert Space vector extending from the origin to the surface of the sphere, with dimensions in $\mathbb{R}^3$.

Additionally, the Bloch sphere visually demonstrates quantum state operations, with rotations representing unitary transformations and projections depicting measurements. It serves as a useful tool for understanding the impact of quantum gates on quantum circuits, facilitating the development and analysis of quantum algorithms and error correction techniques. Finally, the Bloch sphere illustrates entangled states between

two quantum systems, where their combined state is represented by a region on the sphere rather than a single point.

Figure 3 - Bloch Sphere which represents the qubit



Source: (Prakash, 2021)

2.5.5. EEG-based drowsiness detection

EEG is a brain imaging technique utilized to identify various mental states, including distinct emotions, varying levels of mental workload, stress, and fatigue (Cui et al., 2022). Many studies show the association between drowsiness and the oscillation patterns of EEG signals. As previously mentioned, some studies used different EEG channels as information sources to detect drowsiness. For example, W. Li et al. (2012) initially analyzed 16 EEG channels and, using Gray Relational Analysis (GRA) and Kernel Principal Component Analysis (KPCA), decreased the number of electrodes analyzed for the two most significant (Fp_1 and O_1). The drowsiness assessment model was based on a straightforward regression analysis. In another example, Yu et al. (2013) used a model based on support vector machine (SVM) for detecting drowsiness using 11 EEG channels, presenting an accuracy of approximately 95%. However, in practice, the use of multiple electrodes to collect many EEG signals may become invasive when monitoring the cognitive status of human operators, such as in safety-critical environments.

Indeed, single-channel analysis has inspired some research. For example, Garcés Correa et al. (2014) developed a method to detect the stage of drowsiness in EEG records using temporal, spectral, and wavelet analysis, from 19 features extracted from a single EEG channel. The method obtained 83.6% of correct detection rates for drowsiness. In

Belakhdar et al. (2016), the authors used features based on fast Fourier transform (FFT) to infer drowsiness based on EEG. Using a single signal from a differential EEG channel (i.e., C3-O1 electrodes), nine features are extracted from the power spectral density (PSD). Then, the performance of two well-known ML classifiers, SVM and MLP, was compared, with the latter presenting the best result of 83.5% accuracy. In another study, Belakhdar et al. (2018) also reduced the number of EEG features while still achieving a high accuracy level (i.e., 86%), considering the same differential EEG channel. However, this model trains and tests specific models for each person in the database, not addressing a general model.

Relying on the investigations mentioned, there is no consensus about which EEG channel is more appropriate to infer alert and drowsy states. Additionally, we mention that the performance of the drowsiness model is impacted by interindividual variability, and the selection of EEG channels also depends on electrode availability (see Figure 1).

Nevertheless, EEG signals provide large time-series sampled at a particular frequency, typically measured in hertz (Hz). Therefore, pre-processing techniques may be used to deal with complex series, creating more manageable data, even though still carrying important information (C. B. S. Maior et al., 2019).

Conventional methods for drowsiness recognition rely on feature extraction from EEG signals, which requires expertise and/or a priori knowledge of the data in order to model some special EEG characteristics of interest. For example, (Ogino & Mitsukura, 2018) compared three methods for feature extraction, which are Power Spectral Density (PSD), Autoregressive (AR) modeling, and Multiscale Entropy (MSE). They achieved a classification accuracy of 72.7% using the PSD features with Stepwise Linear Discriminant Analysis (SWLDA) for feature selection and SVM for classification. (da Silveira et al., 2016) performed Discrete Wavelet Transform (DWT) on single-channel EEG signals and selected the most significant M−terms from the wavelet expansion as features for classification. Considering that the alpha frequency contains personal information about age and memory performance of a subject, (B & Chinara, 2021) proposed a drowsiness detection model using time-domain features extracted by utilizing Wavelet Packet Transform (WPT).

In contrast to traditional approaches that involve crafting features manually, deep learning offers a distinct advantage by automatically capturing essential characteristics from the data. Models built on deep learning principles can directly learn a cascade of representations from raw and high-dimensional data, fine-tuning parameters through

backpropagation. Deep learning methods have received more attention in the field of EEG signal processing, including for drowsiness. For single-channel EEG signal classification, (Yıldırım et al., 2020) proposed a CNN model with 23 layers for automatic recognition. (Fahimi et al., 2019) explored an end-to-end deep CNN to detect the attentive mental state from single-channel raw EEG data.

2.5.6. CV-based drowsiness detection

In recent years, researchers have made significant strides in developing robust systems for detecting drowsiness, especially in contexts where CV is possible. Studies involving the detection of drowsiness from facial images are primarily in the context of car driving. For instance, (You et al., 2019) have proposed a drowsiness driving detection algorithm that considers individual driver differences. They built a deep cascaded convolutional neural network for the face region, and from the Dlib (Bhandari, 2014; King, 2009) toolkit, the landmarks of a frontal facial driver in a frame are found. In addition, they trained the classifier based on SVMs, which take as input the parameter called eyes aspect ratio. In simulated driving applications, the proposed algorithm detects the drowsy state of the driver quickly with 94.80% accuracy. Yang & Yi (2024) described a Deep Learning-based Intelligent Driver Drowsiness Detection for Advanced Driver-Assistance Systems (DLID3-ADAS) technique, which employs a combination of median filtering, feature extraction with ShuffleNet, hyperparameter optimization with the Northern Goshawk Optimization (NGO) algorithm, and final classification using Extreme Learning Machine (ELM). This system achieved an accuracy of 97.05% and a computational time of only 0.60 seconds, proving to be superior to traditional models.

Meanwhile, (Murthy et al., 2022) proposed a method that involves detecting the movement of the driver's face, eyes and head. Face detection is achieved using the Viola-Jones Algorithm, and other functions are built to invoke methods for eye detection. An SVM classifier is used to analyze eye movement. As a limitation, the authors argue that even if the algorithm detects face movement in all scenarios, it does not detect eye closure if the eyes are not fully visible in the video footage. Another method of detecting drowsiness based on images from driver cameras is seen in the study by Bhandari (2014). In this case, a camera was used to capture yawning characteristics (i.e., mouth shape and lip corner positions) to determine driver alertness. The main limitation of this system is that it is not able to function in poor lighting conditions. (C. B. S. Maior et al., 2020) address the need for vigilance in automated systems, focusing on operators' eye patterns

monitored through video streams. Employing CV and machine learning, the study proposes a real-time, non-intrusive system using a simple web camera. The model, validated on the DROZY database, demonstrates promising results, emitting warnings with 94.44% accuracy for drowsy subjects.

(Bekhouche et al., 2022) delve into the complex issue of driver drowsiness using a CV solution. The proposed framework detects the driver's face, extracts deep features through transfer learning, and applies temporal feature aggregation for drowsiness detection. Experiments on the NTHU Drowsy Driver Detection dataset showcase the framework's outperformance compared to existing approaches. (Rahman et al., 2022) introduce a real-time driver monitoring system utilizing CNN and CV techniques. Operating with an onboard camera module, the system achieves a 97.44% accuracy in fatigue detection, proving its efficacy in real-world scenarios.

Pandey & Muppalaneni (2023) work centers on road safety, employing a sizable drowsiness video dataset to develop two distinct models, emphasizing temporal and spatiotemporal characteristics. While Model-A, incorporating CV techniques and LSTM, exhibits a lower accuracy (86%), Model-B, focusing on spatial information and LSTM, achieves 97.5% accuracy. (Phan et al., 2023) approach integrates the Internet of Things (IoT) and deep neural networks to detect driver drowsiness efficiently. By considering multiple signs of fatigue, including facial landmarks, head tilting, blinking, and yawning, the model achieves an impressive accuracy of up to 98%. Incorporating transfer learning and time-varying factors enhances its robustness across diverse driving conditions, making it a comprehensive solution.

Although these studies contribute significantly to the field of drowsiness detection, it is essential to recognize certain limitations. Most of the research discussed focuses primarily on the context of drivers, and there is a notable absence of work considering attentional tasks such as in safety-critical operations control rooms. The complexities of vigilance and drowsiness may differ between these contexts. Furthermore, the absence of recent models, such as ResNet50, in studies raises questions about the application to contemporary approaches. These limitations highlight opportunities for future research to address existing gaps and improve the robustness and applicability of drowsiness detection systems in diverse environments.

2.5.7. Data fusion-based drowsiness detection

To take advantage of complementary information from different data sources, some data fusion strategies are proposed to concatenate data before classification. In the domain of data fusion-based fatigue detection, several studies have addressed the issue from diverse perspectives. Mehmood et al. (2023) focused on the construction industry, emphasizing that high-attention operations can induce mental fatigue, leading to inefficiencies and accidents. They proposed a novel approach using three machine learning models and multimodal data fusion, incorporating EEG, electrodermal activity, and video signals. The decision tree model, employing multimodal sensor data fusion, outperforms other models with 96.2% accuracy and 96.175%–98.231% F1 scores, showcasing the potential of multimodal sensor data fusion for real-time mental fatigue classification in construction sites.

Shahbakhti et al. (2023) delved into driver fatigue detection using multi-channel EEG, prioritizing the comfort of users by employing a single prefrontal EEG channel. Their method simultaneously analyzes EEG and eye blinks, utilizing an Fp1 EEG channel. The process involves identifying eye blink intervals using the moving standard deviation algorithm, filtering the EEG signal using discrete wavelet transform, and extracting features for classification. The proposed method exhibits robustness in detecting fatigue while driving, demonstrating promising results through the AdaBoost classifier.

L. Zhao et al. (2022) recognize the implications of fatigue on efficiency, especially in educational settings, where it can impact learning outcomes. Their study develops a fatigue detection system using a multimodal approach with ECG and video signals. The system classifies a learner's state into three categories: alert, normal, and fatigued. Tested on both the DROZY dataset and a self-collected dataset from a learning environment, the system achieves high detection accuracy, outperforming state-of-the-art approaches with 99.6% and 91.8% accuracy on the respective datasets.

Cheng et al. (2012) focus on driver drowsiness, a major cause of on-road accidents. They conduct a detailed study of abnormal eye behavior, steering wheel activity, and vehicle trajectory during different drowsiness stages. To overcome the limitations of single-sensor approaches, they developed a two-stage data fusion framework employing Fisher's linear discriminant and Dempster-Shafer evidence theory. The recognition

system achieves 90.7% accuracy, demonstrating the enhanced reliability and accuracy of the fusion method compared to single sensors.

In the context of data fusion-based fatigue detection, in addition to the substantial body of research focusing on driver-related drowsiness, at the best of author's knowledge, the fusion of video and EEG data for drowsiness studies remains an underexplored avenue, especially in the context of control rooms such as in an O&G industry. Furthermore, there is a notable gap in exploring more advanced models, such as quantum machine learning approach for EEG data, in the domain of drowsiness detection. This represents an opportunity for the proposed study to contribute new insights into the fusion of these two distinct data modalities (EEG and facial data), and investigate new models, thus increasing the comprehensiveness of drowsiness detection methodologies.

3. PROPOSED DROWSINESS DETECTION FRAMEWORK

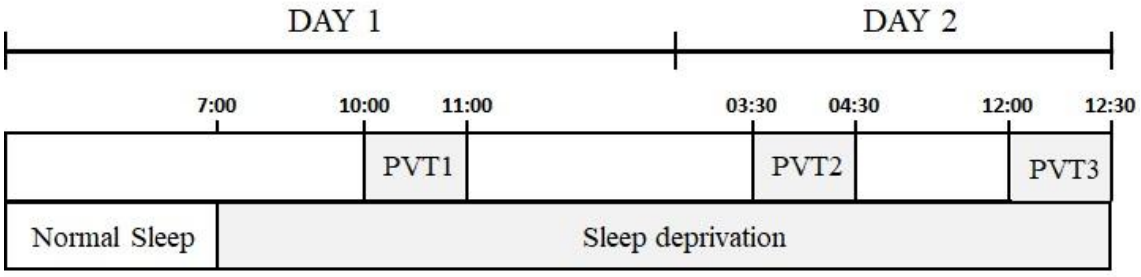
In this chapter, we present a comprehensive visual summary that encompasses the diverse approaches explored throughout this thesis. Figure 5 serves as a roadmap, describing the main components, methodologies, and applications that will be examined. This visual aid serves as a guide, providing a clear and concise visualization of the structure developed and tested in upcoming chapters. Before delving into specific applications, we provide a general overview of the DROZY dataset, which is a curated collection of multimodal data capturing various aspects of drowsiness and stands as a foundational element of our application.

3.1. Dataset

The "ULg multimodality sleepiness database", also called DROZY (Massoz et al., 2016), is a database that contains various types of data related to drowsiness, including EEG and webcam images. For the first case, the system recorded the Fz, Pz, Cz, C3, and C4 channels sampled at 512 Hz. For each video, a pair of near infrared (NIR) intensity and range images (I-image and R-image) are used. The NIR images are 512x424 pixels in size and recorded at 30 fps. In total, fourteen subjects/participants (3 males, 11 females) with a mean age of 22.7 years old and a standard deviation of 2.3 years old were considered (Massoz et al., 2016).

The experimental protocol considered that each subject performed three psychomotor vigilance tests (PVTs) during two consecutive days, in conditions of increased induced and prolonged sleep deprivation. After the subjects performed the first PVT, they were not allowed to sleep before the third PVT, resulting in a total sleep deprivation of 28 to 30 hours. Each PVT was around 10 minutes long. Figure 4 shows a measurement scheme for the three PVTs.

Figure 4 - Scheme for three PVTs in two consecutive days.



Source: Adapted from (Massoz et al., 2016.)

In DROZY, subjects were also asked to fulfill a Karolinska Drowsiness Scale (KSS) form, an established method to measure the subjective level of drowsiness at a particular period of the day (Kaida et al., 2006). The KSS consists of a nine-point scale with drowsiness rated 1 = extremely alert, 2 = very alert, 3 = alert, 4 = fairly alert, 5 = not alert or drowsy, 6 = some signs of drowsiness, 7 = drowsy, but no effort to keep alert, 8 = drowsy, some effort to keep alert and 9 = very drowsy, effort to stay awake (Waage et al., 2012). Here, for the binary classification, we consider two possible categorizations, which have been previously used in the literature and are also suitable for EEG: alert (KSS $\leq$ 3) and drowsy (KSS $\geq$ 7) (C. B. S. Maior et al., 2020; Ogino & Mitsukura, 2018; Sandberg et al., 2011), while for 3<KSS<7 the subject would not be alert or drowsy.

According to these categories, among the 14 participants in the database, only those who classified themselves as a KKS $\leq$ 3 for the first PVT and KKS $\geq$ 7 for the third PVT were selected for analysis. Thus, as shown in Table 2, 8 subjects classified themselves as alert in the first PVT and 10 subjects classified themselves as drowsy in the third PVT. Moreover, there are specific subjects (i.e., subjects 1, 5, 6, 8, and 10) that alternate the classification of alert to drowsy from the first PVT to the third PVT, which enables a direct analysis for those in the two tests. Other subjects did not have the third PVT performed or it was not recorded.

Table 2 - KSS classification for the selected subjects

Subject	PVT 1 (KSS)	PVT 3 (KSS)
1	3	7
2	3	-
3	2	-
4	-	9
5	3	8
6	2	7
7	-	9
8	2	8
9	-	8
10	3	7
11	-	7
12	2	-
13	-	-

14	-	8

Source: Adapted from (Massoz et al., 2016.)

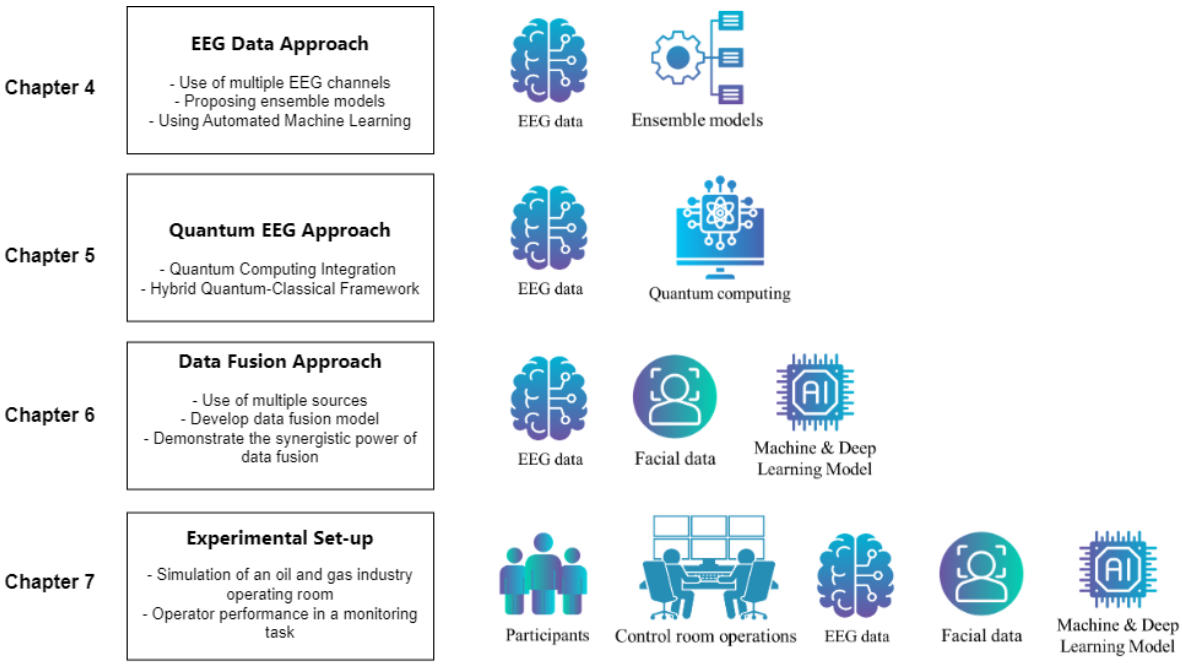
In addition, the EEG signals of all participants were grouped as a ‘single-subject’, as a generalist representation in a sequential manner. In this case, the EEG channel signals belonging to each participant eligible for PVT1 (8 subjects), were grouped as single-subject channels and assigned as alert. Likewise, the signals from 10 subjects in PVT3 were designated as drowsiness signals for that ‘single-subject’. From now on, we call this “single-subject” as ‘multiple subject’, since it gathers available signals from several subjects participating in the experiment. For further descriptions of these subjects (and PVTs) as well as the previously mentioned selection, see (C. B. S. Maior et al., 2020).

During each PVT, a light was randomly turned on every few seconds and the subject should press a button as soon as the light appears; then, their reaction times (RT) were registered. Thus, for each subject and for each PVT, the database contains the raw data perfectly time-synchronized with KSS data, PVT data (including RT), polysomnography signals (PSG) (e.g. EEG and EOG) and NIR intensity and range images of the face. We will use video and EEG data in our study.

3.2. Methodology framework

The methodology framework of this study is structured into four main parts, each dedicated to a distinct aspect of the research objectives. This section outlines the comprehensive framework that guides the upcoming chapters, summarizing the essence of the study.

Figure 5 - Thesis methodologies framework



Source: The Author (2024).

1. EEG Data Approach

Chapter 4 focuses on the detailed exploration of EEG data for effective drowsiness detection. The main components include:

- Use of multiple EEG channels: Employs a multichannel EEG setup to capture a comprehensive view of brain activity.
- Proposition of Ensemble Models: Use of two robust models through the integration of multiple machine learning algorithms, increasing predictive power.
- Use of AutoML: Application of AutoML techniques for efficient model selection and hyperparameter tuning, streamlining the model development process.
- Integration of RT data with EEG patterns: EEG data is aligned with the duration of each RT, predicting lapses and non-lapses based on EEG features at each time point.

This part represents a significant advance in understanding drowsiness patterns using EEG data, exploring ensemble methodologies and incorporating automation for model refinement.

2. Quantum EEG Approach

In Chapter 5, building upon the insights gained from the EEG Data Approach, we delve into the innovative Quantum EEG approach, which leverages quantum computing principles for drowsiness detection. Key aspects of this approach include:

- Quantum Computing Integration: Incorporation of quantum computing techniques to process EEG data, offering the potential for exponential computational speedups and improved pattern recognition capabilities.
- Quantum Feature Extraction: Exploration of quantum algorithms for feature extraction from EEG signals, aiming to capture subtle patterns that are indicative of drowsiness with higher fidelity.
- Quantum Machine Learning Models: Development of quantum machine learning models tailored for EEG data analysis, utilizing quantum circuits and algorithms for classification tasks.
- Hybrid Quantum-Classical Framework: Adoption of a hybrid quantum-classical framework to combine classical machine learning methods with quantum processing, optimizing the balance between computational resources and model accuracy.

This pioneering approach represents a frontier in drowsiness detection research, harnessing the power of quantum technologies to unlock new insights and capabilities in EEG data analysis.

3. Data Fusion Approach

In Chapter 6, based on the knowledge gained from the EEG-focused approach, we introduce the concept of data fusion for a more comprehensive drowsiness detection system. The main components include:

- Approach with three distinct models: one for image data analysis, another for EEG data analysis and a third model for data fusion. A flow of the methodology is shown in Chapter 5 highlighting the sequential steps of the approach.
- Training Data: All models are trained using labeled data generated in the database itself, specifically the alert and drowsy states indicated by PVT1 and PVT3.
- Data Processing: For processing data, it is initially separated, pre-processed and transformed into a manageable format suitable for input into learning algorithms.

- Classifier Development: Classifiers are developed using Multilayer Perceptron (MLP) and the ResNet50 pre-trained model. These classifiers are trained on the extracted data for two main classification tasks: i) drowsiness detection from EEG data and ii) drowsiness detection from images.
- Decision Rule: A decision rule is implemented to combine the results of individual classifiers, with the aim of improving overall prediction performance.

This phase represents an improvement over the previous chapter by combining EEG data and CV, demonstrating the synergistic power of data fusion in improving accuracy and reliability.

4. Experimental Set-up

Chapter 7 shifts the focus to practical experimentation in the control room domain of the O&G industry. That includes:

- Simulation of an O&G industry operating room: Creating a simulated environment representative of real-world conditions in the O&G industry control room.
- Operator Performance in Monitoring Task: Evaluate operator performance through monitoring tasks within the simulated environment, allowing the observation of human responses and behaviors.

This applied phase takes the theoretical advances from the previous sections and tests them in an industry-relevant scenario, offering insights into real-world applicability and effectiveness. Thus, in summary, the methodology framework progresses from an in-depth analysis of EEG data and machine learning, to an exploration of data fusion techniques that incorporate EEG and facial data. Finally, it ends in a practical experience simulating an O&G industry operation, demonstrating the holistic application of the methodologies developed to improve safety and performance. Each step distinctively contributes to the overall goal of advancing safety measures in critical industrial environments.

4. MODELS FOR AUTOMATIC DROWSINESS DETECTION USING EEG SIGNALS

Part of this chapter has been published in Process Safety and Environmental Protection journal (Ramos et al., 2022). This chapter centers on advancing the identification and prediction of operator drowsiness in critical safety systems, such as the O&G industries. We aspire to extract more nuanced and accurate information critical for improving accident prevention measures by EEG patterns. The primary objective is to contribute to accident prevention by developing a robust, multi-source model capable of effectively recognizing diverse features associated with drowsiness. Our focus extends from ensemble machine learning (Ramos et al., 2022), to Automated Machine Learning (AutoML) approaches and advanced deep learning models.

The chapter also conducts an in-depth analysis, leveraging results from the paper published in the Expert Systems with Applications journal (C. B. S. Maior et al., 2020), mainly focusing on reaction times. A unique aspect of our approach involves creating a ground truth for reaction times by aligning EEG data with specific time windows, offering a comprehensive understanding of the temporal dynamics associated with drowsiness.

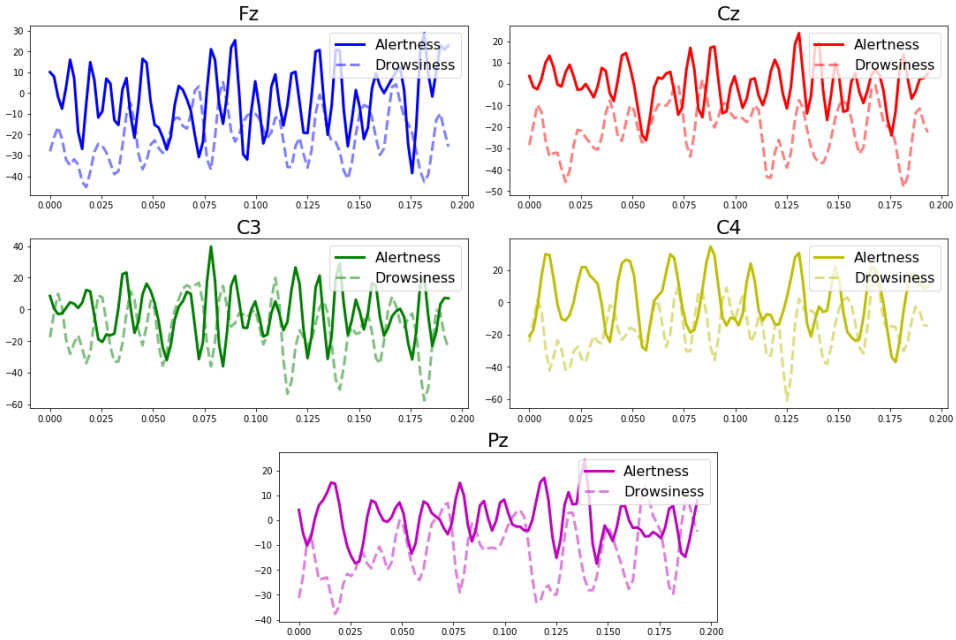
4.1. Proposed Bagging based model

Here, we adapted the classical bagging model to take the different sub-samples monitored synchronously from the same subject in different EEG channels. Bagging, or Bootstrap Aggregating, is an ensemble approach commonly used to increase the accuracy of results in classification problems. Initially developed by Breiman (1996), this technique generates diversity through the creation of subsets of data, known as bootstrap samples. These samples are constructed from the training set, involving replication and omission of some data. Using the same classification algorithm, base learners are generated from these samples and, finally, are combined using strategies such as majority voting.

Thus, it is possible to create a new hybrid model, coupling a single classifier with different sub-samples (i.e., EEG channels) to detect drowsiness. Initially, our model considers five EEG channels (Fz, Cz, C3, C4, and Pz) available in the database. Note that the signal has different behaviors and frequencies, as illustrated in Figure 6. The signals from each channel are separated into training and test sets for the five subjects for which the first and third PVTs are available (i.e., 1, 5, 6, 8, and 10). The remaining subjects,

who have only the first or third PVT available (recall Table 2), are grouped in the test data for their respective states of alertness or drowsiness. As the proposal is a real-time drowsiness detection, different sizes of signal segments were tested from raw data that were equal to or smaller than 1s, but maintained important characteristics.

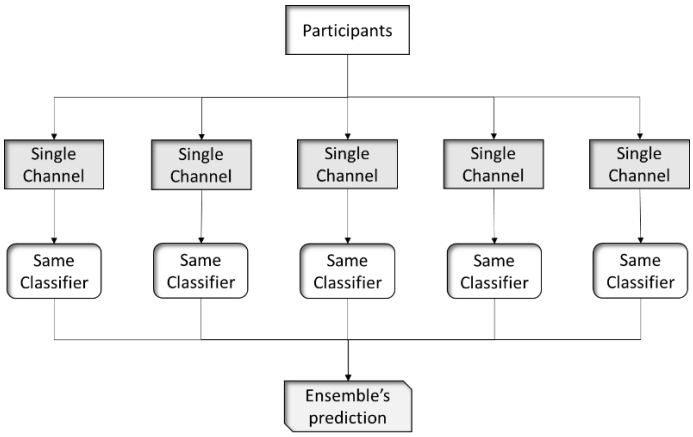
Figure 6 - Difference in signals for the five EEG channels available



Source: The author (2024).

As the ensemble model requires different sub-samples but only a single classifier, the four ML classifiers discussed in section 2.5.1 (KNN, RF, SVM, MLP) were tested in order to verify which one presents the best accuracy. Furthermore, to make the approach less intrusive, different schemes were developed with different numbers of channels in this ensemble model. Figure 7 shows the scheme of the bagging-based ensemble model with five channels and Figure 8 presents the pseudo-code for this scheme.

Figure 7 - Scheme of the proposed bagging model



Source: The author (2024).

Figure 8 - Algorithm for scheme of bagging model

Input:
- EEG channels signals of 512Hz each .
- Function to separate vectors with EEG data into smaller vectors for analysis.
- Functions with classifiers algorithms.

for channel in channels
    if (features):
        feature extraction: 'HDF', 'Complexity', 'Mobility'.
    end
    Separates training and testing data;
    Train the ML model;
    Test the ML model;
    Predictions: Keep the prediction of each EEG channel.
end
Bagging-based model (Predictions):
    Hard voting.
Output:
- Accuracy of the ensemble model.

Source: The author (2024).

Figure 9 explains how the proposed bagging ensemble model works. First, the segments of each EEG channel containing a specific size of values, from a sampling frequency, are trained in the ML model. Through a simple vote, the ensemble model combines the responses of the models of each previously executed channel. The result combines the most votes for a given class, thus producing its final prediction. This model is further tested with the test sets and, finally, the accuracy is measured.

Figure 9 - General example of how the proposed bagging-based works

	Predicted Values							
	S_1	S_2	S_3	...	S_{n-2}	S_{n-1}	S_n	ACC(%)
True Value	0	1	0	...	1	0	1	
Model-FZ	1	0	1	...	0	0	1	85.25
Model-CZ	1	1	1	...	0	0	1	90.67
Model-C3	0	1	1	...	1	0	0	92.20
Model-C4	0	1	0	...	1	0	1	97.08
Model-PZ	0	0	0	...	0	1	1	91.54
Final Pred	0	1	1	...	0	0	1	93.78

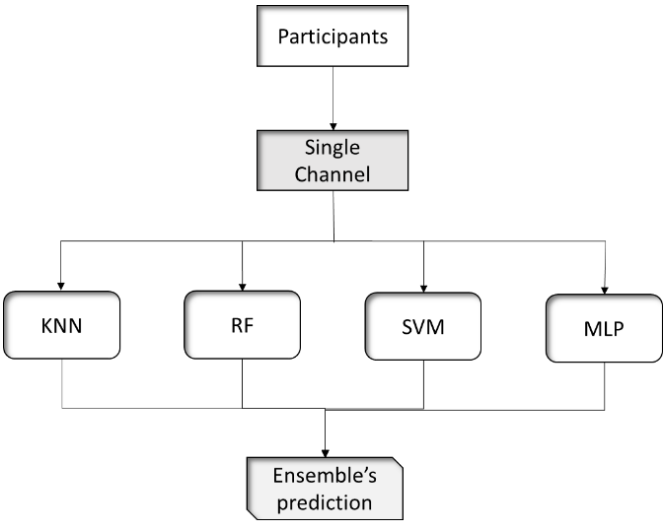
Source: The author (2024).

The set of these classifiers considering different EEG channels makes the bagging-based model a strong classifier with low bias and low variance. Furthermore, bagging has been reported to consistently deliver good performance despite its simplicity (Hassan & Haque, 2016).

4.2. Proposed voting model

Here, the ensemble model considers hard voting using a single set of signals but distinct classification methods trained and tested in all five EEG channels available. As in the previous model, the training data of each subject are also grouped with the other data of subjects who had only one type of state. For voting, the method takes a conservative approach (i.e., if at least two out of the four models classify drowsiness, the ensemble model will classify the signal segment as drowsiness). Figure 10 below shows the scheme of the model described.

Figure 10 - Scheme of the Proposed voting model



Source: The author (2024).

Figure 11 and Figure 12 show the algorithm and a fictitious example of the voting model, respectively.

Figure 11 - Algorithm for scheme of voting model

Input:
- Single-channel of EEG signals of 512Hz each.
- Function to separate vectors with EEG data into smaller vectors for analysis.
- Functions with classifiers algorithms.

for model in models:
  if (features):
    feature extraction: 'HDF', 'Complexity', 'Mobility'.
  end
  Separates training and testing data;
  Train the EEG channel;
  Test the EEG channel;
  Predictions: Keep the prediction of each ML model.
end
Voting model (Predictions):
  Hard voting.
Output:
- Accuracy of the ensemble model.

Source: The author (2024).

Figure 12 - General example of how the voting model works

	Predicted Values							
	S_1	S_2	S_3	...	S_{n-2}	S_{n-1}	S_n	ACC(%)
True Value	0	1	0	...	1	0	1	
Channel-KNN	1	0	1	...	0	0	1	85.25
Channel-RF	1	1	1	...	0	0	1	90.67
Channel-SVM	0	1	1	...	1	0	0	92.20
Channel-MLP	0	1	0	...	1	0	1	97.08
Final Pred	1	1	1	...	1	0	1	93.78

Source: The author (2024).

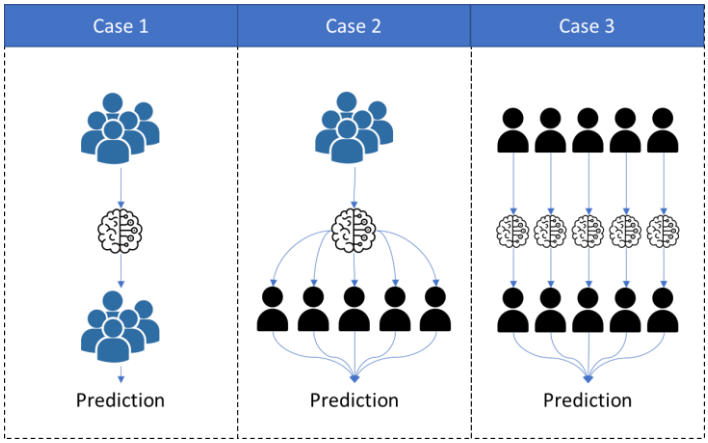
4.3. Results and discussion for the ensemble models

In this study, both ensemble models, Bagging and Voting, were executed considering two representations of the EEG signals: (i) raw data and (ii) the three time-domain EEG features (HDF, complexity, Mobility) presented in Section 2.5.1. Numerous tests were performed to identify the most suitable ML for the bagging mode and the most suitable channel for the voting model. For the sake of brevity, the results presented are for the best-performed one, achieved for the MLP classifier for the bagging case and for the C4 channel for the voting case. To consider real-time drowsiness detection, an EEG time-window containing 100 points was selected, which, at a sampling frequency of 512 Hz, represents <0.2 s. All the experiments were run on a GPU running Python Version 3.9 with a 3.6 GHz Intel CORE i9-9900K processor, 32 GB of RAM .

The results are presented in three parts, as in Figure 13. Case 1 contains information from all subjects, in which, after grouping the data from the alertness and drowsiness categories, they are separated into training and test data to verify if the representations in the time-domain and the ML techniques are appropriate for this detection. Case 2 makes use of this generic training model, containing information from all subjects but tested on data from the five specific subjects discussed in the Chapter 3. Finally, in Case 3, the data from the five specific subjects are separated and tested individually, in which a dedicated model for each subject is created. For the sake of brevity, we only present Case 3 in this chapter, the results for Cases 1 and 2 will be in the APPENDIX A – Case 1: All subjects and

APPENDIX B - Case 2: Specifics subjects with a general model.

Figure 13 - Scheme of the three Cases



Source: The author (2024).

4.3.1. Parameter tuning

Each ML algorithm requires a set of hyper-parameters in its formulation, and their proper estimation demands attention to best adjust the mapping function. Thus, we here applied the grid-search (GS) algorithm to fine-tune their hyper-parameters. GS looks into a specified searching space (a subset of hyper-parameters). Table 3 presents the search spaces for GS approach for each ML model; see (Pedregosa FABIANPEDREGOSA et al., 2011) for further details).

Table 3 - Search space for GS for each classifier and description of the hyper-parameters.

Classifier	Hyper-parameters	Value / Type range
SVM	gamma	[10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}]
	C	[10, 10^{2}, 10^{3}, 10^{4}]
MLP	hidden layers size	[(100), (100, 100), (100, 100, 100)]
	activation	[‘tanh’, ‘relu’]
	solver	[‘sgd’, ‘adam’]
RF	max depth	[20, 30, 40, 50]
	n estimators	[50, 83, 116, 150, ‘None’]
KNN	n neighbors	3, 5, 10
	weights	['uniform', 'distance']
	algorithm	['auto', 'ball_tree', 'kd_tree', 'brute']
	metric	['euclidean', 'manhattan', 'minkowski']

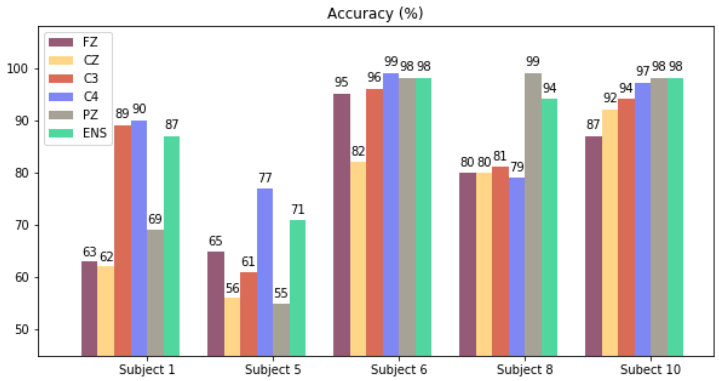
Source: The author (2024).

4.3.2. Specific subjects with dedicated model

- Raw data

The third case considers data from the five specific subjects trained and tested individually, creating a dedicated model for each of them. For the bagging-based method (Figure 14), subject 5 presents the lowest accuracy. This may be explained by the low accuracy from some channels (Fz and Cz), which can reduce performance in the final prediction, as can also be seen in subject 1. For subject 8, even though few channels have accuracy near 80%, the final prediction of the ensemble model was above 90%.

Figure 14 - Bagging-based model for five specifics subjects in a dedicated model from the MLP technique with raw data and five channels



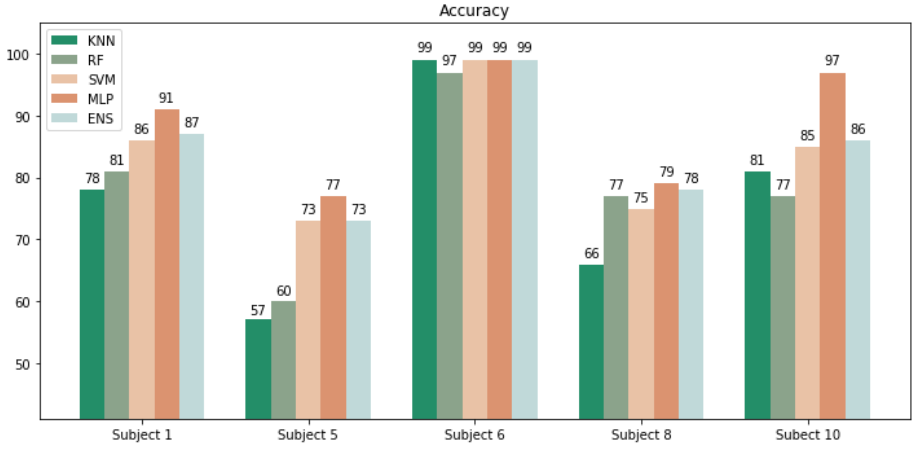
Source: The author (2024).

Therefore, the remaining 3-channel (C3, C4 and PZ) training is performed, and the results are shown in the

APPENDIX C - Case 3: Specifics subjects with dedicated model. In this case, almost all ensemble predictions presented an accuracy above 90%, achieving better performance than the single-channels in nearly all subjects, except for subject 5 (a slightly lower accuracy compared to the single-channel C4) and subject 8 (accuracy of MLP of the single-channel Pz was superior to the bagging-based model). Compared with the previous five-channel bagging-based model, the accuracy of almost all data samples was surpassed by the three-channel model, except for subjects 5 and 8, which presented a similar result.

For the voting model with raw data (Figure 15), when compared to the results of the bagging-based model, the accuracy was slightly lower, being below 80% for subjects 5 and 8. However, the result is somehow expected as the models are fed by data from only a single-channel (C4) and less information is provided, which is, therefore, a satisfactory result.

Figure 15 - Voting model for five specific subjects in a dedicated model from the C4 channel with raw data

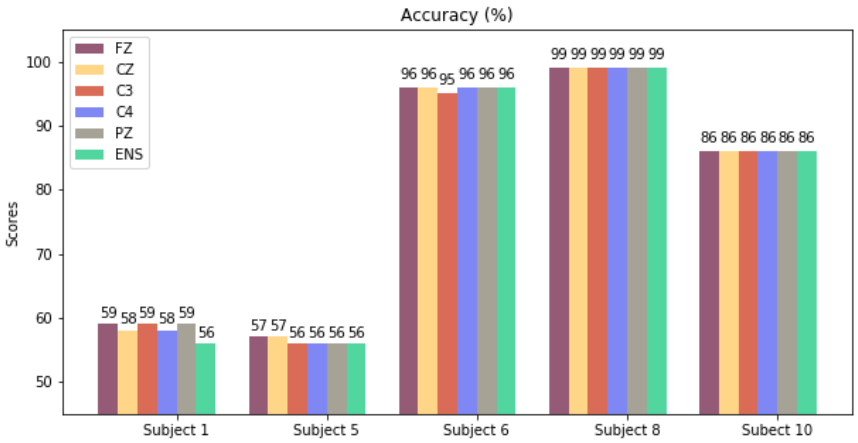


Source: The author (2024).

- EEG features

The same processing was applied to the three features. In Figure 16, subjects 1 and 5 present the poorest results in all channels, while subject 8 provided an accuracy of 99% in the five available channels. This demonstrates the interpersonal variability, considering that even the appropriate EEG features cannot always accurately classify between awake and drowsy. Indeed, this interpersonal variability is in close agreement with the results (C. B. S. Maior et al., 2020), who mentioned that even though the subject 1 is drowsy, it remained quite attentive, with only one lapse (i.e., reaction time greater than 0.5 seconds) during the whole PVT.

Figure 16 - Bagging-based model for five specific subjects in a dedicated model from the MLP technique with features extraction and five-channels



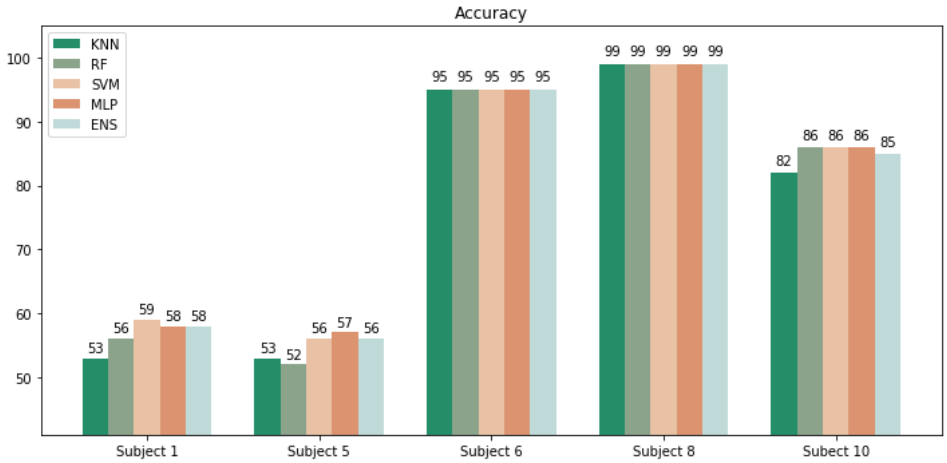
Source: The author (2024).

Then, the channel reduction process was replicated, performing the three-channel model and similar results were obtained when compared to the five-channel model. In fact, the improvement was only made in reducing the number of channels required, and the results can be seen in the

APPENDIX C - Case 3: Specifics subjects with dedicated model. Similar results were obtained for the voting model when compared to the bagging-based model for the three features. Furthermore, we realized that for the three feature extractions, regardless of the ML technique or EEG channel, accuracy achieves similar results. The performances are shown in Figure 17.

However, when comparing the models considering the raw data, the results are clearly better in the latter, especially when considering the raw data through the three-channel model. Nevertheless, even if the ensemble models do not benefit because the single model presented similar performances, the adoption of features can be useful if you consider the use for some specific subjects (e.g., 6, 8 and 10). In addition, the presented results suggest that it is possible to extract significant information even within a short period of time (~ 0.2 s). Finally, it is noteworthy that the superior performance of bagging for our presented scheme benefits from the extraction features dedicated to EEG signals and the consideration of diverse signals as different samples when compared to traditional bagging.

Figure 17 - Voting model for five specific subjects in a dedicated model from the C4 channel with features extraction



Source: The author (2024).

4.4. AutoML approach

In this section, we provide information about three different tools used. Apart from a vanilla implementation of MLPClassifier() from scikit-learn, we used two AutoML APIs: TPOT and AutoKeras. Overall, their usage for classification tasks is straightforward. TPOT uses a generative algorithm to achieve the best pipeline. The most relevant parameters for the TPOTClassifier() are:

- Generations: The number of iterations needed to complete the pipeline optimization procedure. Default value: 100;
- Population size: Number of individuals to keep in the genetic programming population each generation. Default value: 100;
- Offspring size: Number of offspring to be born in each generation of genetic programming. Default value: 100;
- CV: Number of folds to evaluate each pipeline over k-fold cross-validation during the optimization process. Default value: 5;
- Early stop: Maximum number of generations to check whether there is no improvement in the optimization process. Default value: None;
- Scoring: Function used to evaluate the quality of a given pipeline in solving the task. Default value for classification: ‘accuracy’.

AutoKeras is an open-source software library for AutoML. It is developed to make it easier for users to apply deep learning techniques to their own datasets by providing a high-level interface for building and training machine learning models. AutoKeras is built on top of the popular deep learning framework Keras, which is a Deep Learning Application Programming Interface (API).

One of the main advantages of AutoKeras is that it is designed to be user-friendly and requires minimal knowledge of machine learning to use. It allows users to specify their own dataset and the type of model they want to build. Then, it uses an automated search algorithm to find the best model architecture and hyperparameters for the given dataset without having to specify each one manually. We used the AutoKeras StructuredDataClassifier() tool whose most important arguments are:

- num classes: The number of classes. If None, it will be inferred from the data;
- loss: A Keras loss function. Default: ‘binary cross-entropy’ or ‘categorical crossentropy’ based on the number of classes;
- metrics: A Keras metrics. Default: ‘accuracy’;
- objective: Name of model metric to minimize or maximize. Default: to ‘val accuracy’ (computed in the test set);
- seed: An integer seed number, important for reproducibility;
- max model size: Maximum number of scalars in the parameters of a model.

4.4.1. Results and discussions for AutoML

Ramos et al. (2022) performed numerous tests to identify the EEG channel that is most suitable for drowsiness inference for DROZY database, and based on the findings, channel C4 was selected to discuss the results. A 100-point EEG time window was used for real-time drowsiness detection, for which 512 Hz sampling frequency corresponds to ~ 0.2 seconds. Thus, as the DROZY database comprises 18 videos of 10 minute-length, each has 307,200 data points, and overall, there are over 5.5 million data points available. Each time window feature vector has an associated label (either 'drowsy' or 'alert'); the same occurs for each feature vector. Data were split into 80/20 for training/testing (i.e., 2456 points for training and 614 for testing).

After taking the same MLP application, found in (Ramos et al., 2022), as a baseline, we sought AutoML solutions with TPOT and Autokeras. We discuss the results in terms of performance, cost of implementation, and training duration for both TPOT and AutoKeras. For each subject, the comparison contains (i) raw data (RD), (ii) feature vectors (F), and (iii) the combination of both (F+RD). In addition, we adopted two approaches: (a) training a specific model for each subject separately; and (b) training a unique model with all subjects together, as explained in chapter 3 (i.e., five dedicated models and one generalist).

In Table 4, there are the obtained accuracy values for all models (MLP, TPOT, and AutoKeras), input data arrangement (RD, F, F+RD) and subject approach – dedicated models for each subject and a generalist one (All). The best performance by subject is highlighted in gray. We noticed that there is no consensus on the best data arrangement, although F+RD is better in four out of six subjects (1, 6, 10, All).

Table 4 - MLP and AutoML results, in %; RD: raw data, F: Features; best performance per subject highlighted in gray

Subject	MLP			AutoML					
				TPOT			AK		
	RD	F	F+RD	RD	F	F+RD	RD	F	F+RD
1	91.30	58.77	91.29	93.00	57.85	**95.36**	94.06	59.48	94.79
5	75.85	56.97	79.02	77.54	56.14	80.07	**82.83**	57.53	76.89
6	98.69	95.13	99.10	99.59	95.52	**99.92**	99.67	95.85	**99.92**
8	77.15	98.13	90.89	82.42	99.27	98.54	85.53	**99.43**	99.27
10	95.54	86.90	96.66	94.55	86.25	97.64	98.13	87.22	**98.37**
All	66.33	63.40	67.93	75.78	63.92	79.70	77.21	64.29	**82.54**

Source: The author (2024).

Still in Table 4, subjects 1 and 5 had the poorest results when considering features (F) exclusively, regardless of the model used. Indeed, this finding is in close agreement with the results of (C. B. S. Maior et al., 2020), where it was noted that subject 1, despite being drowsy, remained relatively effective throughout the entire PVT. This could potentially account for the low accuracy observed for this subject. Subject 8, in turn, had an accuracy of around 99% for F, regardless of the model employed. This variation in performance among individuals demonstrates the presence of interpersonal variability, highlighting the challenge of accurately classifying between awake and drowsy states based solely on specific EEG features.

Additionally, the best outcomes belong to AutoML models, mainly AutoKeras, which provided better results in four out of six subjects (5, 8, 10, All) presenting in Table 5. In fact, using algorithms to automatically identify the best combination of ML approaches and hyperparameters for the EEG dataset made the models stand out when compared with the MLP models.

Table 5- Best pipelines. In the notation (i, o), i and o represent the input and output sizes, respectively

AutoML	**Subject**	**Input**	**ML model**	**Main Parameters and Hyperparameters**
TPOT	1	F + RD	MLP	1 layer, 100 neurons, lr = 0.001
AK	5	RD	ANN	Normalization (103,103) Dense (103,1024) ReLU Dense (1024,1)
TPOT	6	F + RD	MLP	1 layer, 100 neurons, lr = 0.001
AK	8	F	ANN	Normalization (3,3) Dense (3,16) BatchNormalization (16,16) ReLU Dropout Dense (16,1)
AK	10	F + RD	ANN	Normalization (103,103) Dense (103,1024) BatchNormalization (1024,1024) ReLU Dense (1024,1)
AK	All	F + RD	ANN	Normalization (103,103) Dense (103,256) ReLU Dropout Dense (256,256) ReLU Dropout Dense (256,64) ReLU Dropout Dense (64,1)

Source: The author (2024).

In addition, the presented results reveal interesting information about the performance of EEG resources in the classification between awake and drowsy states. It suggests that it is possible to extract significant information even within a short period of time (~ 0.2 s), demonstrating the potential of the EEG as a tool to assess cognitive states. These insights have implications for developing EEG-based systems for real-time monitoring of cognitive states, which may be useful in several domains beyond those here explained.

In a binary classification problem for risk situations, the accuracy metric (percentage of correct predictions made by the model relative to the total number of predictions) is informative, but the incidence rate of type I and type II errors must be investigated. A type I error, also known as a false positive, happens when the model predicts that an event will occur (e.g., a subject is drowsy), but the event does not actually do (the subject is not drowsy). On the other hand, a type II error, also known as a false negative, arises when the model predicts that an event will not happen (e.g., a subject is not drowsy) but the event does actually do (the subject is drowsy). However, despite minimizing both types of errors would be ideal, there is often a trade-off between them, meaning that they cannot be simultaneously minimized. The optimal balance between type I and type II errors depends on the specific application and their relative costs.

Confusion matrices are used to evaluate the overall performance of classification ML models; they summarize the number of correct and incorrect predictions and enable calculating various performance metrics such as accuracy, precision, recall, and F1 score. Table 6 shows the metrics for the F+RD case. The highlight is the high precision and recall obtained for both classes in all subjects. This indicates that the model was successful in accurately identifying both awake and drowsy states and in recovering the majority of true positive cases for both classes. However, it is important to note that results vary between different subjects. For example, subjects 6 and 10 show particularly high results in terms of precision, recall and F1 score for both classes, indicating robust and reliable detection of drowsiness and wakefulness. On the other hand, subject 5 shows slightly lower results compared to the other subjects, especially in terms of accuracy for the drowsy class.

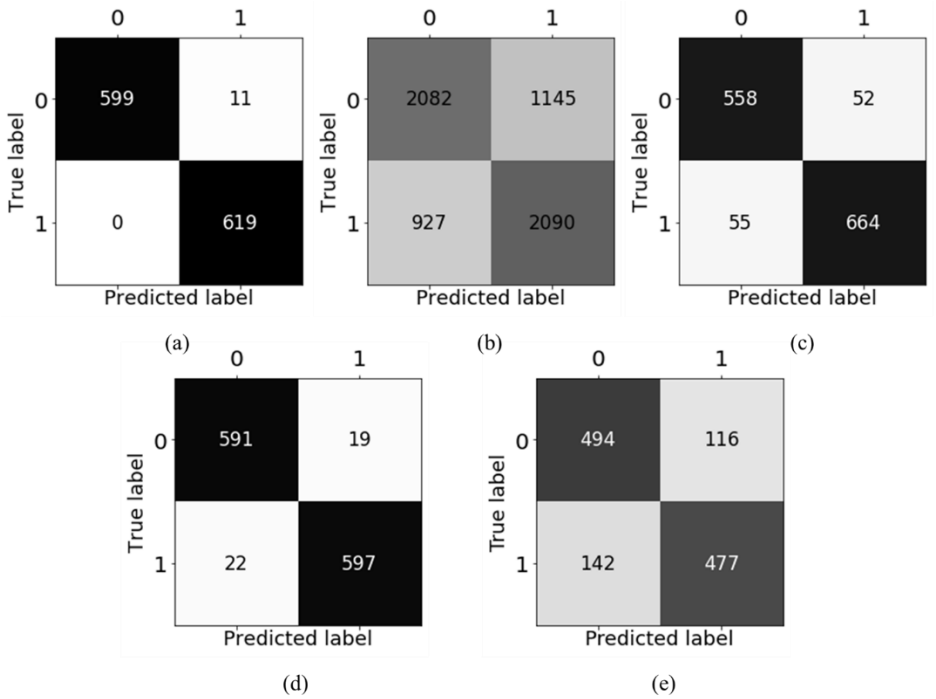
Table 6 - Metrics per subjects with F+RD inputs

Subjects	Label	Precision	Recall	F1 score
1	0	91%	91%	91%
	1	91%	91%	91%
5	0	78%	81%	79%
	1	80%	77%	79%
6	0	100%	98%	99%
	1	98%	100%	99%
8	0	90%	92%	91%
	1	92%	90%	91%
10	0	96%	97%	97%
	1	97%	96%	97%
All	0	69%	67%	68%
	1	67%	69%	68%

Source: The author (2024).

In our case, type I error suggests an unneeded use of human and time resources. For example, as a consequence of a false positive, a manager would unnecessarily check the actual operator status, or the operator could have their attention diverted while performing a task. The type II error, in turn, is the most important situation in the drowsiness detection context. It represents conditions where the lack of an alert emission leads to the drowsy operator's unawareness not only about its own state but also about the tasks they should perform, which may lead to errors and even an accident. In fact, we noticed that the MLP model achieves good results when considering F+RD for subject 6, and all subjects, as we can see in the confusion matrices (a) and (b), respectively (Figure 18). The drowsiness prediction is fairly accurate.

Figure 18- Confusion matrices per subject (1, 5, 6, 10, and all, for (c), (a), (e), (d), and (b), respectively); all cases with F + RD as input.



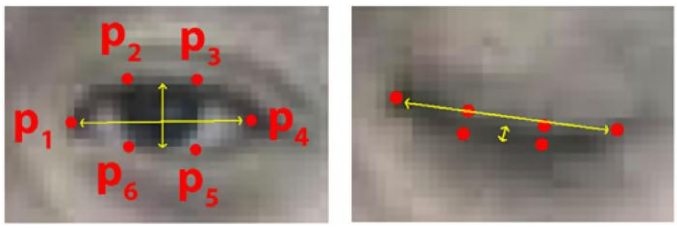
Source: The author (2024).

For subjects 1, 5, 8, and 10, the Type II error was slightly higher when compared to the Type I error; however, there is no notable difference, as we can see in Figure 18 the confusion matrices (c) and (d) that show the confusion matrices of subjects 1 and 10, respectively, for F+RD as input.

4.5. Integrating Reaction Time and EEG patterns

In the study conducted by C. B. S. Maior et al. (2020), a drowsiness detection model was proposed based on the Eye Aspect Ratio (EAR), which is a variable determined by the proportion between the width and height of the eye, computed from eye landmarks (Figure 19). This method provided a low computational cost for extracting eye-based variables, allowing for efficient inference. Specifically, EAR was employed to identify blinks and assess eye openness. To establish a ground truth, EEG data was synchronized with reaction time (RT) data in each PVT, categorizing lapses (RT >= 0.5s) and non-lapses (RT < 0.5s). Subject 6 was chosen for this integration, exhibiting the highest number of lapses (40) and non-lapses (52) during PVT 3.

Figure 19 - Eyes landmarks



Source: (Soukupova & Cech, 2016)

For the long blink detection model, (C. B. S. Maior et al., 2020) employed a threshold method, defining fixed limits for EAR. If the user's EAR fell below the threshold for a specific number of frames, a blink was detected. An alternative method involved concatenating EARs from 15 consecutive frames, creating a user state feature classified as open eye, short blink, or long blink. The model was tested on the DROZY dataset, where subjects performed PVTs with randomized light stimuli. Lapses, defined as reactions exceeding 500 msec, were primary outcomes. The model used real-time video input to detect drowsiness, with specific analysis focusing on transitions from alertness to drowsiness states in subjects, exemplified by subject 6 shown in Figure 20.

Figure 20 - Subjective and objective drowsiness-related metrics for subject 6 analyzed.

Subject	PVT	Subjective metric	Objective metric			# warnings emitted by our model
		KSS Level	# Lapses	RT mean (s)	RT variance (10^{-3} s)	
1	PVT1	3	0	0.25	2.10	0
	PVT3	7	1	0.29	8.34	11
6	PVT1	2	2	0.34	1.94	0
	PVT3	7	40	0.50	12.63	19

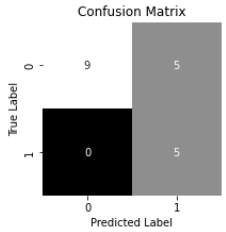
Source: Adapted from (C. B. S. Maior et al., 2020)

Quantitative metrics such as mean RT, RT variance, and the number of lapses were considered. Subject 6 exhibited a 47% increase in mean RT from PVT1 to PVT3, correlating with an increase in lapses from two to 40. The long blink detection model, (C. B. S. Maior et al., 2020), issued 19 drowsiness warnings in PVT3 for subject 6, indicating its effectiveness in capturing the declining performance during drowsiness.

In our case, we applied a similar integration approach, aligning EEG data with the duration of each RT and using the model to predict lapses and non-lapses based on EEG features at each reaction time point. Considering the 3 specific EEG features (HFD, Complexity, and Mobility), we input the number of varied points for each recorded reaction time (for example, 2000 points of EEG signals for RT 1 and 2500 for RT2) and extract the characteristics of each of them. To create the labels, we considered lapses to be those that took more than 0.5s and non-lapses to be those shorter than that, as previously mentioned, summing up 40 lapses and 52 non-lapses.

The confusion matrix indicates a 74% accuracy. In the confusion matrix depicted in Figure 21, it is evident that the detection of non-lapses achieved a precision of 100%, as also illustrated in Table 7. The detection of lapses reached a precision of 50%. This outcome may be attributed to the imbalance in the training dataset (40 lapses and 52 non-lapses), causing the model to better learn from non-lapse data. However, in this small sample, it implies that non-lapse data were erroneously considered as a lapse, indicating that what the model identifies as a lapse indeed is, even though it may not have successfully identified all instances. This aligns with part of the results presented by (C. B. S. Maior et al., 2020) model, which managed to identify half of the lapses using a blink detection model, demonstrating that other potential factors might contribute to lapses that were not detected in both the EEG-based model and the blink detection model.

Figure 21 - Confusion matrix for subject 6 in classification of lapses and non-lapses



Source: The author (2024).

Table 7 - Metrics of confusion matrix for subject 6 of lapses and non-lapses

Label	Precision	Recall	F1-score
0	1.00	0.64	0.78
1	0.50	1.00	0.67

Source: The author (2024).

5. QUANTUM EEG

This chapter delves deeper into the realm of drowsiness detection QML. Part of this chapter has been published in Process Safety and Environmental Protection journal (Lins et al., 2024).We present a hybrid classical-quantum algorithm based on a Quantum Neural Network for extracting features from EEG signals and, then, classifying individuals as drowsy or awake. To that end, parameterized quantum circuits are obtained from the extracted features, allowing the detection of drowsiness.

Although traditional machine learning methods such as MLP have been explored for EEG-based drowsiness detection, the integration of quantum mechanical concepts presents new ways to address the problem. Various quantum circuit architectures here are employed with different layer configurations. The findings highlight the suitability of QML models for analyzing sleepiness-related EEG data, indicating potential advances as quantum computing evolves.

5.1. Variational Quantum Circuits

One particular Noisy Intermediate Scale Quantum (NISQ)-friendly QML technique is that of variational/parameterized quantum circuits, leading to hybrid (classical-quantum) machine learning models (Benedetti et al., 2019). If these models are applied to classification problems, they are known as Variational Quantum Circuits (VQCs). In a classification problem, the trained model aims to assign one out of several discrete classes to a given observation according to a pattern learned from a set of labeled (previously classified) examples. In our case, VQC should categorize EEG data into either “drowsy” or “awake” classes.

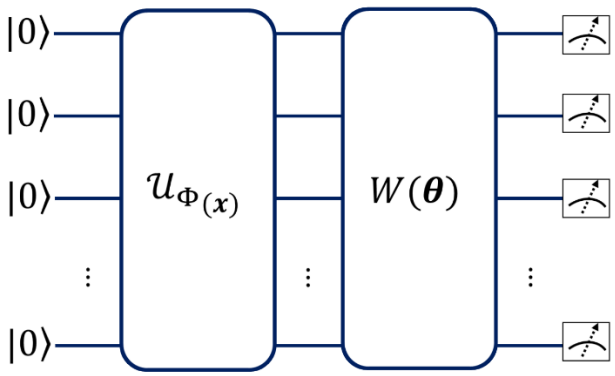
A VQC is also called a Parameterized Quantum Circuit (PQC), which refers to the way these models work: they are quantum circuits whose gates (most often, rotations) are parameterized by a free parameter vector θ. In this sense, we can think of the quantum circuit as implementing a hypothesis function whose functional form is given by the particular quantum circuit architecture and whose parameterization arises directly using parameterized quantum gates. One important point is that the loss function is classically optimized, so we often refer to VQCs as hybrid (classical-quantum) machine learning models. This point will be detailed further on.

The quantum circuits associated with VQCs may be schematically thought of as composed of three layers:

- The feature map $\mathcal{U}_{\Phi}(x)$: part of the circuit responsible for encoding classical data x into quantum states that will be processed in the quantum circuit;
- Variational layer $W(\theta)$: part of the circuit parameterized by θ. These are the parameters learned in the training process;
- Measurement: the final part of the circuit, consisting of measurements of the quantum register(s), thus producing classical information.

In Figure 22, the feature map and variational layers were schematically represented by boxes (or algebraically by $\mathcal{U}_{\Phi}(x)$, $W(\theta)$), but, naturally, they represent unitary operations. This comes with a very important caveat: the feature map and variational layers can be constructed in several ways, using different ansätze, consisting of diverse quantum gates in different compositions. The choice of particular constructions is often referred to as the architecture of the VQC.

Figure 22- Basic VQC architecture



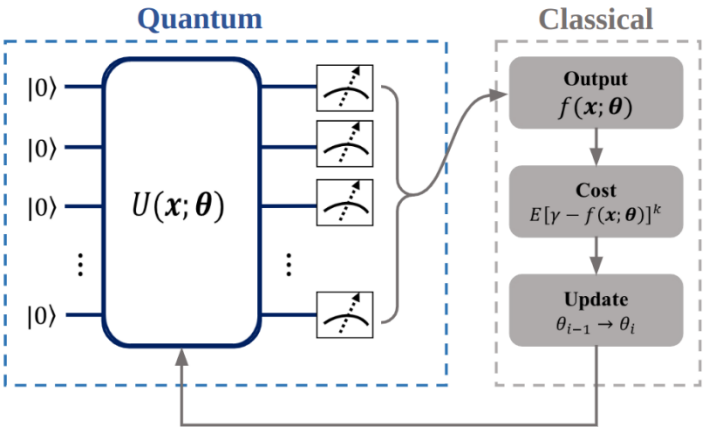
Source: The author (2024).

As aforementioned, VQC is a hybrid algorithm because it uses classical routines to optimize the parameters θ according to the minimization of a loss function calculated on the training data; this process is illustrated in Figure 23. Notice that the construction of a classical loss function is achieved directly from the measurements of the quantum registers in the last layer of the circuit, which produces classical data.

The main components of a VQC are detailed below. First, we initialize the computation with a standard initial state where all N qubits are in the computational basis $|0\rangle$ (eigenstate of the Pauli matrix σ_z with eigenvalue +1). That is, the initial N-qubit state is given by

$$|\psi_0\rangle = |0\rangle^{\otimes N} \tag{8}$$

Figure 23 - Hybrid VQC routine. Here, we are representing the composition of the variational layer and the feature map by $U(\boldsymbol{x};\boldsymbol{\theta}) = W(\boldsymbol{\theta})U_{\Phi}(x)$



Source: The author (2024).

The encoding of classical data into a quantum state vector can be done in different manners. Here, we choose an encoding such that each feature from a given data point is mapped to a rotation angle. Each qubit in the initial state evolves by a rotation R_x^i around the X axis of the Bloch sphere (for the i-th qubit), where

$$R_x(\phi) = e^{-i\frac{\phi}{2}X} = \begin{bmatrix} \cos\frac{\phi}{2} & -i\sin\frac{\phi}{2} \\ -i\sin\frac{\phi}{2} & \cos\frac{\phi}{2} \end{bmatrix}, \tag{9}$$

resulting in the state

$$|\psi_{1,k}\rangle \Rightarrow \prod_{i=0}^{N-1} R_x^i(\phi_{i,k})\,|\psi_0\rangle, \tag{10}$$

where $\phi_{\{i,k\}}$ is the angle used to encode the i-th feature ($i = 0, 1, \ldots, N-1$) of the k-th training observation. Notice that $\phi_{\{i,k\}}$ can be equal to (a normalized value of) the feature itself $x_{\{i,k\}}$, or it may be a function thereof. In our case, as seen in Eq. (13), we use $\phi_{i,k} = \frac{\pi}{2}x_{i,k}$. This is known as the angle encoding feature map.

The variational layer is the part of the quantum circuit made of entanglers and parameterized rotations, whose parameters will be learned by classically minimizing a loss function calculated on the training data. The parameters of the rotations will be collectively represented via the variational parameters vector θ, whose dimension depends on the chosen number of parameters. As with classical neural networks, the

particular topology (architecture) of the variational circuit will influence the final model, and one should iterate different architectures to tune and improve it.

As with any quantum algorithm, the qubits must be measured, and the measurement results must be post-processed. The answer to a classification problem, i.e., the predicted class, may be encoded in the measured outcome of the first qubit when measured in the σ_z basis or by a combination of the measurement of different qubits, which determines the specific function $f(x,\theta)$ indicated in Figure 23. Then, the measurement may be directly identified with the model prediction with no post-processing, or one can additionally process the circuit measurements, yielding a hybrid model in the sense that even the prediction mechanism has a quantum and a classical part.

The training of a VQC is done by choosing a loss function $L_{k(\hat{y},y)}$ to be minimized with respect to the variational parameters vector θ. Here, $\hat{y}$ is the predicted class, whilst y is the true class. For instance, one could employ the squared error $L_{k(\hat{y},y)} = \left(y_k - P(y_k)\right)^2$, where $P(y_k) \equiv P(y_k|x_k)$ is the probability of observing the specified target value of the k-th observation, y_k, given its features x_k. In Figure 23, the predicted probability is identified by $f(,\theta)$, a classical loss built from the measurements of the quantum circuit, which clearly depends on the features x and the variational parameters θ and indirectly on the circuit architecture chosen for both the feature map and the variational layer.

In the context of neural networks, this optimization is often achieved with the gradient descent algorithm, which consists of the update of the variational parameter vector according to

$$\theta \to \theta - \eta \partial_\theta \mathrm{L_k} \tag{11}$$

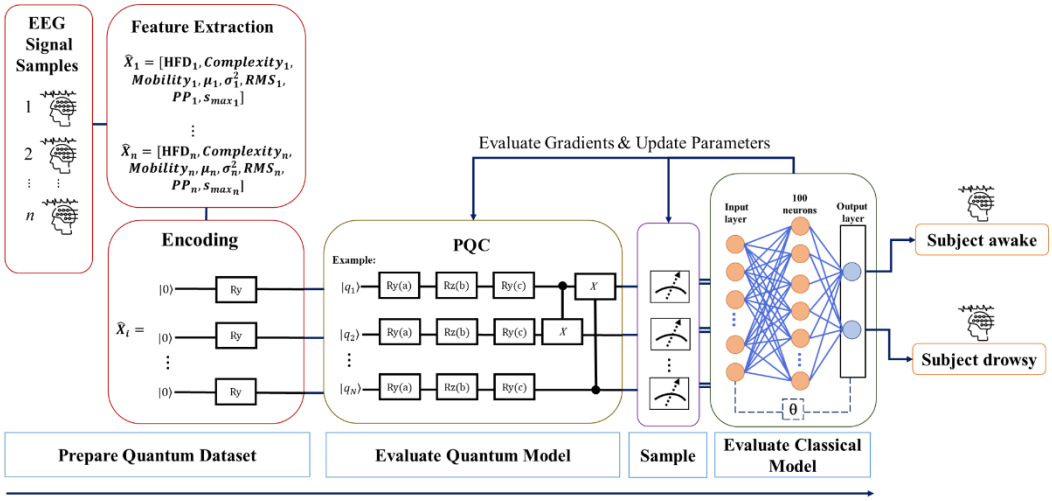
where η is the learning rate and $\partial\theta \equiv \frac{\partial}{\partial\theta}$ denotes the directional derivative with respect to the variational parameters vector θ. More specifically, in our implementation, we perform a stochastic gradient descent for the Loss optimization, which is characterized by the fact that the weight update step outlined above is performed for every observation of the training dataset. Here, we are using the TensorFlow default for the optimization algorithm.

5.2. Quantum EEG approach

The TensorFlow-Quantum® library and Cirq simulator were used to create the circuits and run the models without considering the quantum noise. In this study, we use

a QML framework to classify the subjects according to Figure 24. This approach is composed of four main steps.

Figure 24- QML framework for drowsiness classification



Source: The author (2024).

The first step consists of the quantum data set preparation. The usual procedure of normalization and feature extraction used for ML or DL as pre-processing is adopted. Here, we used eight features to extract from the EEG signal. Three of them are related specifically to EEG data: HFD, Complexity, and Mobility and the others are common statistical features: Mean, Variance, Root Mean Square (RMS), Peak-to-Peak, and Maximum Amplitude. The peak-to-peak (PP) value of a waveform is defined as the difference between its maximum positive and maximum negative amplitudes. For a time series of length L, s_i, the variance is given by $\sigma^2 = \frac{1}{L}\sum_{i=1}^{L}(s_i - \bar{s})^2$. The mean is written as $\bar{s} = \frac{1}{L}\sum_{i=1}^{L}(s_i)$. The RMS is calculated as the square root of the mean square. It is calculated as $PP = s_{max} - s_{min}$, where s_{max} represents the maximum positive amplitude and s_{min} represents the maximum negative amplitude. The maximum amplitude refers to the greatest displacement or distance a point travels on a vibrating body or wave, measured from its equilibrium position. It is denoted by the symbol s_{max}, representing the maximum displacement or amplitude.

As shown in Figure 24, the vectors $\widehat{X_1}$, $\widehat{X_2}$ and $\widehat{X_3}$ contain the number of columns relative to the number of features extracted(Sierra-Sosa et al., 2020). Eq. (12) is the mathematical formulation of the encoding process, where the tensor product of the vector space S is achieved.

$$x \to |\psi\rangle = \mathrm{S}(x_0) \otimes S(x_1) \otimes \cdots \otimes \mathrm{S}(x_{N-1}) \quad (12)$$

The operation performed in $\mathrm{S}(x_i)$ can be developed as follows in Equation (13):

$$\mathrm{S}(x_i) = \cos\left(\frac{\pi}{2}x_i\right) \cdot |0\rangle + \sin\left(\frac{\pi}{2}x_i\right) \cdot |1\rangle \quad (13)$$

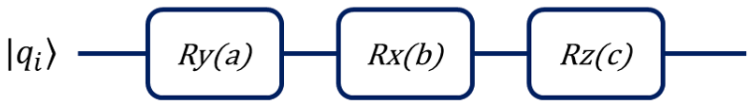
In the second step, the objective is to evaluate the quantum circuit. The PQC is created by defining the logic gates that will operate the qubits. The gates' parameters, for example, the rotation angles, will be adjusted during the QML model training according to the backpropagation effect. For performance comparison purposes, different PQCs architectures will be used in this study, based on the (Rasmussen & Zinner, 2022) model, which combines Euler rotations with entanglement. The first one consists of a sequence of three rotation gates (see Figure 25), namely R_y, R_x, and R_y, where the rotations have a matrix representation given by

$$R_y(\xi) = \begin{bmatrix} cos\frac{\xi}{2} & -sin\frac{\xi}{2} \\ sin\frac{\xi}{2} & cos\frac{\xi}{2} \end{bmatrix}, \quad (14)$$

and

$$R_z(\xi) = \begin{bmatrix} e^{-i\frac{\xi}{2}} & 0 \\ 0 & e^{i\frac{\xi}{2}} \end{bmatrix} \quad (15)$$

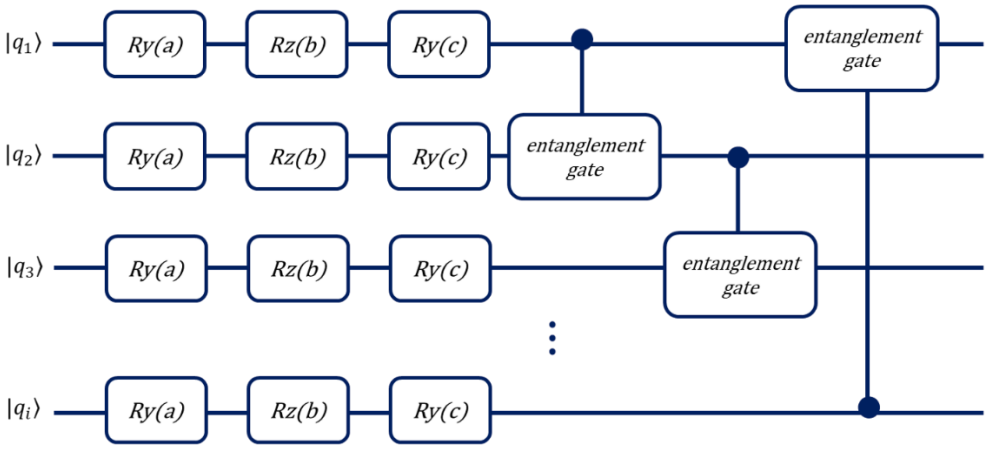
Figure 25 - PQC with rotation gates architecture



Source: The author (2024).

The second configuration is based on the Variational Quantum Eigensolver (VQE). An Euler rotation is defined, i.e., a sequence of R_y, R_x, and R_y (C. B. S. Maior et al., 2023; Rasmussen & Zinner, 2022). This study tested three two-qubit gates: $CNOT$, CZ, and $iSWAP$. Their matrices are shown in Eq. (16) to (18) respectively. A generic visualization of the PQC scheme can be observed in Figure 26.

Figure 26 - PQC with VQE and generic entanglement gates architecture



Source: The author (2024).

$$\text{CNOT} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{16}$$

$$\text{CZ} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \tag{17}$$

$$i\text{SWAP} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & i & 0 \\ 0 & i & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{18}$$

The third step performs the measurement operations corresponding to Pauli Z-gate given by

$$Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \tag{19}$$

In the fourth and last step, the traditional ML structure of the model is defined. In this work, we consider a traditional Neural Network (NN), which receives features that have been encoded and extracted from the database. Backpropagation is introduced into the PQC and flows through the NN's weights. We here employ the numeric finite difference method to compute the derivatives. That is, we perform the approximation

$$\frac{\partial L_k}{\partial_{\theta_i}} \approx \frac{L_k(\hat{y}(\theta_i + \delta), y) - L_k(\hat{y}(\theta_i), y)}{\delta} \tag{20}$$

where θ_i denotes the i-th component of θ, and δ is a small deviation. Naturally, as long as δ is finite, this difference will only approximate the actual derivative, achieved as $\delta \rightarrow 0$. If we choose a small enough δ, we, in general, obtain a good approximation, although some care is needed to avoid numerical blowups.

5.3. Results and discussions for QEEG

The results obtained from the QML models in this chapter were tested on five individuals separately (subjects 1, 5, 6, 8 and 10), as in the previous chapter, as well as on all individuals together. For drowsiness detection, an EEG time window of 100 points was used, where the sampling frequency of 512 Hz corresponds to approximately 0.2 seconds, similar to the previous chapter.

Here, the metric evaluated is the test accuracy. The 12 PQCs settings and the classic MLP were run 10 times, and the mean and standard deviation of the accuracy were extracted, as shown in Table 8. Importantly, another metric extracted was the loss of the models; for more details see

APPENDIX D - Kruskal-Wallis tests

Subject 5. Referring to Table 8, the first set of results pertains to Subject 1 (following the same identification as in the previous chapter). Among the average accuracy values obtained, the hybrid quantum configuration achieved the highest result utilizing the $iSWAP$ entanglement gate with only 1 repetition layer in the circuit (80.08%). This value is close to the second-best configuration, which is the classic MLP model (79.99%). Conversely, the lowest average accuracy was observed in the configuration utilizing the CNOT entanglement gate with only one layer (75.92%). In our study, configurations involving the CNOT entanglement gate yielded the lowest results, along with CZ with ten layers.

Table 8-Accuracy results in percentage for all trained models, including the mean value

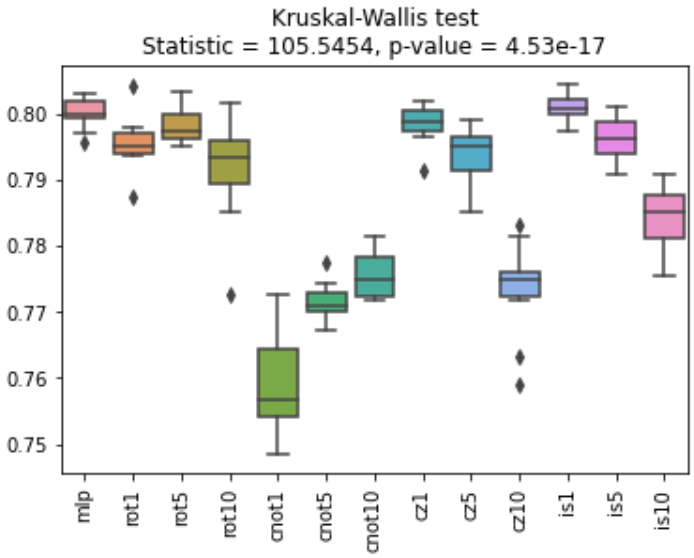
Accuracy (%) by Subject												
Model	Subject 1		Subject 5		Subject 6		Subject 8		Subject 10		All Subjects	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Classic MLP	79.99	0.22	60.50	0.80	97.47	0.42	99.29	0.29	**89.26**	0.57	**72.50**	0.74
$R_yR_zR_y$ 1 Layer	79.55	0.40	60.02	0.99	97.59	0.35	99.21	0.08	86.08	0.49	71.32	1.78
$R_yR_zR_y$ 5 Layers	79.84	0.28	60.29	0.60	97.72	0.26	99.24	0.08	86.31	0.55	71.74	1.21
$R_yR_zR_y$ 10 Layers	79.19	0.81	60.34	0.88	97.52	0.35	99.23	0.23	85.18	0.49	70.27	0.51
$R_yR_zR_y$ $CNOT$ 1 Layer	75.92	0.80	55.61	0.49	93.16	0.35	96.43	0.13	77.80	0.79	65.38	0.65
$R_yR_zR_y$ $CNOT$ 5 Layers	77.14	0.29	57.39	0.47	96.36	0.17	99.15	0.25	86.03	0.81	69.69	0.60
$R_yR_zR_y$ $CNOT$ 10 Layers	77.56	0.34	57.61	0.50	94.43	0.14	98.95	0.36	84.97	0.52	68.70	1.49
$R_yR_zR_y$ CZ 1 Layer	79.84	0.28	59.80	1.00	**97.90**	0.17	99.32	0.13	86.13	0.77	67.94	0.38
$R_yR_zR_y$ CZ 5 Layers	79.37	0.41	57.79	0.62	96.50	0.25	98.33	0.20	85.19	0.64	70.89	0.40
$R_yR_zR_y$ CZ 10 Layers	77.33	0.70	**61.13**	0.56	96.17	0.20	98.75	0.23	86.34	0.64	70.89	0.40
$R_yR_zR_y$ $iSwap$ 1 Layer	**80.08**	0.20	60.37	0.71	97.67	0.26	99.27	0.10	85.90	0.43	72.19	0.89
$R_yR_zR_y$ $iSwap$ 5 Layers	79.62	0.32	59.35	0.65	97.19	0.18	98.64	0.28	87.15	0.52	69.09	0.71
$R_yR_zR_y$ $iSwap$ 10 Layers	78.41	0.46	60.04	1.03	96.66	0.24	**99.35**	0.09	87.20	0.53	69.73	0.79

Source: The author (2024).

Here, we applied the Kruskal-Wallis test to assess if there is a significant difference among all the models employed, including the classic MLP. The Kruskal-Wallis test is a non-parametric statistical method suitable for situations when data does not follow a normal distribution or when there is heterogeneity in variance among the groups(Lybeck et al., 2007). It allows us to determine if there is statistical evidence to reject the null hypothesis, thus helping us ascertain if there are significant differences in accuracy among the different models. The null hypothesis being evaluated is whether the medians of the accuracy obtained from each model are equal.

Using a significance level of 5%, we assessed whether a difference exists between the models. As depicted in Figure 27, the p-value obtained from the test was found to be lower than the aforementioned significance level. As a result, we have sufficient statistical evidence to reject the null hypothesis. Therefore, we conclude that there is indeed a significant difference between the models. Note that in the graphs, the labels with "rot" are related to the configurations R_y, R_x, R_y, and the numbers following them are associated with the amount of layers 1, 5, or 10. On the other hand, the labels "cnot," "cz," and "is" correspond, respectively, to the configurations R_y, R_x, $R_y + CNOT$, R_y, R_x, $R_y + CZ$, and R_y, R_x, $R_y + iSWAP$. The box plot presented in Figure 27 further illustrates this observation, with the boxes representing the models with $iSWAP$ and one circuit layer (is1), CZ, and one circuit layer (cz1), and the classic MLP (mlp) being the three highest. Here, the CNOT-based models exhibit considerably lower values than the others.

Figure 27 - Kruskal-Wallis test to assess the factor architecture models of Subject 1



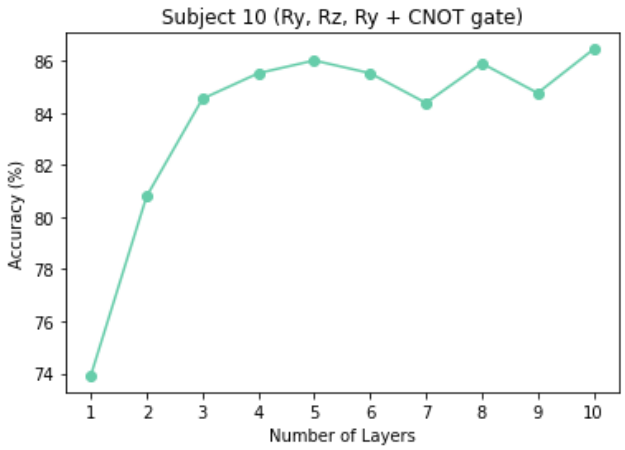
Source: The author (2024).

Finally, we also conducted the Kruskal-Wallis statistical test to assess the effect of the number of layers inserted into the circuit on the results to observe if increasing the number of layers leads to improvement. Our null hypothesis is as follows: The medians of the accuracies, considering the factor of circuit layers, are equal. We performed the test for each type of circuit structure. There was not enough statistical evidence to conclude that there is a difference between the number of layers in the configuration with only rotation gates (R_y, R_x, R_y). However, the null hypothesis was rejected in the

configurations $R_y, R_x, R_y + CNOT$, Ry $R_y, R_x, R_y + CZ$, and $R_y, R_x, R_y + iSWAP$ (p-values, respectively, 0.0003, 0.00001, and 0.00001). We observe that for CNOT, the higher the number of layers, the higher the result of the median of the accuracy. On the other hand, for CZ and iSWAP, the median is the highest in the configuration with 1 layer, followed by 5 layers, and then 10 layers.

Aiming to observe if the increase of circuit layers in the QML models causes a better accuracy performance, a graph was plotted for Subject 10, as shown in Figure 28. In this illustration, the configuration used was with the two-qubit CNOT gate since the accuracy behavior was increased when running the model with 1, 5, and 10 layers. Thus, we ran the model varying the layers by units from 1 to 10. It can be seen that, from layers 1 to 2, the increase is approximately 7 percentage points. From that to layer 5 the accuracy continues to increase but in smaller intervals. Between layers 5 and 10 the behavior is unstable; that is, there was growth, as expected. This is because the accuracy decreases from layer 5 to layer 7, increases in the eighth layer, decreases in the ninth and increases again in the tenth. Therefore, it is not possible to assume that by increasing the number of layers in the circuit, the model improves.

Figure 28 - Subject 10 mean accuracy with the Ry, Rz, Ry + CNOT



Source: The author (2024).

One crucial insight gleaned from this chapter is the pivotal role of heuristics in determining optimal parameterized circuit architectures for VQCs. In fact, as it was shown, more entanglement does not invariably translate to superior predictive performance. Moreover, while this study focused on noiseless simulations, computations performed on actual quantum hardware are still deeply affected by noise, and, in fact, noise also has an effect in the trainability of VQCs.

Our research, which focus on application-oriented classical noiseless simulations, are important to navigate through the enormous variety of VQCs architectures, as the pursuit of true quantum advantage demands the identification of VQC architectures capable of not only outperforming classical counterparts but doing so while eluding efficient classical simulation. Moving forward, further research should focus on refining parameterized circuit designs and mitigating quantum noise, which should be explored via deployments on actual quantum hardware.

The results obtained by applying quantum methods, although still limited by the current stage of development of quantum technologies, indicate a promising path for drowsiness detection in critical industrial systems. The comparison between traditional and quantum algorithms reveals that, even in NISQ (Noisy Intermediate-Scale Quantum) devices, quantum methods can offer advantages in terms of classification accuracy when analyzing large volumes of data. These results reinforce the potential of quantum algorithms as a complementary or even superior tool for complex machine learning tasks, especially in contexts where accuracy is critical.

6. *DATA FUSION METHODOLOGY TO DETECT DROWSINESS IN INDUSTRIAL ENVIRONMENTS WITH CRITICAL SAFETY SYSTEMS*

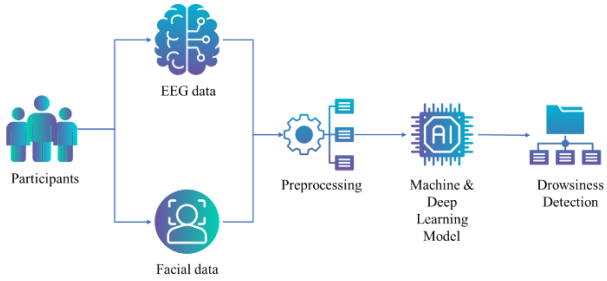
This chapter focuses on enhancing the identification and prediction of drowsiness for operators of critical safety systems, as an operation of O&G industries. The goal is to aid the detection and prevention of accidents resulting from human fatigue. To that end, our objective is to develop a combined and multi-source model that can effectively learn and recognize various features associated with drowsiness. By integrating multiple data sources, we aim to extract more accurate information, which can be utilized to improve accident prevention measures.

6.1. Proposed Methodology

We propose two models for drowsiness detection based on data fusion: one operating at the decision level and another at the feature level. Our approach leverages the strengths of both facial and EEG data to enhance the accuracy and reliability of drowsiness detection in industrial environments.

We utilized CV techniques to extract visual data of participants from a publicly available drowsiness database. Simultaneously, feature extraction was applied to the EEG raw data to facilitate identifying brain patterns related to two states of alertness during an attention task. Our approach involved the implementation of three distinct models: one for image data analysis, another for EEG data analysis and, finally, a model for data fusion. An overview of the proposed methodology is presented in Figure 29, highlighting the sequential flow of our approach. To ensure efficient processing, the collected data was initially separated, pre-processed, and transformed into a manageable format suitable for input into the learning algorithms.

Figure 29 – General overview of the proposed methodology.



Source: The author (2024).

6.2. Decision level model

At the decision level, classifiers were developed using the MLP and the pre-trained model ResNet50, which was selected after some model tests, presenting a deep ResNet architecture with 50 layers and which has obtained good results when facing performance degradation problems. These classifiers were trained on the extracted data to perform two main classification tasks: i) drowsiness detection from images, and ii) drowsiness detection from EEG data. Furthermore, a decision rule was implemented to combine the outputs of the individual classifiers and enhance the overall prediction performance. It is worth noting that all models were trained using labeled data generated from the database itself, that is, the state of alertness and drowsiness given by PVT1 and PVT3. In the following sections, we provide a more detailed description of each of these steps.

6.2.1. Preprocessing EEG data

In the analysis of EEG data for drowsiness detection, preprocessing plays a crucial role in enhancing the quality of the data and extracting relevant features. Another commonly used technique (different from those covered in chapter 4) is the DWT, which allows for the decomposition of the EEG signal into different frequency sub bands. This section provides an overview of the preprocessing steps involving the application of DWT to EEG data.

6.2.1.1. EEG preprocessing steps

The steps are described below:

- Data Acquisition: Initially, our model considers five EEG channels (Fz, Cz, C3, C4, and Pz) available in the database.
- Segmentation: Before applying DWT, the EEG signal is divided into smaller segments, typically of fixed length, to facilitate analysis and feature extraction. Each segment represents a specific time window of the EEG recording. In our case, the DROZY database comprises 18 videos of 10 minute-length, each has 307,200 data points, and overall, there are more than 5.5 million data points available. Thus, a 512-point EEG time window was used for real-time drowsiness detection, for which 512 Hz sampling frequency corresponds to ~ 1 second.
- DWT Decomposition: The segmented EEG signal is subjected to the DWT, which decomposes the signal into approximation coefficients (low-

frequency components) and detail coefficients (high-frequency components) across different scales or levels.

- Feature Extraction: DWT provides valuable information about the distribution of different frequency components in the EEG signal. After conducting numerous tests, we have found that considering only the approximation coefficients is sufficient for our application due to better performance.

6.2.2. Preprocessing Facial data

Preprocessing plays a crucial role in analyzing video data for drowsiness detection. In this section, we discuss pre-processing steps performed using CV techniques to extract relevant information for further analysis.

6.2.2.1. Computer Vision Preprocessing Steps

The following steps are performed using the cv2 library (Bradski, 2000) in Python:

- Video Processing: In the video processing step, the code loads the video using cv2.VideoCapture. It then reads frames from the video using cap.read(), allowing access to each individual frame. To enhance the visibility or brightness of the frames, the image is normalized using cv2.normalize.
- Face Detection: To detect faces in the frames, the image is converted to grayscale using cv2.cvtColor. The code utilizes the dlib library, specifically the pre-trained face detector dlib.get_frontal_face_detector(), which employs advanced algorithms to identify and locate faces in the image.
- Facial Landmark Detection: Once the faces are detected, the shape predictor model (dlib.shape_predictor) comes into play. This model analyzes the detected faces and identifies specific facial landmarks, such as the corners of the eyes or the nose. By extracting the coordinates (x, y) of these facial landmarks, the code obtains valuable information for subsequent image cropping.
- Image Cropping: Using the extracted facial landmark coordinates, the code defines a region of interest (ROI) that focuses on the relevant facial area to be cropped. The ROI is adjusted to include a specific area around

the detected region of interest, ensuring that the important facial features are captured. Finally, the cropping operation is performed using the defined ROI, extracting the desired portion of the image for further analysis.

These preprocessing steps using CV techniques are crucial for isolating and extracting the necessary visual information from the video frames. Through the combination of face detection, facial landmark detection, and image cropping, the preprocessing step prepares the video frames for subsequent analysis, such as classification, improving the overall accuracy and performance of the drowsiness detection system.

6.2.3. Data-driven models used

After conducting numerous tests with various machine learning classifiers, it has been determined that utilizing the MLP Classifier for the classification of EEG data is an effective approach (Ramos et al., 2022). This classifier demonstrates the ability to extract pertinent information and identify meaningful patterns within brain signals. Thus, we use MLP Classifier of scikit-learn (Pedregosa et al., 2011) in conjunction with features extracted from DWT to perform EEG data classification. For the facial data, we explore the utilization of the ResNet50 architecture in CV tasks, followed by the implementation of an LSTM model for sequential data analysis. The ResNet models are a powerful DL architecture for image recognition tasks, while LSTM is commonly used for sequential data analysis, such as frames of video.

6.2.3.1. MLP classifier

To use the MLP Classifier with EEG data, the first step is to extract relevant features from the DWT of the EEG signals. The extracted features include spectral characteristics such as power and peak frequency in each subband, as cited previously. Once the features are extracted from the DWT of the EEG data, they are used as input to train and evaluate the MLP classifier. The scikit-learn MLP classifier has various parameters that can be adjusted to customize the model's behavior, such as the number of hidden layers, the number of neurons in each layer, and the activation function used in each neuron.

The MLP classifier is created with a single hidden layer containing 100 neurons and the 'relu' activation function. The model is, then, trained with the training data, and

predicted labels are obtained for the test data. Finally, the performance of the model is evaluated.

6.2.3.2. ResNet50

The ResNet model was implemented using the ResNet50 architecture from the TensorFlow Keras library (Abadi et al., 2015). Using the model, we created an input tensor of the desired shape and, then, instantiated the ResNet50 class as the input. In our case, the model expects input images of size 125x125 pixels with three color channels.

After processing the images using the ResNet50 model, we proceed with the utilization of the LSTM model for sequential data analysis. The LSTM model was implemented using the Sequential class from the TensorFlow Keras library (Abadi et al., 2015). We defined the model architecture by adding LSTM layers, followed by dense layers with dropout regularization. We defined an LSTM layer with 1000 neurons to pass the output sequences to the subsequent layers. The Flatten layer was, then, added to convert the output into a one-dimensional vector. The model continued with dense layers. The final dense layer used the activation function to obtain the classification probabilities.

In our example, the LSTM model was defined with an input shape of (75, 1000), indicating that it expected input sequences of length 75 with 1000 features. The training was implemented using the Adam optimizer (Kingma & Ba, 2014), which is a stochastic gradient descent method based on adaptive estimation of first-order and second-order moments. It is computationally efficient and suitable for complicated models with many trainable parameters. The learning rate was set at 0.00005. The loss function was binary cross-entropy, which is the most widely used in classification purposes. It measures the performance of a classification model whose output is a probability value between 0 and 1. Cross-entropy loss increases as the predicted probability diverges from the actual label. Parameters were initialized using ImageNet (Deng et al., 2009).

6.2.4. Data fusion

In order to achieve a more robust and reliable drowsiness classification, a decision rule is employed to combine the results obtained from both the MLP and the ResNet50 models. The goal of this strategy is to leverage the strengths of each model and make a collective decision based on their combined predictions (Ramos et al., 2022).

During the classification process, each model independently assigns a class label to the input data, indicating whether the subject is alert or drowsy. These individual

predictions are then combined through a decision mechanism to determine the final classification. In this study, the decision-level fusion was used, as discussed in Cai et al. (2020). The decision rule operates as follows:

- For each instance, the predictions from both models are considered. If there is a majority where one class receives more votes than the other, that class is chosen as the final classification.
- However, in the event of a tie where an equal number of votes is obtained for both classes, a predefined rule is applied. In this case, the decision is made to classify the subject as drowsy.

By employing this decision rule, our aim is to improve the metrics and overall performance of the drowsiness classification system. The combination of the MLP model's analysis of EEG features and the CV model's evaluation of visual signals provides complementary information, leading to a more comprehensive evaluation of drowsiness levels.

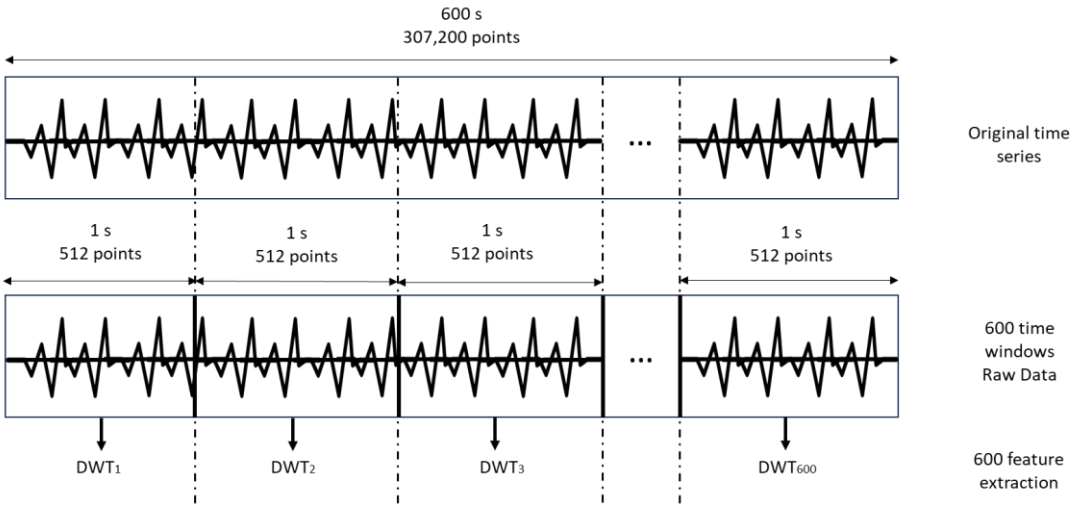
6.3. Results and discussion for decision level model

This section presents the results and discusses our proposed drowsiness detection system. The section is divided into three parts, and each one focuses on a specific model as well as its contributions to system performance: (1) drowsiness detection based on EEG data, (2) drowsiness detection based on facial image data, and (3) the fusion model for combined data detection. All the experiments were run on a GPU running Python Version 3.9 with a 3.6 GHz Intel CORE i9-9900K processor, 32 GB of RAM.

6.3.1. Drowsiness detection model from EEG Data

Numerous tests were performed to identify the most suitable ML as well as the most suitable channel for representation. The results presented are for the best-performed one, achieved for the MLP classifier and for the C4 channel. To consider real-time drowsiness detection, an EEG time-window containing 512 points was selected, which, at a sampling frequency of 512 Hz, represents <1 s, as can be seen in Figure 30.

Figure 30 - Representation of EEG data to feed the model.



Source: The Author (2024).

To assess the learning and generalization ability of the drowsiness detection model from EEG Data, the test data, representing 1s, was aligned with the seconds of test data from the video image window. In this way, each subject presented a different amount of test data for different alert states, as shown in Table 9.

This same amount of alertness and drowsiness test data was also considered for the CV model. Thus, the accuracy of the EEG data, which passed through the DWT feature extraction, is presented in Table 10.

Table 9 - Amount of test data for the five subjects.

Subject	Alertness test data	Drowsiness test data
1	84	87
5	84	81
6	81	87
8	81	87
10	87	90

Source: The author (2024).

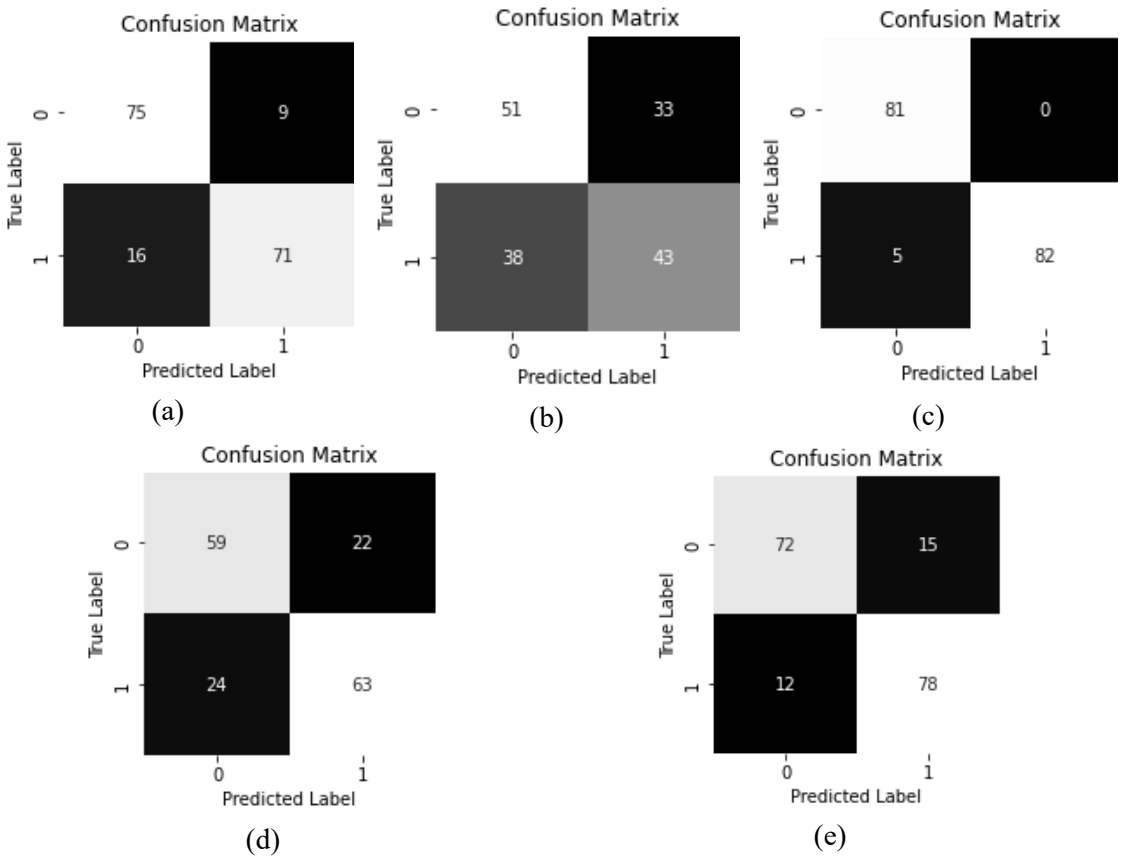
Table 10 - Results of the accuracy of the EEG data

Subjects	1	5	6	8	10
Accuracy	85.38%	57.10%	97.12%	72.62%	84.71%

Source: The author (2024).

Three out of 5 subjects achieved a more significant performance (i.e., subjects 1, 6 and 10), with an accuracy above 80% of detection. Still in Table 10, subject 5 had the worst result, when exclusively considering the DWT feature extraction in the MLP model. In fact, this finding agrees with the results of (Ramos et al., 2022), where it was observed that subject 5, despite the use of other extraction features, presented the poorest results compared to other subjects. Subject 8, in turn, presented an accuracy of around 70%. This variation in performance among individuals demonstrates the presence of interpersonal variability, highlighting the challenge of accurately distinguishing between awake and drowsy states based solely on EEG features. The confusion matrices and metrics for each subject in this case are shown in the Figure 31 and Table 11, respectively.

Figure 31- Confusion matrix from EEG model per subject (1, 5, 6, 8, 10, for (a), (b), (c), (d), and (e), respectively)



Source: The Author (2024).

Table 11 - Per-subject metrics for the EEG model

Subject	Label	Precision	Recall	F1-score
1	0	0.82	0.89	0.86

	1	0.89	0.82	0.85
5	0	0.57	0.61	0.59
	1	0.57	0.53	0.55
6	0	0.94	1.00	0.97
	1	1.00	0.94	0.97
8	0	0.71	0.73	0.72
	1	0.74	0.72	0.73
10	0	0.86	0.83	0.84
	1	0.84	0.87	0.85

Source: The Author (2024).

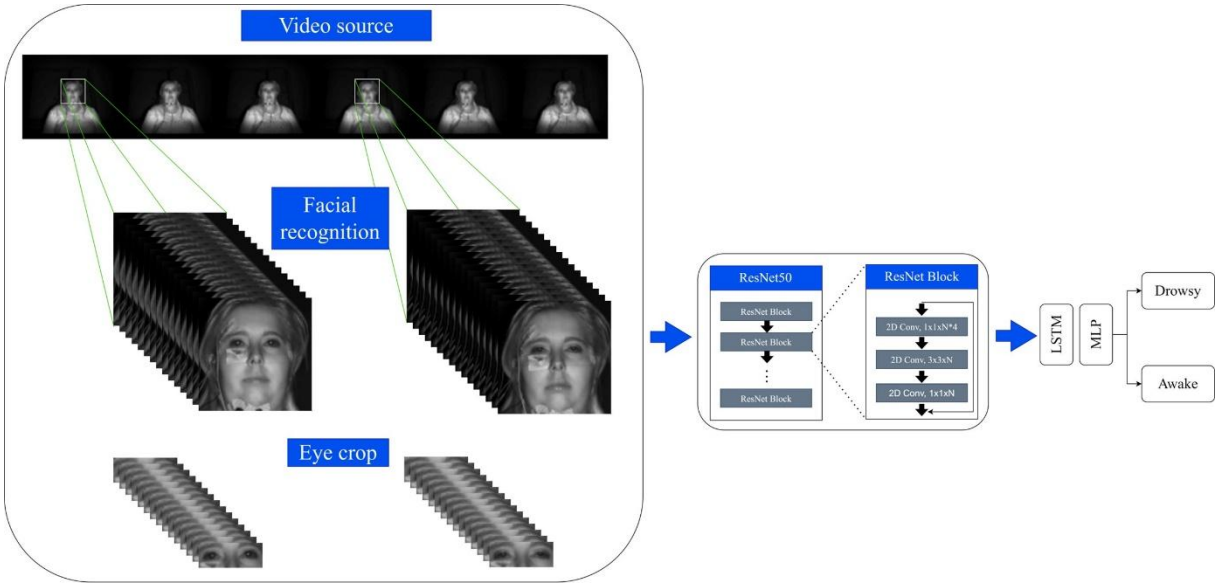
For subjects 1, 6, and 10, we observed consistently high results across all metrics, indicating robust performance of the model in detecting both states. These subjects achieved accuracies, recalls, and F1-scores above 0.80 for both states, demonstrating a reliable ability to distinguish between alertness and drowsiness. However, for subject 5, the results were significantly lower, with precisions, recalls and F1-scores around 0.57 for both states. This suggests an additional difficulty in detecting drowsiness for this specific subject, which may be related to unique individual characteristics in their EEG patterns. Subject 8 presented intermediate results, with metrics around 0.70 for both states. This indicates reasonable model performance, but with room for improvement in terms of precision and recall. Highlighting the importance of personalization and adaptation of the model to deal with interpersonal variability and achieve accurate and reliable detection of drowsiness in diverse individuals.

6.3.2. Drowsiness Detection Model from Facial Data

In this part, we evaluate the performance of the ResNet50 model using the images extracted from CV-based techniques. In this analysis, a 5-second time window was employed for each test data to enhance the detection of facial information comprehensively. To prevent any time overlap between the test and training data, the seconds within each test data time window were excluded from the training data. This approach ensures that the model is evaluated on unseen facial (more specifically, eye crop) images and provides a more accurate assessment of its performance in detecting drowsiness, as can be seen in Figure 32. Table 12 shows the computational times, measured in seconds, necessary to process the data for training and testing the models.

We highlight that the processing time for subject 10 is almost double the average for the other subjects because the camera recorded in the experiment was different, having twice as many frames per second.

Figure 32 - Representation of image data to feed the model.



Source: The Author (2024).

Table 12- Run times in seconds by model for the five subjects.

Subject	Train	Test
1	2510.845	850.693
5	2267.739	769.462
6	2269.260	752.608
8	2270.812	744.312
10	4925.148	1593.598

Source: The Author (2024).

We measure the performance of the ResNet50 model using various evaluation metrics, including accuracy, precision, recall and F1 score. These metrics provide information about the model's ability to correctly classify drowsiness and alertness states based on facial features. In addition, we analyzed the model's confusion matrix to understand any misclassifications and identify potential areas for improvement. Thus, accuracy of the facial data is presented in Table 13.

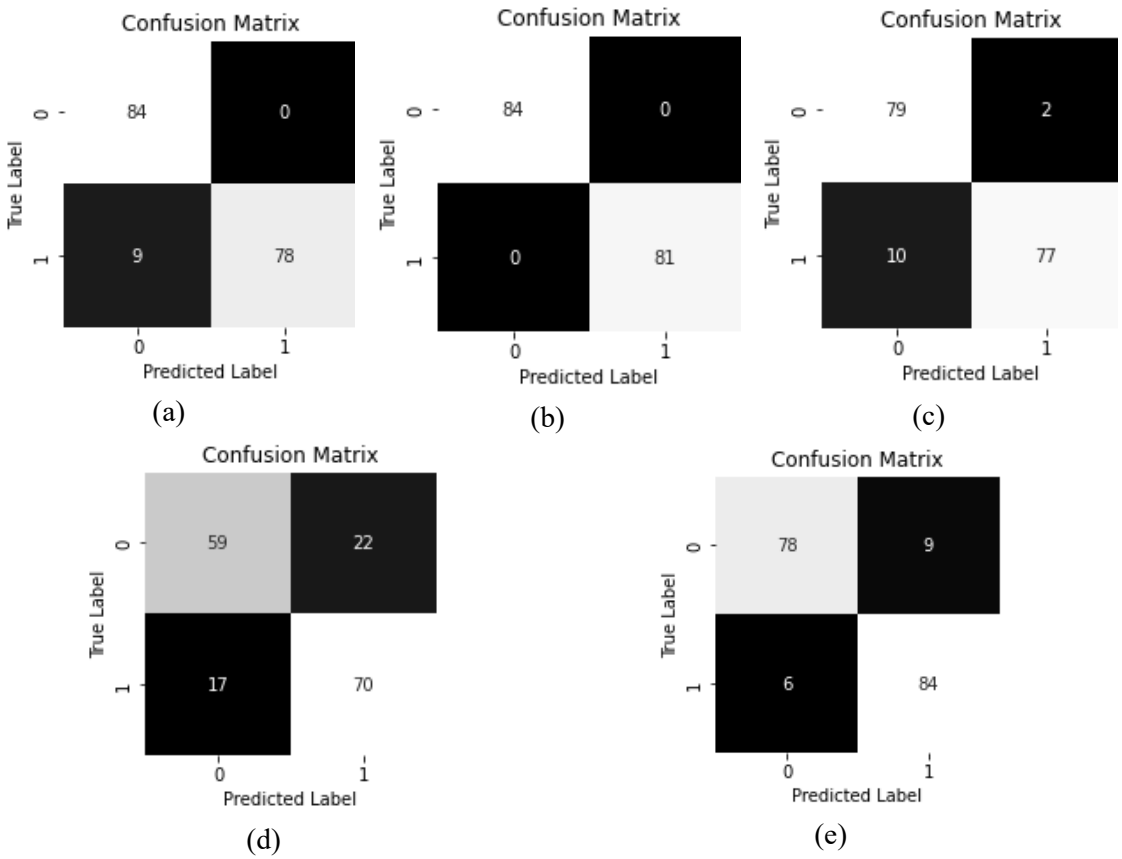
Table 13 - Results of the accuracy of the facial data

Subjects	1	5	6	8	10
Accuracy	94.73%	100%	92.86%	76.78%	91.80%

Source: The author (2024).

CV-based drowsiness detection results show promising accuracy rates for subject tests. Four out of the five evaluated subjects achieved accuracy levels above 90%. Subject 5 demonstrated a perfect accuracy rate of 100%, indicating that the model correctly classified all cases of drowsiness and alertness. Subjects 1, 6 and 10 also achieved high accuracy rates of 94.73%, 92.86% and 91.80%, respectively, indicating a reliable performance and validating the effectiveness of the CV approach from the model ResNet50. However, Subject 8 continued to show a low accuracy rate (i.e., 72.61%), suggesting some challenges in accurately classifying drowsiness in this specific case. The confusion matrices and metrics for each subject in this case are shown in the Figure 33 and Table 14, respectively.

Figure 33 - Confusion matrix from CV model per subject (1, 5, 6, 8, 10, for (a), (b), (c), (d), and (e), respectively)



Source: The author (2024).

It is worth noting that the test dataset used to evaluate the CV-based drowsiness detection model is relatively small, with a limited number of instances for each subject (see Table 9). The number of samples of alertness and drowsiness test data varies from 81 to 90 depending on the subject, with the smallest being for subject 5. Although these

sample sizes may be considered small, it is important to recognize that the model still achieved high accuracy rates despite the limited amount of data. However, it is essential to further validate the performance of the model in different formats from the datasets to ensure its robustness and generalization ability across different individuals and scenarios.

Table 14 - Per-subject metrics for the CV model

Subject	Label	Precision	Recall	F1-score
1	0	0.90	1.00	0.95
	1	1.00	0.90	0.95
5	0	1.00	1.00	1.00
	1	1.00	1.00	1.00
6	0	0.89	0.98	0.93
	1	0.97	0.89	0.93
8	0	0.70	0.75	0.73
	1	0.75	0.70	0.73
10	0	0.93	0.90	0.91
	1	0.90	0.93	0.92

Source: The author (2024).

For subjects evaluated in the CV model, we consistently observed high results across all metrics, indicating robust performance of the model in detecting both states. Specifically, four of the five subjects achieved precision, recall, and F1-score rates above 90%.

However, as with the EEG-based model, subject 8 continued to show a relatively low accuracy rate (i.e., 72.61%), indicating some difficulties in accurately classifying drowsiness in this specific case. This variation in performance across subjects once again highlights the importance of considering interpersonal variability when evaluating the effectiveness of drowsiness detection models. It is also worth mentioning that these models were trained for each individual subject, with no generalization of facial information.

6.3.3. Drowsiness Detection Fusion Data Model

In the data fusion approach, we combined the predictions from both the MLP model using features extracted from DWT of EEG data and the CV-based model. By assessing the recall of class 1 (Table 15), that is, the ratio between the number of true positives for drowsiness and the total number of positive cases for drowsiness, the data

fusion model demonstrated superior performance to both individual models (EEG and CV). Despite a slightly lower accuracy compared to the CV model in most cases (Table 17), the drowsiness detection reached values close to or equal to 100% in 4 out of 5 subjects and surpassed the CV result for Subject 8. For instance, for Subject 6, the CV model made incorrect predictions for 10 test data of the drowsiness class, whereas the data fusion model got all of them right, showcasing the significant benefit of incorporating EEG data.

However, it's crucial to mention that, considering the recall for the alert state, we observed a decrease in almost all subjects in the data fusion model compared to the separate CV model. Subject 5, in particular, showed a significant drop, going from a recall of 100 in the CV model to 61 in the data fusion model. This decline can be attributed to the less robust performance of the EEG model and the conservative nature of the decision rule in the data fusion model. However, the recall per subject for drowsiness had a significant improvement in the data fusion model, compared to previous models, as can be seen in Table 15.

Table 15 - Per-subject recall for drowsiness state for the three models

Subject	**Recall**		
	EEG	**CV**	**Data fusion**
1	0.82	0.90	**0.98**
5	0.53	**1.00**	**1.00**
6	0.94	0.89	**1.00**
8	**0.72**	0.70	**0.72**
10	0.87	0.93	**1.00**

Source: The author (2024).

In Table 16, we observed an accuracy performance of the data fusion model compared to individual EEG and CV models. In terms of accuracy, the data fusion model outperformed EEG results in 4 out of 5 subjects, matching only Subject 8. Regarding CV results, the data fusion model surpassed only Subject 6, and was comparable to Subject 1.

Table 16 – Accuracy performance of the three models

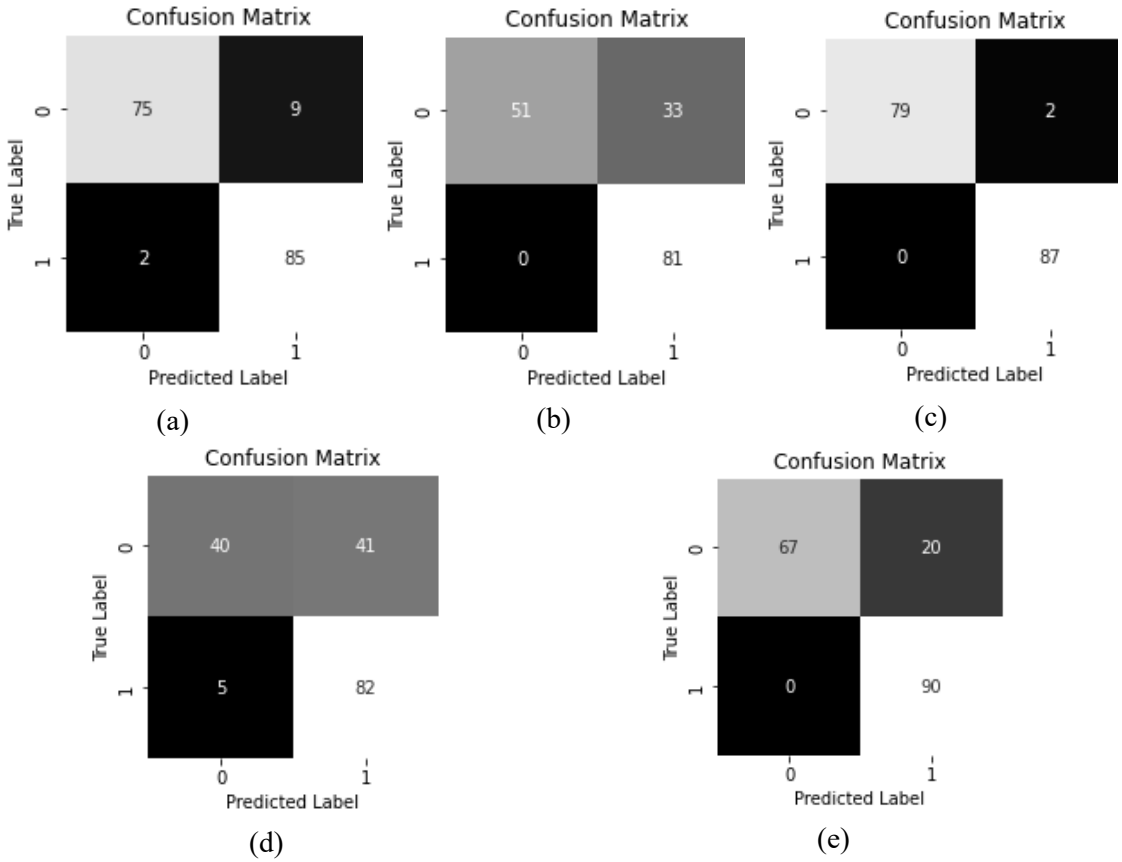
Data source	Subjects				
	1	5	6	8	10
EEG	85.38%	57.10%	97.12%	72.62%	84.71%

CV	**94.73%**	**100%**	92.86%	**76.78%**	**91.80%**
Data fusion	93.57%	80%	**98.81%**	72.62%	88.70%

Source: The author (2024).

These results highlight the effectiveness of combining both EEG-based and facial-based features for detecting drowsiness. The data fusion approach yielded improved accuracy in most subjects compared to individual models of the EEG data. This suggests that leveraging the complementary information extracted from both modalities can enhance the overall performance of drowsiness detection systems. It is important to note that the data fusion approach achieved the highest accuracy in Subject 6, indicating the significance of integrating multiple sources of information for accurate drowsiness detection in different individuals. The confusion matrices and metrics for each subject in this case are shown in Figure 34 and Table 17, respectively.

Figure 34 - Confusion matrix from data fusion model per subject (1, 5, 6, 8, 10, for (a), (b), (c), (d), and (e), respectively)



Source: The author (2024).

Table 17 - Per-subject metrics for the data fusion model

Subject	Label	Precision	Recall	F1-score
1	0	0.97	0.89	0.93
	1	0.90	0.98	0.94
5	0	1.00	0.61	0.59
	1	0.71	1.00	0.76
6	0	1.00	0.98	0.83
	1	0.98	1.00	0.99
8	0	0.71	0.73	0.72
	1	0.74	0.72	0.73
10	0	1.00	0.77	0.87
	1	0.82	1.00	0.90

The results from Table 17 suggest that the fusion model generally outperformed the individual EEG and CV models in terms of accuracy, precision, recall, and F1-score. This improvement in performance underscores the potential benefits of combining multiple modalities for drowsiness detection. Additionally, the fusion model may be less prone to biases inherent to individual methods, as it integrates diverse sources of information, thereby providing a more holistic and neutral assessment of drowsiness.

However, it is important to note that, similar to the individual models, the fusion model still exhibited some variability in performance across different subjects. Subject 8 continued to show relatively lower accuracy rates compared to other subjects, indicating ongoing challenges in accurately detecting drowsiness in this specific case. This highlights the continued importance of considering interpersonal variability and fine-tuning model parameters to optimize performance for different individuals.

Discussion of the results focused on the overall performance of the fusion model compared to individual models, highlighting the improvements achieved through the integration of EEG and facial image data. We also address the implications of decision rule, its impact on decision-making, and the potential benefits of using multiple modalities in enhancing drowsiness detection accuracy.

By analyzing the results, we gained valuable insight into the effectiveness of the ResNet50 model and the benefits of integrating EEG and CV techniques for drowsiness detection. The findings provide a solid foundation for further advances in the

development of robust and reliable systems to prevent accidents caused by drowsy operators in critical safety systems.

6.4. Feature level model

At the feature level, the approach involves transforming EEG signals into spectrograms, preprocessing both EEG and facial data, and aggregating the features before feeding them into a deep learning model for training and evaluation.

6.4.1. Processing EEG Data

The first major step in our methodology is the transformation of raw EEG signals into spectrograms. This involves the following steps:

- Data Acquisition: EEG data is loaded from EDF files using the MNE library. For this study, we focus on the C4 channel as it has shown significant relevance in previous tests.
- Segmentation: The continuous EEG data is segmented into 1-second or 5-second windows. This is achieved by dividing the data into segments of a fixed length corresponding to the sampling frequency.
- Spectrogram Generation: Each EEG segment is transformed into a spectrogram using the SciPy library's spectrogram function. This transformation involves applying a Hamming window and converting the power spectral density to a decibel scale.
- Saving Spectrograms: The spectrograms are saved as images in specific folders, each representing a 1-second or 5-second window of consecutive EEG segments.

6.4.2. Aggregating Data and Model Training

After preprocessing the EEG data into spectrograms and organizing the facial frames, the next step involves aggregating these features and passing them to a deep learning model for training. The steps include:

- Loading and Preprocessing Images: Both the frames from video data and the spectrograms are loaded and preprocessed using the OpenCV library. Each image is resized to a target size suitable for the ResNet50 model and normalized using the preprocess_input function from TensorFlow.

- Feature Extraction: Features are extracted from both the frames and spectrograms using the ResNet50 model pre-trained on ImageNet. The model is used without its top layer, and the output features are pooled using global average pooling.
- Data Aggregation: The extracted features from the frames and spectrograms are concatenated to form a single feature vector for each instance. This aggregated feature set provides a comprehensive representation of both visual and EEG data.
- Model Training: The aggregated feature vectors are used to train a dense neural network. The network architecture includes several dense layers with ReLU activation, followed by a sigmoid activation function in the output layer for binary classification (alert vs. drowsy).
- Evaluation: The model's performance is evaluated using standard metrics such as accuracy, precision, recall, and the confusion matrix. The evaluation helps assess the model's effectiveness in detecting drowsiness based on the data fusion.

By following this methodology, we aim to leverage the complementary information from EEG spectrograms and visual frames to improve the robustness and accuracy of the drowsiness detection system. The combination of features from multiple data sources is expected to provide a more holistic assessment of the operator's cognitive state, leading to more reliable drowsiness detection in industrial environments.

6.5. Results and discussion for feature level model

This section presents the results and discusses our proposed feature-level drowsiness detection system. The section is divided into three parts: (1) presentation of the results using confusion matrices and classification metrics for 1s data, (2) presentation of the results using confusion matrices and classification metrics for 5s data, and (3) discussion of the findings and comparison with the results of the decision-level model.

6.5.1. Presentation of results (1-second window)

The results for the feature level model with a 1-second window, focusing on the C4 channel, are summarized in the following confusion matrices (Figure 35) and classification reports (Table 18) for each subject. Confusion matrices for additional

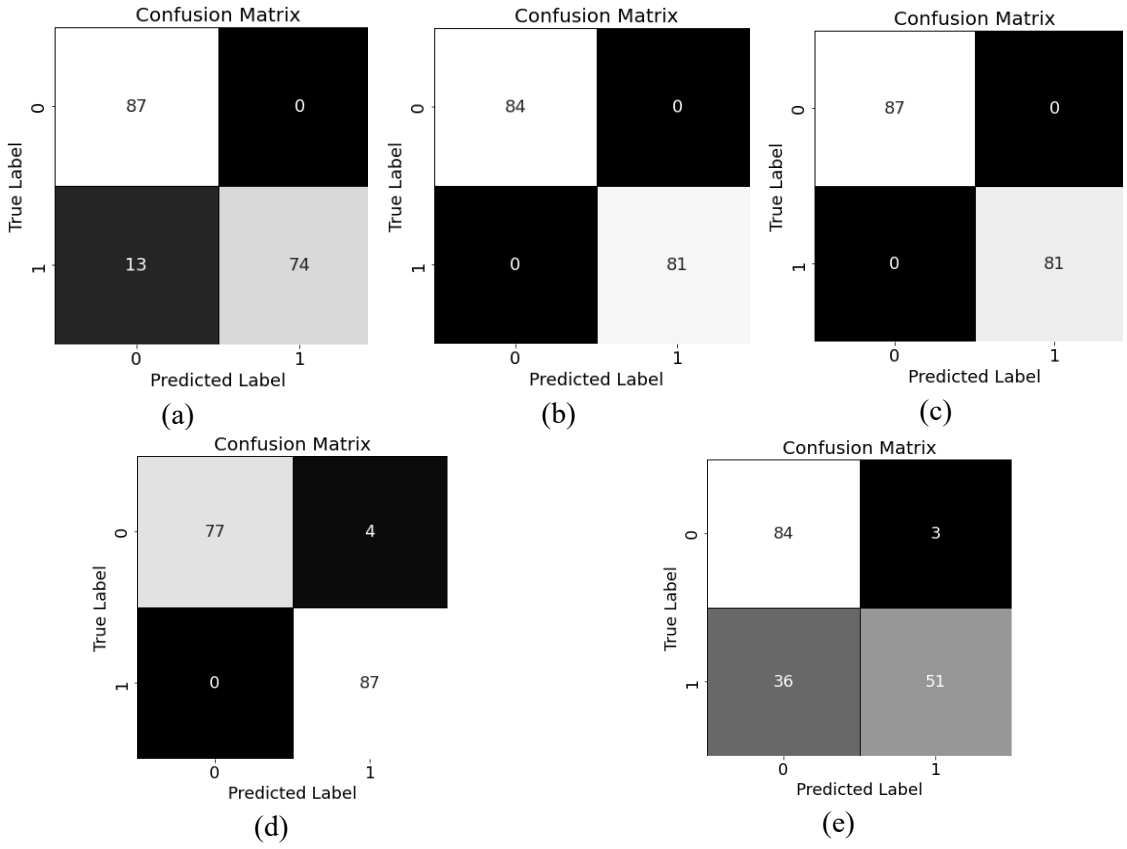
channels (FZ, C3, CZ, and PZ) for the same subjects, are provided in APPENDIX E - Extended performance of feature-level models.

The results for the feature-level model demonstrate strong performance in classifying drowsiness across multiple subjects. The model consistently achieved high precision, recall, and F1 scores, particularly for Subjects 5 and 6, who achieved perfect scores across all metrics.

For Subject 1, the model achieved an overall accuracy of 93%, with a precision of 1.00 for awake states and a recall of 0.85. The slightly lower recall for the drowsy state indicates that some instances of drowsiness were missed, which is a critical consideration for further improving the model.

Subjects 5 and 6 exhibited exceptional performance, with 100% precision, accuracy, recall, and F1 scores. This indicates that the model was highly effective in distinguishing between drowsy and awake states for these subjects, demonstrating the potential of the feature-level model in real-world applications.

Figure 35 - Confusion matrix from data fusion feature level model (1s-w) per subject (1, 5, 6, 8, 10, for (a), (b), (c), (d), and (e), respectively)



Source: The author (2024).

Table 18 - Per-subject metrics for the data fusion feature level model (1s-w)

Subject	Label	Precision	Recall	F1-score
1	0	0.87	1.00	0.93
	1	1.00	0.85	0.92
5	0	1.00	1.00	1.00
	1	1.00	1.00	1.00
6	0	1.00	1.00	1.00
	1	1.00	1.00	1.00
8	0	1.00	0.95	0.97
	1	0.95	1.00	0.98
10	0	0.70	0.97	0.81
	1	0.94	0.59	0.72

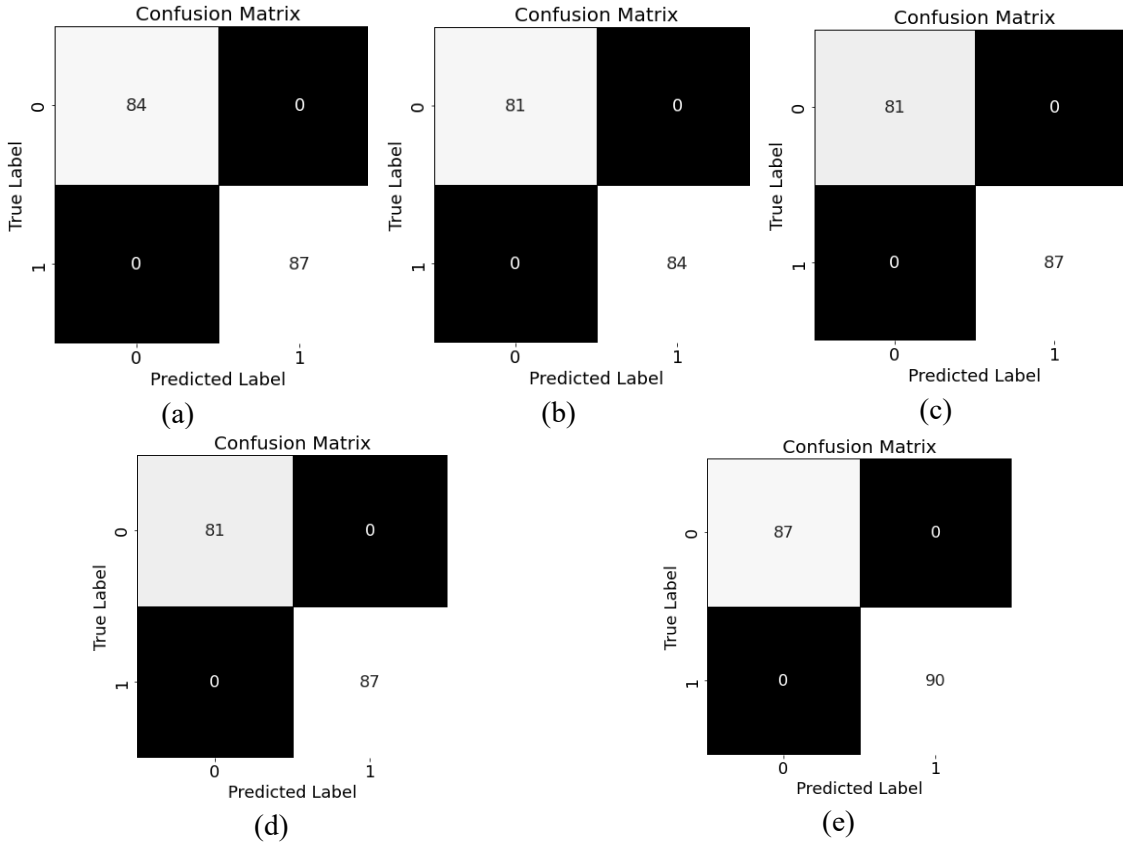
Source: The author (2024).

Subject 8 also showed high performance, with an accuracy of 98%, precision of 0.95 for awake states, and an F1 score of 0.98. Subject 10 exhibited a lower accuracy of 78%, with a precision of 0.94 for awake states, but a significantly lower recall of 0.59. This suggests that the model struggled to accurately identify all instances of drowsiness for this subject.

6.5.2. Presentation of results (5-second window)

The results for the feature level model with a 5-second window are presented in the following confusion matrices (Figure 1) and classification reports (Table 19) for each subject:

Figure 36 - Confusion matrix from data fusion feature level model (5s-w) per subject (1, 5, 6, 8, 10, for (a), (b), (c), (d), and (e), respectively)



Source: The author (2024).

Table 19 - Per-subject metrics for the data fusion feature level model (5s-w)

Subject	Label	Precision	Recall	F1-score
1	0	1.00	1.00	1.00
	1	1.00	1.00	1.00
5	0	1.00	1.00	1.00
	1	1.00	1.00	1.00
6	0	1.00	1.00	1.00
	1	1.00	1.00	1.00
8	0	1.00	1.00	1.00
	1	1.00	1.00	1.00
10	0	1.00	1.00	1.00
	1	1.00	1.00	1.00

Source: The author (2024).

The feature level fusion model with a 5-second window also demonstrates high accuracy across all subjects. Key observations include:

- Subjects 1, 2, 6, 8, and 10: Achieved near-perfect precision, recall, and F1-scores, indicating the model's ability to maintain high performance over longer windows.
- The 5-second window allows for the capture of more comprehensive patterns over time, which may contribute to the slight improvement in performance metrics compared to the 1-second window model but suffers from a decrease in operational performance due to having to process five times more data.

Overall, the feature level fusion model, whether using a 1-second or 5-second window, shows significant improvements over the decision level model. This approach leverages the combined strengths of both EEG and facial data, leading to a more robust and reliable drowsiness detection system.

6.5.3. Comparison with decision-level model

When comparing the results from the feature-level model to the decision-level model presented in section 6.3, several key differences and improvements can be observed:

- Overall Accuracy:

The feature level model generally shows higher accuracy across most subjects compared to the decision-level model. For example, subjects 5 and 6 achieved 100% accuracy with the feature-level model (1-second window), while their performance was slightly lower with the decision-level approach. For the 5-second window models, the results are better, but with a computational increase compared to the 1-second model.

- Precision and Recall:

The feature level model provides higher precision and recall for most subjects. Specifically, for Subjects 5 and 6, indicating an excellent ability to correctly identify both drowsy and awake states without any false positives or false negatives.

For Subject 10, however, the feature level model shows a noticeable drop in recall for the awake state (0.59) compared to the decision-level model; for the 5-second window model, this issue is resolved.

- F1-Score:

The F1-scores, which balance precision and recall, are generally higher for the feature level model. This indicates a more balanced performance in identifying drowsy states accurately.

Subject 8, for example, had an F1-score of 0.98 with the feature level model, demonstrating a significant improvement over the decision-level approach.

- Error Analysis:

The feature level model reduces the type II errors (false negatives) significantly for most subjects, ensuring that fewer instances of drowsiness go undetected. This is critical in safety-critical environments where missing a drowsy state can lead to severe consequences.

The feature level model's ability to aggregate and leverage both EEG and facial data features results in more accurate and reliable drowsiness detection. This enhanced performance is particularly valuable in industrial environments, where maintaining high levels of alertness is crucial for safety and operational efficiency.

By integrating data at the feature level, the model benefits from a richer representation of the underlying patterns associated with drowsiness. This approach can significantly improve the robustness of drowsiness detection systems, reducing the risk of accidents and enhancing overall safety.

6.6. DrowsinessNET: A proposal of a web-based friendly interface to detect drowsiness

Overall, the proposed methodology for identifying drowsiness risk factors has shown promising results, indicating its potential usefulness as a tool to support safety personnel and operators in high-risk industries. In line with this objective, we have developed a user-friendly web application called DrowsinessNET that incorporates trained classifiers based on our proposed model.

The aim of DrowsinessNET is to assist operators and safety personnel in identifying and assessing drowsiness-related scenarios, particularly in environments where sustained attention and vigilance are crucial, such as control rooms in industries like oil and gas. The application operates by allowing users to input real-time biological data, including video images and/or EEG signals, to predict drowsiness levels accurately. DrowsinessNET utilizes advanced machine learning models to provide real-time predictions of drowsiness levels, enabling users to take proactive measures and, then, prevent potential safety incidents.

In addition to real-time drowsiness prediction, DrowsinessNET includes a feature that allows users to upload recorded video and/or EEG data for testing and training the

prediction model. This feature provides users with the flexibility to customize and refine the prediction model based on their specific requirements and scenarios.

By offering a streamlined approach to drowsiness prediction, DrowsinessNET simplifies the process of assessing and managing drowsiness-related risks in high-risk environments. Rather than relying solely on subjective assessments, operators and safety personnel can harness the application's predictive capabilities to identify and mitigate potential safety hazards proactively.

DrowsinessNET is developed in Python, leveraging libraries such as Streamlit for web application development, Keras for building machine learning models, and other essential tools like TensorFlow for deep learning tasks (Chollet et al., 2015). The application is structured into four main sections: i) What is DrowsinessNET, ii) User's Guide, iii) Data Upload, and iv) Real-Time Drowsiness Detection. As depicted in Figure 37, a sidebar menu on the left side of the screen allows users to navigate between these sections.

Figure 37 - What is DrowsinessNET

Source: The author (2024).

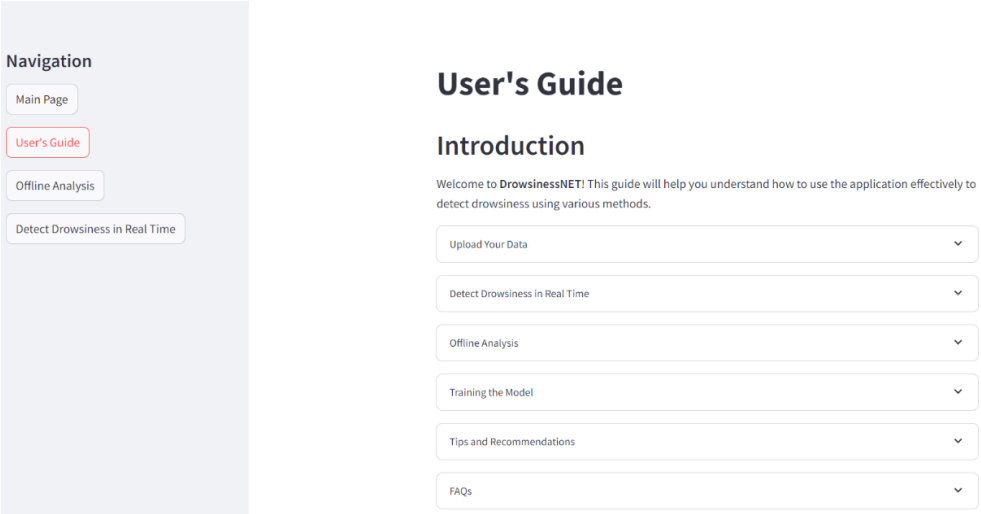
The first section provides an overview of DrowsinessNET and its capabilities. The primary objective of the application is to predict drowsiness levels based on real-time biological data input by users. By analyzing video images and EEG signals, DrowsinessNET identifies potential drowsiness episodes, enabling users to take preventive measures to ensure safety.

6.6.1. User's Guide

The User Manual section provides comprehensive guidance on how to navigate and utilize DrowsinessNET effectively. It includes step-by-step instructions on accessing different features of the application, such as uploading data, configuring settings,

interpreting results, and troubleshooting common issues (Figure 38). Users can refer to this section to understand the functionalities of DrowsinessNET and make the most out of its capabilities.

Figure 38 - User's guide interface

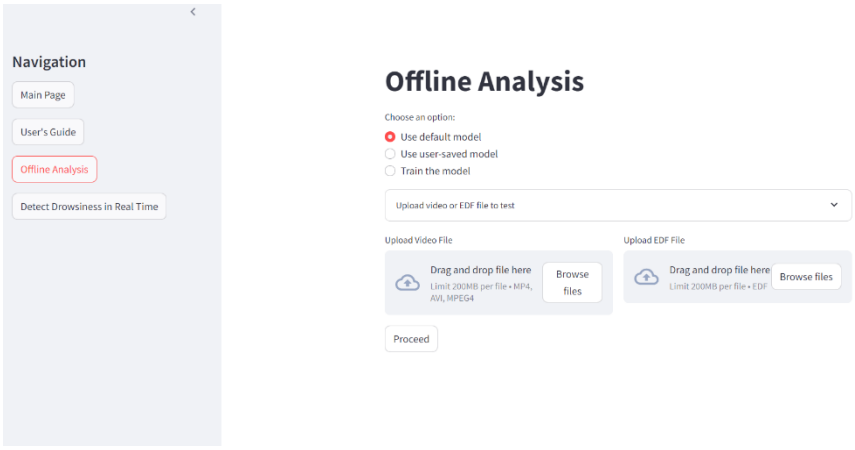


Source: The author (2024).

6.6.2. Data Upload

In the Data Upload section, users can easily upload their biological data, typically consisting of video images and/or EEG signals, to initiate the drowsiness detection process. The interface allows users to select and upload files from their local storage or input data from compatible external devices (Figure 39). DrowsinessNET seamlessly processes the uploaded data, extracting relevant features and utilizing machine learning algorithms to predict drowsiness levels accurately. Upon selecting the files, DrowsinessNET performs a synchronization check to ensure temporal alignment between the two input streams. For instance, if the EEG data consists of a 10-minute recording and the video data also spans 10 minutes, the synchronization process ensures that corresponding moments in both streams are matched, which is essential for accurate drowsiness detection.

Figure 39 - Data upload interface

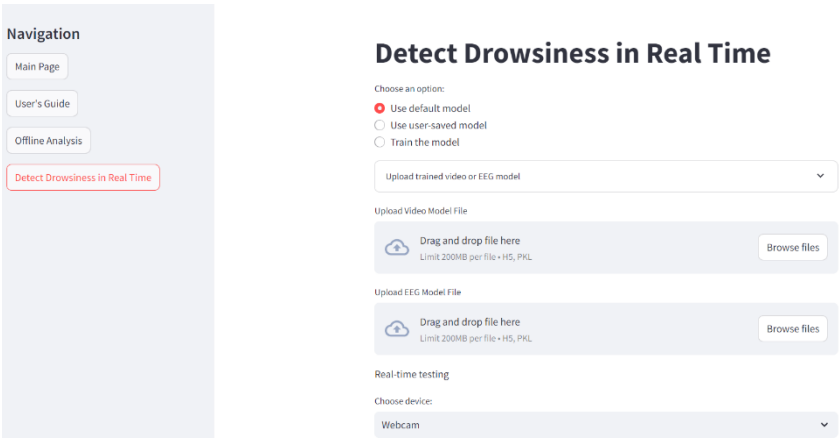


Source: The author (2024).

6.6.3. Real-Time Drowsiness Detection

The Real-Time Drowsiness Detection feature enables users to monitor drowsiness levels continuously in real-time. By interfacing with live video feeds and EEG devices, DrowsinessNET analyzes incoming data streams to detect signs of drowsiness as they occur (Figure 40). Users can observe drowsiness predictions in real-time through intuitive visualizations and receive timely alerts or notifications to take appropriate actions, such as initiating rest breaks or alerting supervisors. This functionality enhances situational awareness and enables proactive interventions to mitigate the risks associated with drowsiness in high-risk environments.

Figure 40 - Real-time drowsiness detection interface



Source: The author (2024).

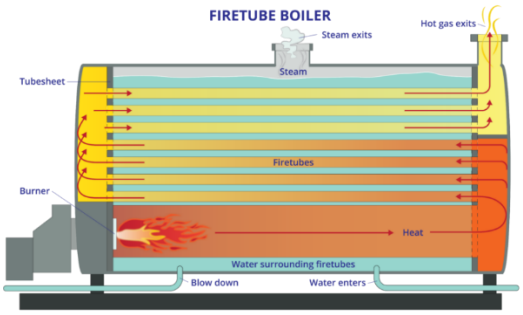
7. *OPERATOR PERFORMANCE IN AUTOMATED OIL AND GAS OPERATION*

Part of this chapter has been presented at 33rd European Safety and Reliability Conference (Esrel) (Ramos et al., 2023), and describes the proposed experiment that aims at analyzing the human performance of control room operators of automated O&G operations. The experiment objective is to evaluate the impact of automation-related factors on operator performance based on a microworld simulation, with variables controlled and measured. The experiment in an O&G process proposes to investigate whether operators can maintain a water level within the pre-established limits to have an acceptable vapor pressure level in the High Pressure (HP), Medium Pressure (MP) and Low Pressure (LP) steam collectors. Additionally, we will collect information on automation complacency, human errors, and assess the level of attention/drowsiness among operators.

7.1. Experiment description

The experiment, inspired by real oil refineries, simulates a hypothetical steam system comprising two gas turbines with dedicated Heat Recovery Steam Generators (HRSG) and a three-pressure steam distribution system. In addition, the system includes a steam generator boiler, a boiler drum supply and drain system, a burning system, and a heated gas output system; Figure 41 shows the steam system.

Figure 41-Schematic representation of the steam system in a refinery.



Source: Adapted from MIURA (2023).

The fired boiler is part of a process in which a liquid is heated and vaporized. Normally, water serves as the working fluid, and the separation into liquid water and steam takes place in the boiler drum. In this type of process, it is extremely important to

control the water level in the boiler drum, which cannot be too high or too low. Thus, for our simulation, some simplified scenarios were proposed:

- If the water level is low: it can disrupt the circulation of steam within the system, potentially causing adverse effects on the pipes;
- If the water level is extremely low: the boiler can run dry, causing mechanical damage to the equipment and not taking steam to the parts needed in the rest of the process (causing major consequences);
- If the water level is high, it can affect the steam's purity and result in more water droplets entering the superheater with the saturated steam. To vaporize these water droplets, more heat will be required, increasing the thermal load, also causing adverse effects on pipes;
- If the water level is extremely high, some amounts of liquid may be carried downstream and damage the equipment.

Thus, the operator and/or the system must act to maintain standard operational control. Keeping the boiler drum level at 50% during operation is normally the standard. Thus, the actions taken are presented in Table 20 depending on the scenario.

Table 20 – Summary of scenarios, consequences and actions that should be taken.

Scenarios	Consequences	Actions
Low water level	Tubes can be affected by a lack of optimal steam circulation;	Open water injection valve
Too low water level	Boiler running dry, causing mechanical damage to the equipment;	Open water injection valve
High water level	Steam purity is affected, increased heat load to vaporize water droplets, pipes affected;	Open drain valve
Too high water level	Transported liquids causing mechanical damage to the equipment.	Open drain valve

Source: The author (2024).

7.2. Process Plant Simulation

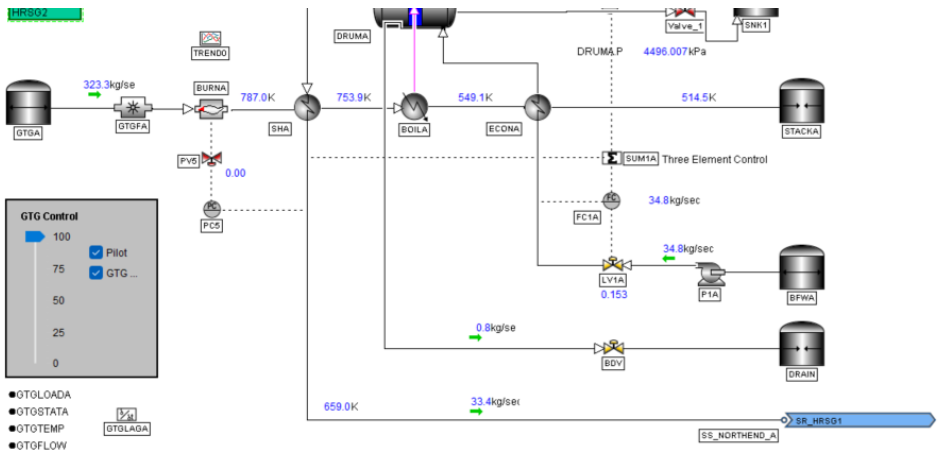
The plant is simulated using AVEVA, a dynamic simulation software (AVEVA 2020), which focuses on dynamic simulation studies and emphasizes key modeling assumptions and expected results. The software was selected for the process plant

simulation due to its robustness and industry recognition as a reliable tool for dynamic modeling in the oil and gas industry. AVEVA provides an intuitive interface and features designed to replicate complex operating environments, which align well with the needs of this study.

Other options, such as Aspen HYSYS (AspenTech, 2024) and CHEMCAD (Chemstations, 2024), were considered during the study planning phase. While these tools also offer interesting features, AVEVA was specifically chosen for its ease of integration with other software used in this study (subsection 7.4) and its ability to simulate dynamic process behaviors with a high degree of accuracy. From this, among the thirteen default processes available in the software, we specifically utilize the Steam Drum Three Element Control simulation for our study.

A gas turbine generator/heat recovery steam generator (GTG/HRSG) is shown in Figure 42. High-pressure steam is produced in the HRSG and sent to the Refinery North End via the gas turbine exhaust, which has a temperature of 958ºF. A superheater, a boiler, an economizer, and a steam drum constitute the HRSG that supports the supplemental firing of natural gas using excess oxygen in the gas turbine exhaust. The steam drum level is maintained with a three-element control system. Table 21 comments on the main components of the system.

Figure 42- Dynamic simulation of a refinery steam from AVEVA



Source: The author (2024).

Table 21 - Main components of the system

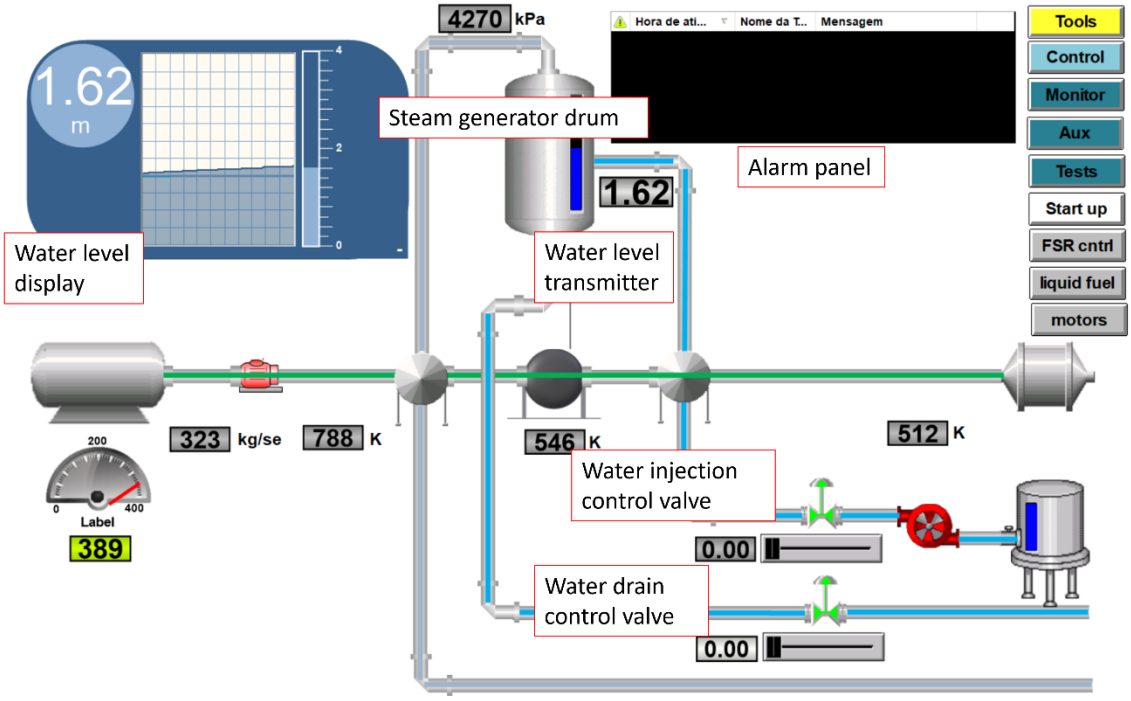
Name	Component	Unit
SHA	Superheater A	K
DRUMA	Drum A	kPa
BOILA	Boiler A	K
ECONA	Economizer A	K
HRSG	Heat recovery steam generator	-
GTG	Gas turbine generator	Kg/sec

Source: The author (2024).

7.3. Human-System Interface

The system is deemed to be highly automated, but it still needs to be monitored by an operator. The control and management of a system in a petrochemical industry can be performed with a graphical user interface known as a control panel or dashboard. Thus, we implemented through Wonderware InduSoft web studio software (AVEVA, 2023) a control panel that schematizes the simulation (Figure 43).

Figure 43-Preliminary dashboard proposed in InduSoft



Source: The author (2024).

In this type of control mechanism, water is fed into the boiler drum through one or more pipes and, therefore, through one or more control valves. Only the water level in the drum is measured using a level transmitter and the information is sent to the controller. The information is compared with the set point and, then, the control valves are manually operated to increase or decrease the flow of water inside the boiler drum.

7.4. Integration with Matrikon OPC Server for Simulation

To ensure seamless communication between the simulated plant components and the control system interface, the Matrikon OPC Server for Simulation was utilized. Matrikon OPC is an industry-standard software that facilitates Open Platform Communications (OPC), enabling real-time data exchange between different systems, including simulation environments and human-machine interfaces (HMIs).

7.4.1. Role of Matrikon OPC in Simulation

The Matrikon OPC Server for Simulation acts as an intermediary between the AVEVA dynamic simulation software and the Wonderware InduSoft Web Studio control panel. In this setup, Matrikon OPC serves as a bridge, ensuring that data from the simulated plant's key variables—such as water level, steam pressure, and valve status—are communicated to the operator's dashboard in real time.

- Data Flow: The OPC server continuously transfers critical data points such as boiler drum water levels and HRSG component statuses to the control panel. The operator can then monitor and adjust these variables based on the real-time feedback provided by the system.
- Interoperability: One of the advantages of Matrikon OPC is its interoperability with various control systems and platforms, ensuring compatibility between the AVEVA software and the InduSoft interface. This allows for the creation of a realistic control room environment, where real-time decisions can be simulated based on the data generated by the plant model.

7.4.2. Importance of OPC for Human Performance Analysis

The use of the Matrikon OPC Server is critical for accurately simulating the operator's role in controlling the refinery system. By providing real-time data, the server ensures that operators are exposed to realistic operational conditions, including varying water levels in the boiler drum and potential system errors, as reflected in the dashboard

interface. This level of integration allows for a thorough evaluation of human performance under both normal and emergency conditions.

- Event Handling: The OPC server also facilitates the triggering of system events, such as automation failures or process disturbances, which the operator must respond to. This allows for a detailed analysis of reaction times, decision-making processes, and human errors in a highly automated environment.
- Scalability for Future Experiments: Additionally, Matrikon OPC provides scalability for future experiments. More complex scenarios, including multiple process variables or the introduction of additional error conditions, can be integrated into the simulation environment with minimal modifications to the OPC server configuration.

7.5. Control and measurement variables of the experiment

The experiment aims to assess the impact of certain factors related to an automated system on human performance. Firstly, human performance must be defined in the context of the task. In this task, an adequate performance is to open or close the valves when needed by the plant, where the water level significantly deviates from the optimal range, either exceeding predetermined upper (water level is high or too high) or lower (water level is low or too low) thresholds.

Human Failure Events (HFEs) in this task are: (1) opening/closing the valve too early or too late, (2) manipulating the incorrect equipment, or (3) not performing the action. These HFEs are translated into measurable variables: reaction time and correct action (Table 22). Reaction time refers to the time interval between the moment a deviation (e.g., water level change) occurs in the system and the moment the operator initiates an action. Correct action refers to the appropriateness and accuracy of the operator's response to a given deviation. A correct action is characterized by performing the right operation (e.g., opening or closing the correct valve) at the appropriate time to address the system issue effectively. Furthermore, the factors to be analyzed must be translated into control variables, namely, monitoring time, task complexity, and failure in automation, further described in following sub-sections.

Table 22 – Experiment's control and measurable variables

Measurable variables	Reaction time
	Correct action
Control variables	Monitoring time
	Task complexity
	Failures in automation

Source: The author (2024).

7.5.1. Monitoring time

In automated operations, it is common for operators to go for extended periods without any actions required on their part, leading to a potential decrease in situational awareness. Situational awareness, as defined by Endsley (1995), involves perceiving elements in the environment, understanding their meaning, and projecting their status into the near future. The lack of immediate stimuli or occurrences can cause complacency, making it challenging to be prepared to quickly recognize newly emerging problems.

When monitoring automated systems, it frequently takes people a long time to notice that a situation calls for action, and even longer to fully comprehend it and respond appropriately (Endsley, 2017). This gap between problem detection and action can result in critical delays and negatively impact the effectiveness of needed interventions. Therefore, it is essential to adopt strategies to mitigate the decrease in situational awareness.

7.5.2. Task complexity

In general, procedures lessen the likelihood that human operators would forget or skip an activity they need to do, reduce their physical and/or cognitive workload by providing explicit instructions, and maintain their performance over time. However, complicated procedures (i.e., incomplete, inaccurate, inconsistent, or difficult to understand) reduce human performance, suggesting that tasks' complexity has an impact on it (Jang et al., 2021).

The level of complexity also can negatively affect human performance. Hence, there are points of attention that the operator must focus on; however, there are several other points that bring information about the system but do not necessarily impact its action. For example, when the water level is rising (or decreasing), the operator can see the variation in the water level display (right side of the steam generator drum) and in the

water level over time (left side of the steam generator drum) shown in Figure 43. The system is also proposed to alarm (box at the bottom left) if the level exceeds certain pre-established limits.

7.5.3. Failures in automation

In our example, automation is designed to function correctly for a period, to build operator' confidence in the system, which may lead to complacency. Automation complacency is a critical topic in automation safety. Complacency manifests in inadequate monitoring and checking of automated functions, as exemplified by pilots who overly rely on their autopilot's proper functioning and consequently fail to monitor and check it appropriately (Bahner et al., 2008). According to (Parasuraman & Manzey, 2010), automation complacency is found for both inexperienced employees and experts and cannot be overcome with simple training practice and can affect decision-making in individuals as well as in teams.

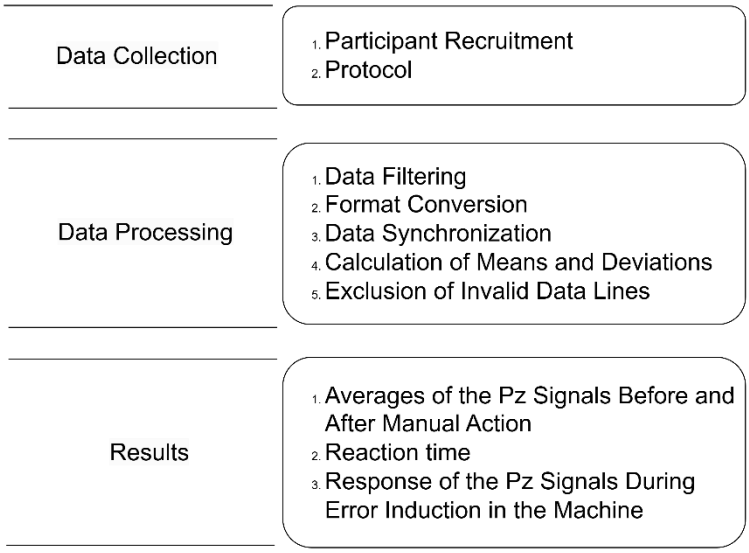
Thus, we randomly induce an automation error over time, which can only be noticed when one of the events occurs and the visual alarm does not go off. The event is implemented in the simulation itself, to induce a reduction in the water level, behaving like a leak.

In our simulation, the automation error will perhaps be interpreted by a communication error between the monitored equipment and the control panel, where even with the water level exceeding the pre-established limits, the alarm is not activated. However, the variation in the dashboard display (right side of the steam generator drum) and in the water level over time (left side of the steam generator drum) shown in Figure 43 will continue to be reported correctly. We can assess participant' reaction time and actions to determine their perception of water level changes without relying on the alarm and evaluate their accuracy.

7.6. Material and methods

We conducted an experiment to collect EEG and facial data to observe participants' reactions to the outcomes. The data collected will also be used to create a database for monitoring alertness levels. The research methodology flowchart is presented in Figure 44. Details of the experiment's data collection, processing, and results are described below.

Figure 44- Flow chart of research methodology



Source: The author (2024).

7.6.1. Data collection

- Participants

Twelve researchers participated in the experiment (mean age ± standard deviation: 24.47 ± 3.59 years). They were instructed to get their usual amount of sleep the night before the experiment to avoid any effects on drowsiness. Individuals without any alcohol dependence, medication use, or any sleep disorder were required. They were not informed in advance of the automation failure in the experiment and participated without any preconceptions about the expected results.

- Protocol

The experiment was conducted in Recife - PE - Brazil, with the approval of the internal Ethics Review Committee, process number 69415423.8.0000.5208, confirming the protection of all participants in scientific research to safeguard the dignity, rights, safety and well-being of each one of them without distinction, in an ethical and responsible manner.

Initially, a brief training was carried out with individuals who were willing to participate in the experiment. The training consisted of explaining, through a video recording, which parameters (temperature, pressure, flow) the individual needs to pay great attention to in relation to their limits, activating a button in the simulated program when observing any abnormality (parameters above or below the pre-established specification limit).

The data collection procedure began shortly after the participant setup and training time; collection took place while the participants carried out the process of responding to the simulation through inspection operations with different complexities when they agreed that they were in a state of alert, without any signs of drowsiness. This simulation took around 30 min.

A webcam installed in front of the participants captured their face at 30 frames per second (fps), while they were in the simulation. Furthermore, EEG headsets EMOTIV INSIGHT (EMOTIV, 2024) are designed for scalable and contextual research of the human brain and provides access to brain data (Figure 45). The EMOTIV product is intended for research and personal use applications only. It is not designed or intended to be used for the diagnosis or treatment of diseases, nor for medical purposes; therefore, it does not require a specialized professional to configure it. The EMOTIV INSIGHT, with its long-life semi-dry polymer and user-friendly design, is ideal for research contexts and personal use applications.

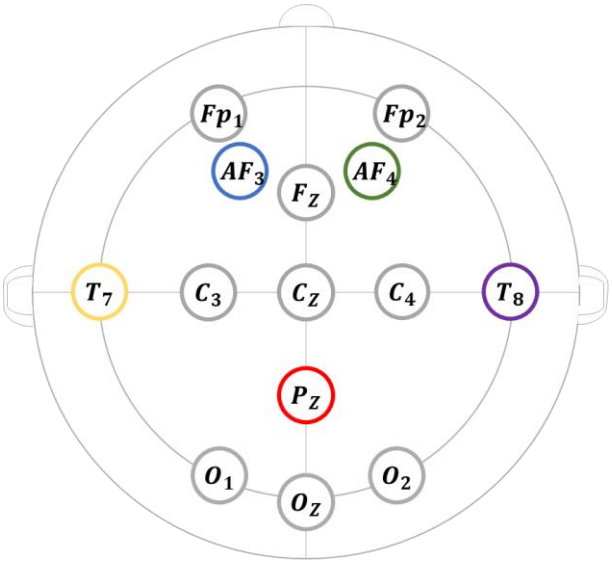
Figure 45 - Participant experimental setup

Source: The author (2024).

The Emotiv INSIGHT 2 features 5 EEG signal capture channels, positioned at AF3, AF4, T7, T8, and Pz locations, according to the 10-20 electrode placement system (Figure 46). It performs sequential sampling, with a rate of 128 SPS (internal rate of 2048 Hz) and a 16-bit resolution, where 1 LSB equals 0.128 µV. The device operates within a bandwidth of 0.5 to 45 Hz, using digital notch filters at 50 Hz and 60 Hz to reduce interference. Connectivity is via Bluetooth 5.0, which is backward compatible, and the 430 mAh lithium polymer (LiPo) battery provides up to 20 hours of autonomy. The device also measures impedance in real-time to evaluate the quality of electrode-skin contact, using a patented system. These characteristics make the Emotiv INSIGHT 2 a robust and

efficient tool for EEG data collection in research contexts, especially in environments like the proposed experiment, where monitoring cognitive states is essential.

Figure 46 - Electrode Locations in the 10-20 EEG System Used by the Emotiv INSIGHT 2



Source: The author (2024).

However, in industrial settings like oil and gas, there's a growing interest in wearable EEG technology, including dry electrode systems or helmet-mounted devices. These innovations offer potential advantages for operators, providing non-invasive and comfortable monitoring of cognitive states, such as drowsiness, without disrupting their workflow. AVEVA Software – Dynamic Simulation 2021 was used to execute critical safety operation simulation, while Wonderware InduSoft web studio software was adopted to present a control panel that outlines the simulation and where the experiment participants performed the actions.

7.6.2. Data processing

To ensure the accuracy and integrity of the data collected during the experiments, several processing steps were adopted, which include:

- Data Filtering: Initially, the raw EEG data and process variables were filtered to remove noise and artifacts that could distort subsequent analyses.
- Format Conversion: The collected data were converted from CSV files to Excel format, facilitating manipulation and analysis using Python language tools and programs.

- Data Synchronization: The synchronization of EEG data and process variables was performed using timestamps. This step is crucial to ensure that subsequent analyses correctly reflect the experimental conditions at specific times.
- Calculation of Means and Deviations: For each timestamp, the means and standard deviations of the EEG signals that were sampled at 128 μV were calculated. This allows for a more detailed analysis of the variations in the signals during the experiment.
- Exclusion of Invalid Data Lines: Data lines that contained missing or inconsistent values were excluded to ensure the quality of the analyzed data.

To ensure transparency and reproducibility, the entire dataset used in this thesis will be made available online, enabling other researchers to access and cite it. The database can be found at the following link: DrowsinessNET database.

7.6.3. Results

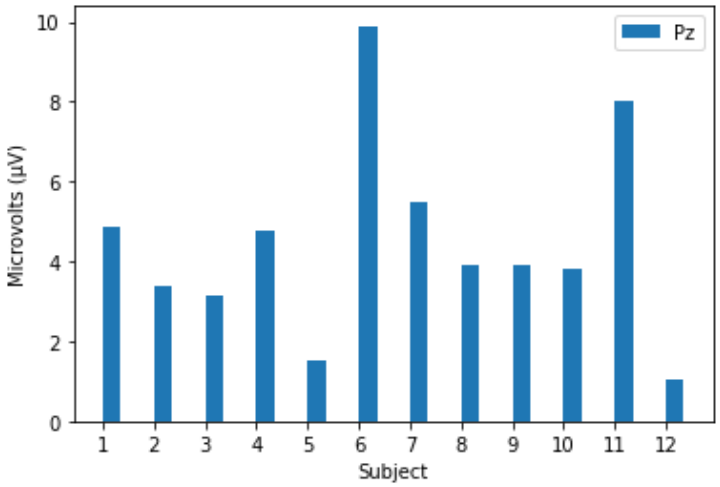
This section is divided into three parts to discuss the results obtained based on the changes in the EEG signals and the process variables.

7.6.3.1. Averages of the Pz Signals Before and After Manual Action

During the analysis of the collected data, it was observed that the Pz signals presented significant changes before and after the manual action of the operator, that is, when there was a manual increase in the water level in the drum. For each change point identified, the data were segmented into two periods: (1) the 10 records prior to the change, and (2) the 10 records after the change, with each record representing one second.

The average of the EEG signals (AF3, T7, Pz, T8, AF4) was calculated for these periods, and the differences between the averages were analyzed for each subject. Figure 47 shows the changes in Pz signals, which are primarily associated with sensory processing and information integration.

Figure 47 - Average changes in Pz signals for subjects



Source: The author (2024).

For the sake of brevity, our analysis focused on the Pz signal. The decision to prioritize Pz stems from previous sections (see Section 4), where the central-parietal region was identified as one of the key areas associated with sensory integration and attention in similar tasks. The Pz signal is considered representative in detecting cognitive engagement during manual interventions, making it a suitable choice for this analysis. The increase in the central-parietal (Pz) signal following manual action reflects a heightened level of attention and decision-making by the operator. This brain region is known to be involved in integrating sensory information and directing motor actions, which aligns with the cognitive processes needed to manually adjust the water levels in the system. These findings suggest that operators become more engaged when transitioning from passive monitoring to active control.

While the Pz signal changes were observed across all subjects, the degree of change may have varied. This could be attributed to individual differences in attention, experience, or fatigue levels. Subjects with more significant changes in the Pz signals might have had higher engagement or were more attentive to the task at hand. In contrast, those with lower changes could indicate less cognitive activation, potentially due to complacency or fatigue. Investigating the variability between subjects could help identify patterns in operator performance and situational awareness under different levels of automation.

In addition, although the Pz signal was a focal point in this analysis, other channels such as AF3, T7, T8, and AF4 also provide valuable insights into operator performance.

These regions correspond to frontal and temporal areas associated with executive function, attention, and motor control. For the sake of completeness, the graphs representing the data from these additional channels have been included in the APPENDIX F - Bar Plot Representation of EEG Channels (AF3, T7, T8, AF4), allowing for further exploration and comparison of their relevance in relation to operator engagement and task performance.

7.6.3.2. Reaction Time

The analysis of operator reaction time focused on two simulated events designed to assess their responsiveness to abnormalities in water levels. These events were intended to evaluate the operators' situational awareness, particularly under conditions where system automation failed to provide explicit alerts.

- Events Conducted:

1. First Event:
    - A simulated abnormality in water levels required manual intervention by the operator.
    - Operators were expected to detect the deviation and take corrective actions to stabilize the system promptly.
    - This event established a baseline for reaction times under normal conditions.
2. Second Event:
    - A simulated automation error was introduced at a later stage of the simulation.
    - During this event, the water level exceeded pre-established limits without triggering any alarms.
    - Operators had to rely solely on dashboard displays and visual cues to detect the anomaly and respond appropriately.
    - This event specifically tested the operators' ability to identify and respond to system errors without automated support, highlighting their situational awareness and cognitive engagement.

Throughout the experiment, two out of twelve subjects did not react to the second event, while other two reacted before the expected time. This reduced the number of valid subjects for subsequent statistical analysis to eight, as shown in Table 23.

Table 23 - Reaction Times in seconds of Subjects Before and After the Second Event

Event	Subj 2	Subj 3	Subj 4	Subj 6	Subj 8	Subj 9	Subj 10	Subj 11
1st	6	4	5	2	2	2	3	4
2nd	9	5	7	3	5	2	4	5

Source: The author (2024).

To investigate whether there was a significant change in reaction times before and after the second event, a Wilcoxon signed-rank test was conducted. The Wilcoxon test is appropriate for this type of analysis because it evaluates differences between two distinct moments for the same group of subjects without assuming that the data follow a normal distribution.

- $\boldsymbol{H_0}$: There is no significant difference in operator reaction times before and after the second event.
- $\boldsymbol{H_1}$**:** There is a significant difference in operator reaction times before and after the second event.

The test results indicated a p-value of 0.0158. This p-value, being lower than the conventional significance level of 0.05, allows us to reject the null hypothesis. Thus, the analysis suggests that the operators indeed exhibited different reaction times after the second event.

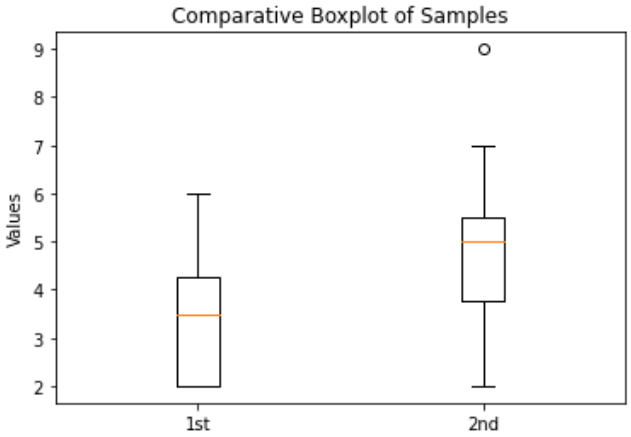
To visualize the differences in reaction times for the two events, a comparative boxplot was created (Figure 48), which displays the distribution of reaction times for both periods. The median reaction time is visibly higher after the second event, indicating an overall increase in response time. The interquartile range (IQR) for the first event is slightly larger, suggesting greater variability in reaction times. Additionally, an outlier is observed in the second event, which represents a subject who had an atypically high reaction time compared to others.

The increase in reaction time and variability following the second event may be attributed to multiple factors. One possible explanation is fatigue, which could have been accumulated as the simulation progressed, slowing down operator responses. Similarly, cognitive overload may have set in, as operators were required to process some tasks over time, leading to reduced responsiveness.

Another important factor to consider is situational awareness, as defined by Endsley (1995), refers to an operator's ability to perceive, understand, and project elements of the environment in time to respond effectively. In automated environments, situational awareness can degrade over time, particularly when operators are exposed to

repetitive tasks or prolonged periods of passive monitoring. This could explain the longer reaction times observed after the second event, where operators may have struggled to maintain the same level of attentiveness as in the earlier stages of the simulation.

Figure 48 - Comparative Boxplot of Reaction Times Before and After the Second Event



Source: The author (2024).

Furthermore, the failure to react by some operators, as well as premature reactions, may suggest a breakdown in their situational awareness, where they either missed the critical cues or misinterpreted system information. Such lapses in awareness can delay or hinder necessary corrective actions, highlighting the importance of continuously monitoring and maintaining a high level of situational awareness in automated environments.

However, it is important to recognize that the nature of the simulation itself may have also influenced the operators' behavior. Since participants were aware that the experiment posed no real-world consequences, some may have adjusted their strategies accordingly. For instance, they may have delayed their reactions intentionally during the second event, or approached it with less urgency, knowing it was part of a controlled environment.

These results underline the need for mitigation strategies that enhance situational awareness and reduce the cognitive load on operators, such as incorporating decision support systems, providing regular stimuli to maintain alertness, or improving training programs focused on handling automation complacency. Addressing these factors can lead to better performance and increased safety in critical operational settings.

7.6.3.3. Cognitive Load and Pz Signal Responses During Automation Errors

When the machine was induced into an error state, it was observed in subjects 1 and 12 that the Pz signals exhibited increases, even though no manual intervention was performed by either subject. The change that occurred in the water variable may have been perceived by the display and/or dashboard, but it was not shown in the alarm. The results are presented below:

Table 24 - Pz Signal Changes Before and After Perception of Machine Error

Subjects	**Before Perception of Change (µV)**	**After Perception of Change (Without Manual Actuation) (µV)**	**Differences in Means (µV)**
1	4155.67	4158.70	3.03
12	4080.06	4082.44	2.38

Source: The author (2024).

The increase in the Pz and T7 signals during the error state, despite no manual intervention, suggests that both operators experienced an increase in cognitive load as they processed the information presented by the system. This increase may reflect the mental processing required to interpret the information displayed on the dashboard, even in the absence of immediate action. The changes in this signal could indicate that the operators were preparing for a possible intervention but chose not to act due to the lack of a corresponding alarm.

The absence of manual action in response to the error detected may reflect a reliance on automation or alarm systems. Operators may delay their response when they do not receive a clear signal (such as an alarm) that intervention is necessary. This could be an indication of automation complacency, where the operator trusts the automated system to handle the situation. The increase in EEG signals may reflect the operator's uncertainty about whether to intervene, which could explain why no action was ultimately taken.

However, this behavior was not consistent among the other subjects. Eight operators responded to the automation error despite the absence of an alarm, indicating that not all participants relied on the alarm system to the same extent. These operators acted based on the visual displays or their perception of changes in the system, without

waiting for a formal alarm, demonstrating a higher level of situational awareness and a more proactive approach to managing system errors. Additionally, two operators adjusted the water levels before the thresholds were exceeded, effectively preempting the error but failing to recognize the automation malfunction. This variability in responses underscores the differences in operator performance and situational awareness when dealing with system errors, highlighting the importance of individual cognitive engagement and attention in automated environments.

Finally, this raises important considerations regarding the design of alarm systems in automated environments. Situational awareness can be compromised if operators become overly reliant on alarms rather than on their own interpretation of system data. This also indicates that visual displays and dashboards play a critical role in maintaining situational awareness, even when alarms are not activated. These results highlight the importance of monitoring EEG signals and process variables to better understand operator performance in automated oil and gas environments.

8. CONCLUDING REMARKS

This thesis focused on developing a comprehensive methodology that can be applicable to various companies with critical safety systems, aiming to enhance both work safety and process safety. The findings of this study support the notion that valuable information extracted from human operators can be automatically captured and processed using innovative data-driven solutions, to support human reliability studies and consequently improve risk management and safety in the work/industrial environment. By capturing data from facial information and EEG recordings and processing them with ML/QML/DL algorithms, it is possible to identify early signs of fatigue/drowsiness, such as states that could compromise safety. In this way, companies can proactively implement preventive measures, reducing the risk of accidents and improving both worker safety and the integrity of industrial processes. The results highlight the potential of this methodology to impact safety practices across different industries positively.

Our research advances the identification and prediction of operator drowsiness with emphasis on the development of a robust, multi-source model capable of recognizing several characteristics associated with drowsiness. Chapter 4 explores ensemble machine learning models, automated machine learning approaches. Notably, our unique approach involves creating a ground truth for reaction times by aligning EEG data with specific time windows, providing a comprehensive understanding of the temporal dynamics associated with drowsiness.

The inclusion of quantum methods in this thesis (Chapter 5) represents a groundbreaking contribution to the field of drowsiness detection and human reliability in critical industrial environments. The results obtained suggest that quantum computing, despite still being in its early stages of development, has significant potential to transform the way we approach complex machine learning problems. The use of these models significantly improved the accuracy of drowsiness and inattention detection, showcasing their effectiveness in dealing with real-world datasets.

In this study, we also developed a data fusion approach that combines EEG and facial data (Chapter 6). This fusion is implemented at both the decision and feature levels, resulting in significant improvements in drowsiness detection systems. The experimental results show that combining EEG and facial data provides enhanced recall and accuracy compared to single-modality approaches. The methodology was divided into three parts for the model in decision level: i) the model based on EEG data, ii) the model based on

CV and iii) the final model of data fusion. The first model used features extracted from EEG data using DWT within an MLP, while the second used a pre-trained model ResNet50 and CV techniques focusing on facial images. The experimental results confirmed the effectiveness of the fusion approach. For example, subject-specific results showed accuracy levels exceeding 97% when using EEG-based models, while CV-based models reached over 90% accuracy. The data fusion model, which combined both modalities, achieved recall rates between 72.61% and 98.81%, further validating the complementary nature of EEG and CV data for robust drowsiness detection.

The second approach in data fusion involves, at the feature level, aggregating the features before feeding them into a deep learning model for training and evaluation. The 1-second and 5-second window approaches were compared, and it was found that the 5-second window consistently outperformed the shorter time window in terms of detection accuracy. This approach provided a more detailed understanding of an operator's state, highlighting how data fusion enhances the system's ability to capture subtle signs of drowsiness over time. The results were considered satisfactory, demonstrating the effectiveness of the data fusion approach to improve the accuracy and reliability of drowsiness detection systems. Both fusion approaches offer a more holistic view of an operator's state.

Another contribution is the creation of DrowsinessNET, a real-time web-based application that integrates the developed detection models for practical use in high-risk environments such as control rooms in O&G industries. This application provides real-time monitoring and allows for early detection of drowsiness, enabling timely interventions to prevent accidents. By making the system accessible through a web-based platform, this thesis bridges the gap between theoretical research and practical application, ensuring that safety technologies are readily deployable in real-world industrial operations.

In addition to technological advancements, this thesis also conducted a simulator-based experiment to assess operator performance in automated O&G operations (Chapter 7). The experiment specifically evaluated how automation-related factors, such as overconfidence, boredom, and inattention, influence operator performance. The results revealed that automation, while beneficial in many aspects, can lead to reduced attentiveness and increased human errors in monotonous tasks. This highlights the pressing need for integrating human reliability technologies, such as the developed drowsiness detection systems, into automated safety-critical environments. The findings

emphasize that supervision supplemented by advanced monitoring systems remains crucial even in highly automated industries.

It is worth highlighting that, in addition to its contributions to safety enhancement in critical industrial settings, this thesis acknowledges the paramount importance of adhering to data protection regulations. Specifically, in the context of the Brazilian General Data Protection Law (LGPD), the methodologies developed herein prioritize privacy and ethical considerations. The extraction and processing of valuable information from human operators, involving facial images and EEG recordings, must ensure the confidentiality and protection of sensitive data. Our study is aligned with the principles of the LGPD, emphasizing transparency, security and respect for individuals' privacy rights.

Overall, our study underscores the importance of integrating multiple sources of information for robust drowsiness detection. The combination of EEG data analysis and CV techniques provides a comprehensive understanding of the operator's cognitive state and facial expressions, leading to enhanced accuracy in predicting drowsiness. This approach has significant implications for accident prevention in critical safety systems, as it enables timely interventions to mitigate the risks associated with human error. Furthermore, the implementation of drowsiness detection technologies can have significant environmental, social and financial impacts by supporting the reduction of accidents, thereby minimizing damage to lives, the environment, infrastructure and equipment. By promoting operator alertness, drowsiness detection technologies improve the overall safety of critical industrial operations, creating a safer work environment for operators and all individuals involved, as well as contributing to incident prevention, safeguarding ecosystems and mitigating environmental degradation. In addition, implementing drowsiness detection can help reduce interruptions to industrial operations caused by accidents, whether temporary interruptions or complete shutdowns.

Despite promising results, our study recognizes certain limitations and identifies avenues for future research. Further validation of the proposed approach is needed on larger datasets and in real-world scenarios to ensure its effectiveness in diverse contexts. An important area for future work involves enhancing the existing database with alertness information, which can be used to train unsupervised models, such as autoencoders. This approach would enable the detection of anomalies and the development of more robust and generalized drowsiness detection systems. Additionally, exploring additional modalities such as ECG and EOG, and employing advanced fusion techniques could

potentially increase the accuracy and robustness of the drowsiness detection system Moreover, future research could focus on studying individuals with insomnia, providing insights into how sleep disorders impact drowsiness detection and system performance. The integration of QML models for handling combined data could be explored. These considerations underscore the need for continued research and development efforts to refine and expand the application of drowsiness detection technologies in safety-critical systems.

REFERENCES

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., … Research, G. (n.d.). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems*. www.tensorflow.org.

Abílio Ramos, M., López Droguett, E., Mosleh, A., & Das Chagas Moura, M. (2020). A human reliability analysis methodology for oil refineries and petrochemical plants operation: Phoenix-PRO qualitative framework. *Reliability Engineering & System Safety*, *193*, 106672. https://doi.org/10.1016/j.ress.2019.106672

Acharya, J. N., Hani, A. J., Cheek, J., Thirumala, P., & Tsuchida, T. N. (2016). American Clinical Neurophysiology Society Guideline 2: Guidelines for Standard Electrode Position Nomenclature. *Neurodiagnostic Journal*. https://doi.org/10.1080/21646821.2016.1245558

Almeida, A. G. de, & Vinnem, J. E. (2020). Major accident prevention illustrated by hydrocarbon leak case studies: A comparison between Brazilian and Norwegian offshore functional petroleum safety regulatory approaches. *Safety Science*, *121*, 652–665. https://doi.org/10.1016/j.ssci.2019.08.028

Andrade, M. V. P. (2018). *A DATA COLLECTING FRAMEWORK FOR HUMAN RELIABILITY ANALYSIS VIA GAME ENGINE BASED SIMULATORS*. https://repositorio.ufpe.br/handle/123456789/33596

Andrei, D. M., Griffin, M. A., Grech, M., & Neal, A. (2020). How demands and resources impact chronic fatigue in the maritime industry. The mediating effect of acute fatigue, sleep quality and recovery. *Safety Science*, *121*, 362–372. https://doi.org/10.1016/j.ssci.2019.09.019

ANP. (2023). *Agência Nacional do Petróleo, Gás Natural e Biocombustíveis-ANP SUPERINTENDÊNCIA DE SEGURANÇA OPERACIONAL-SSO Coordenação de Segurança Operacional*.

Arefnezhad, S., Eichberger, A., Fruhwirth, M., Kaufmann, C., & Moser, M. (2020). Driver Drowsiness Classification Using Data Fusion of Vehicle-based Measures and ECG Signals. *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 451–456. https://doi.org/10.1109/SMC42975.2020.9282867

Arshad, S., Akinade, O., Bello, S., & Bilal, M. (2023). Computer vision and IoT research landscape for health and safety management on construction sites. *Journal of Building Engineering*, *76*, 107049. https://doi.org/10.1016/j.jobe.2023.107049

AspenTech. (2024). *Aspen HYSYS: Process Simulation Software.* https://www.aspentech.com.

AVEVA. (2023). *InduSoft Web Studio*. https://www.aveva.com/en/products/edge/

B, V. P., & Chinara, S. (2021). Automatic classification methods for detecting drowsiness using wavelet packet transform extracted time-domain features from single-channel EEG signal. *Journal of Neuroscience Methods*, *347*. https://doi.org/10.1016/j.jneumeth.2020.108927

Bahner, J. E., Hüper, A. D., & Manzey, D. (2008). Misuse of automated decision aids: Complacency, automation bias and the impact of training experience. *International Journal of Human Computer Studies*, *66*(9), 688–699. https://doi.org/10.1016/j.ijhcs.2008.06.001

Bajaj, V., Taran, S., Khare, S. K., & Sengur, A. (2020). Feature extraction method for classification of alertness and drowsiness states EEG signals. *Applied Acoustics*, *163*. https://doi.org/10.1016/j.apacoust.2020.107224

Bekhouche, S. E., Ruichek, Y., & Dornaika, F. (2022). Driver drowsiness detection in video sequences using hybrid selection of deep features. *Knowledge-Based Systems*, *252*, 109436. https://doi.org/10.1016/j.knosys.2022.109436

Belakhdar, I., Kaaniche, W., Djemal, R., & Ouni, B. (2018). Single-channel-based automatic drowsiness detection architecture with a reduced number of EEG features. *Microprocessors and Microsystems*, *58*, 13–23. https://doi.org/10.1016/j.micpro.2018.02.004

Belakhdar, I., Kaaniche, W., Djmel, R., & Ouni, B. (2016). A comparison between ANN and SVM classifier for drowsiness detection based on single EEG channel. *2016 2nd International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, 443–446. https://doi.org/10.1109/ATSIP.2016.7523132

Benedetti, M., Lloyd, E., Sack, S., & Fiorentini, M. (2019). Parameterized quantum circuits as machine learning models. *Quantum Science and Technology*, *4*(4), 043001. https://doi.org/10.1088/2058-9565/ab4eb5

Bhandari, G. M. , A. D. , A. B. , & U. A. (2014). YAWNING ANALYSIS FOR DRIVER DROWSINESS DETECTION. *International Journal of Research in Engineering and Technology*, *03*(02), 502–505. https://doi.org/10.15623/ijret.2014.0302087

Birjandtalab, J., Baran Pouyan, M., Cogan, D., Nourani, M., & Harvey, J. (2017). Automated seizure detection using limited-channel EEG and non-linear dimension reduction. *Computers in Biology and Medicine*. https://doi.org/10.1016/j.compbiomed.2017.01.011

Bradski, G. (2000). The openCV library. *Dr. Dobb's Journal: Software Tools for the Professional Programmer*, *25*(11), 120–123.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*(2), 123–140. https://doi.org/10.1007/BF00058655

Bye, A. (2023). Future needs of human reliability analysis: The interaction between new technology, crew roles and performance. *Safety Science*, *158*. https://doi.org/10.1016/j.ssci.2022.105962

Cabañero-Gomez, L., Hervas, R., Gonzalez, I., & Rodriguez-Benitez, L. (2021). eeglib: A Python module for EEG feature extraction. *SoftwareX*, *15*, 100745. https://doi.org/10.1016/j.softx.2021.100745

Cai, H., Qu, Z., Li, Z., Zhang, Y., Hu, X., & Hu, B. (2020). Feature-level fusion approaches based on multimodal EEG data for depression recognition. *Information Fusion*, *59*, 127–138. https://doi.org/10.1016/j.inffus.2020.01.008

Caldwell, J. A. (2005). Fatigue in aviation. *Travel Medicine and Infectious Disease*, *3*(2), 85–96. https://doi.org/10.1016/j.tmaid.2004.07.008

Camden, M. C., Medina-Flintsch, A., Hickman, J. S., Bryce, J., Flintsch, G., & Hanowski, R. J. (2019). Prevalence of operator fatigue in winter maintenance operations. *Accident Analysis and Prevention*, *126*(January 2018), 47–53. https://doi.org/10.1016/j.aap.2018.01.009

Cerezo, M., Arrasmith, A., Babbush, R., Benjamin, S. C., Endo, S., Fujii, K., McClean, J. R., Mitarai, K., Yuan, X., Cincio, L., & Coles, P. J. (2021). Variational quantum algorithms. *Nature Reviews Physics*, *3*(9), 625–644. https://doi.org/10.1038/s42254-021-00348-9

Chemstations. (2024). *CHEMCAD: Chemical Process Simulation Software*. https://www.chemstations.com.

Cheng, B., Zhang, W., Lin, Y., Feng, R., & Zhang, X. (2012). Driver drowsiness detection based on multisource information. *Human Factors and Ergonomics in Manufacturing & Service Industries*, *22*(5), 450–467. https://doi.org/10.1002/hfm.20395

Cui, J., Lan, Z., Liu, Y., Li, R., Li, F., Sourina, O., & Müller-Wittig, W. (2022). A compact and interpretable convolutional neural network for cross-subject driver drowsiness detection from single-channel EEG. *Methods*, *202*, 173–184. https://doi.org/10.1016/j.ymeth.2021.04.017

da Silveira, T. L. T., Kozakevicius, A. J., & Rodrigues, C. R. (2016). Automated drowsiness detection through wavelet packet analysis of a single EEG channel. *Expert Systems with Applications*, *55*, 559–565. https://doi.org/10.1016/j.eswa.2016.02.041

Daubechies, I. (1992). *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9781611970104

Deng, J., Dong, W., Socher, R., Li, L.-J., Kai Li, & Li Fei-Fei. (2009). ImageNet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–255. https://doi.org/10.1109/CVPR.2009.5206848

Du, L., Lu, Z., & Li, D. (2022). Broodstock breeding behaviour recognition based on Resnet50-LSTM with CBAM attention mechanism. *Computers and Electronics in Agriculture*, *202*, 107404. https://doi.org/10.1016/j.compag.2022.107404

EMOTIV. (2024, January 20). *Insight Insight rotated Insight travel case closed Insight travel case open Insight sensor tips Insight sensor tips with saline solution Insight - 5 Channel Wireless EEG Headset*. Https://Www.Emotiv.Com/Collections/Headsets-Collection/Products/Insight.

Endsley, M. R. (1995). Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *37*(1), 32–64. https://doi.org/10.1518/001872095779049543

Endsley, M. R. (2017). Autonomous Driving Systems: A Preliminary Naturalistic Study of the Tesla Model S. *Journal of Cognitive Engineering and Decision Making*, *11*(3), 225–238. https://doi.org/10.1177/1555343417695197

Eskandarian, A., & Mortazavi, A. (2007). Evaluation of a Smart Algorithm for Commercial Vehicle Driver Drowsiness Detection. *2007 IEEE Intelligent Vehicles Symposium*, 553–559. https://doi.org/10.1109/IVS.2007.4290173

Fahimi, F., Zhang, Z., Goh, W. B., Lee, T.-S., Ang, K. K., & Guan, C. (2019). Inter-subject transfer learning with an end-to-end deep convolutional neural network for EEG-based BCI. *Journal of Neural Engineering*, *16*(2), 026007. https://doi.org/10.1088/1741-2552/aaf3f6

Farhangi, F. (2022). Investigating the role of data preprocessing, hyperparameters tuning, and type of machine learning algorithm in the improvement of drowsy EEG signal modeling. *Intelligent Systems with Applications*, *15*, 200100. https://doi.org/10.1016/j.iswa.2022.200100

Feng, X., Gao, X., & Luo, L. (2021). A ResNet50-Based Method for Classifying Surface Defects in Hot-Rolled Strip Steel. *Mathematics*, *9*(19), 2359. https://doi.org/10.3390/math9192359

Ferraro, J. C., & Mouloua, M. (2021). Effects of automation reliability on error detection and attention to auditory stimuli in a multi-tasking environment. *Applied Ergonomics*, *91*. https://doi.org/10.1016/j.apergo.2020.103303

Figueiredo, M. G., Alvarez, D., & Adams, R. N. (2018). O acidente da plataforma de petróleo P-36 revisitado 15 anos depois: Da gestão de situações incidentais e acidentais aos fatores organizacionais. *Cadernos de Saude Publica*, *34*(4). https://doi.org/10.1590/0102-311x00034617

Folkard, S. (2003). Shift work, safety and productivity. *Occupational Medicine*, *53*(2), 95–101. https://doi.org/10.1093/occmed/kqg047

Forsman, P., Pyykkö, I., Toppila, E., & Hæggström, E. (2014). Feasibility of force platform based roadside drowsiness screening - A pilot study. *Accident Analysis and Prevention*. https://doi.org/10.1016/j.aap.2013.09.015

Gander, P., Hartley, L., Powell, D., Cabon, P., Hitchcock, E., Mills, A., & Popkin, S. (2011). Fatigue risk management: Organizational factors at the regulatory and industry/company level. *Accident Analysis and Prevention*, *43*, 573–590. https://doi.org/10.1016/j.aap.2009.11.007

Garcés Correa, A., Orosco, L., & Laciar, E. (2014). Automatic detection of drowsiness in EEG records based on multimodal analysis. *Medical Engineering & Physics*, *36*(2), 244–249. https://doi.org/10.1016/j.medengphy.2013.07.011

García, D. P., Cruz-Benito, J., & García-Peñalvo, F. J. (2022). *Systematic Literature Review: Quantum Machine Learning and its applications*. http://arxiv.org/abs/2201.04093

Georgescu, I. M., Ashhab, S., & Nori, F. (2014). Quantum simulation. *Reviews of Modern Physics*, *86*(1), 153–185. https://doi.org/10.1103/RevModPhys.86.153

Greenblatt, A. S., Beniczky, S., & Nascimento, F. A. (2023). Pitfalls in scalp EEG: Current obstacles and future directions. *Epilepsy & Behavior*, *149*, 109500. https://doi.org/10.1016/j.yebeh.2023.109500

Gruenhagen, J. H., Parker, R., & Cox, S. (2021). Technology diffusion and firm agency from a technological innovation systems perspective: A case study of fatigue monitoring in the mining industry. *Journal of Engineering and Technology Management*, *62*, 101655. https://doi.org/10.1016/j.jengtecman.2021.101655

Gu, J., & Guo, F. (2022). How fatigue affects the safety behaviour intentions of construction workers an empirical study in Hunan, China. *Safety Science*, *149*, 105684. https://doi.org/10.1016/j.ssci.2022.105684

Guarda, L., Tapia, J. E., Droguett, E. L., & Ramos, M. (2022). A novel Capsule Neural Network based model for drowsiness detection using electroencephalography signals. In *Expert Systems with Applications* (Vol. 201). Elsevier Ltd. https://doi.org/10.1016/j.eswa.2022.116977

Hajinoroozi, M., Mao, Z., Jung, T. P., Lin, C. T., & Huang, Y. (2016). EEG-based prediction of driver's cognitive performance by deep convolutional neural network. *Signal Processing: Image Communication*, *47*, 549–555. https://doi.org/10.1016/j.image.2016.05.018

Hassan, A. R., & Haque, M. A. (2016). Computer-aided obstructive sleep apnea screening from single-lead electrocardiogram using statistical and spectral features and bootstrap aggregating. *Biocybernetics and Biomedical Engineering*, *36*(1), 256–266. https://doi.org/10.1016/j.bbe.2015.11.003

He, K., Zhang, X., Ren, S., & Sun, J. (2015). *Deep Residual Learning for Image Recognition*. http://arxiv.org/abs/1512.03385

Higuchi, T. (1988). Approach to an irregular time series on the basis of the fractal theory. *Physica D: Nonlinear Phenomena*, *31*(2), 277–283. https://doi.org/10.1016/0167-2789(88)90081-4

Hjorth, B. (1970). EEG analysis based on time domain properties. *Electroencephalography and Clinical Neurophysiology*, *29*(3), 306–310. https://doi.org/10.1016/0013-4694(70)90143-4

Hong, S., & Baek, H. J. (2021). Drowsiness detection based on intelligent systems with nonlinear features for optimal placement of encephalogram electrodes on the cerebral area. *Sensors (Switzerland)*. https://doi.org/10.3390/s21041255

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *International Conference on Machine Learning*, 448–456.

Iridiastadi, H. (2021). Fatigue in the Indonesian rail industry: A study examining passenger train drivers. *Applied Ergonomics*, *92*, 103332. https://doi.org/10.1016/j.apergo.2020.103332

Jacobé de Naurois, C., Bourdin, C., Bougard, C., & Vercher, J.-L. (2018). Adapting artificial neural networks to a specific driver enhances detection and prediction of drowsiness. *Accident Analysis & Prevention*, *121*, 118–128. https://doi.org/10.1016/j.aap.2018.08.017

Jang, I., Kim, Y., & Park, J. (2021). Investigating the Effect of Task Complexity on the Occurrence of Human Errors observed in a Nuclear Power Plant Full-Scope Simulator. *Reliability Engineering and System Safety*, *214*. https://doi.org/10.1016/j.ress.2021.107704

Ji, Q., Zhu, Z., & Lan, P. (2004). Real-Time Nonintrusive Monitoring and Prediction of Driver Fatigue. *IEEE Transactions on Vehicular Technology*, *53*(4), 1052–1068. https://doi.org/10.1109/TVT.2004.830974

Kaida, K., Takahashi, M., Åkerstedt, T., Nakata, A., Otsuka, Y., Haratani, T., & Fukasawa, K. (2006). Validation of the Karolinska sleepiness scale against performance and EEG variables. *Clinical Neurophysiology*. https://doi.org/10.1016/j.clinph.2006.03.011

Karayiannis, N. B., Mukherjee, A., Glover, J. R., Ktonas, P. Y., Frost, J. D. Jr., Hrachovy, R. A., & Mizrahi, E. M. (n.d.). Quantifying and visualizing uncertainty in EEG data of neonatal seizures. *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 423–426. https://doi.org/10.1109/IEMBS.2004.1403184

Kartsch, V. J., Benatti, S., Schiavone, P. D., Rossi, D., & Benini, L. (2018). A sensor fusion approach for drowsiness detection in wearable ultra-low-power systems. *Information Fusion*, *43*, 66–76. https://doi.org/10.1016/j.inffus.2017.11.005

Katsis, C. D., Katertsidis, N., Ganiatsas, G., & Fotiadis, D. I. (2008). Toward Emotion Recognition in Car-Racing Drivers: A Biosignal Processing Approach. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, *38*(3), 502–512. https://doi.org/10.1109/TSMCA.2008.918624

Kesić, S., & Spasić, S. Z. (2016). Application of Higuchi's fractal dimension from basic to clinical neurophysiology: A review. *Computer Methods and Programs in Biomedicine*, *133*, 55–70. https://doi.org/10.1016/j.cmpb.2016.05.014

Kim, K., Hong, H., Nam, G., & Park, K. (2017). A Study of Deep CNN-Based Classification of Open and Closed Eyes Using a Visible Light Camera Sensor. *Sensors*, *17*(7), 1534. https://doi.org/10.3390/s17071534

King, D. E. (2009). Dlib-ml: A machine learning toolkit. *The Journal of Machine Learning Research*, *10*, 1755–1758.

Kingma, D. P., & Ba, J. (2014). *Adam: A Method for Stochastic Optimization*. http://arxiv.org/abs/1412.6980

Kolus, A. (2024). A Systematic Review on Driver Drowsiness Detection Using Eye Activity Measures. *IEEE Access*, *12*, 97969–97993. https://doi.org/10.1109/ACCESS.2024.3424654

Lavie, P., Kremerman, S., & Wiel, M. (1982). Sleep disorders and safety at work in industry workers. *Accident Analysis and Prevention*, *14*(4), 311–314. https://doi.org/10.1016/0001-4575(82)90043-4

Lee, B.-G., & Chung, W.-Y. (2012). A Smartphone-Based Driver Safety Monitoring System Using Data Fusion. *Sensors*, *12*(12), 17536–17552. https://doi.org/10.3390/s121217536

Lee, C., & An, J. (2023). LSTM-CNN model of drowsiness detection from multiple consciousness states acquired by EEG. *Expert Systems with Applications*, *213*. https://doi.org/10.1016/j.eswa.2022.119032

Li, G., Lee, B. L., & Chung, W. Y. (2015). Smartwatch-Based Wearable EEG System for Driver Drowsiness Detection. *IEEE Sensors Journal*. https://doi.org/10.1109/JSEN.2015.2473679

Li, J., Li, H., Umer, W., Wang, H., Xing, X., Zhao, S., & Hou, J. (2020). Identification and classification of construction equipment operators' mental fatigue using wearable eye-tracking technology. *Automation in Construction*, *109*, 103000. https://doi.org/10.1016/j.autcon.2019.103000

Li, W., He, Q., Fan, X., & Fei, Z. (2012). Evaluation of driver fatigue on two channels of EEG data. *Neuroscience Letters*, *506*(2), 235–239. https://doi.org/10.1016/j.neulet.2011.11.014

Li, Y., Zhou, R.-G., Xu, R., Luo, J., & Jiang, S.-X. (2022). A Quantum Mechanics-Based Framework for EEG Signal Feature Extraction and Classification. *IEEE Transactions on Emerging Topics in Computing*, *10*(1), 211–222. https://doi.org/10.1109/TETC.2020.3000734

Lin, C. T., Wu, R. C., Liang, S. F., Chao, W. H., Chen, Y. J., & Jung, T. P. (2005). EEG-based drowsiness estimation for safety driving using independent component analysis. *IEEE Transactions on Circuits and Systems I: Regular Papers*. https://doi.org/10.1109/TCSI.2005.857555

Lins, I. D., Araújo, L. M. M., Maior, C. B. S., Ramos, P. M. da S., Moura, M. J. das C., Ferreira-Martins, A. J., Chaves, R., & Canabarro, A. (2024). Quantum machine learning for drowsiness detection with EEG signals. *Process Safety and Environmental Protection*, *186*, 1197–1213. https://doi.org/10.1016/j.psep.2024.04.032

Liu, Z., Yang, C., Huang, J., Liu, S., Zhuo, Y., & Lu, X. (2021). Deep learning framework based on integration of S-Mask R-CNN and Inception-v3 for ultrasound image-aided diagnosis of prostate cancer. *Future Generation Computer Systems*, *114*, 358–367. https://doi.org/10.1016/j.future.2020.08.015

Lotte, F., Bougrain, L., Cichocki, A., Clerc, M., Congedo, M., Rakotomamonjy, A., & Yger, F. (2018). A review of classification algorithms for EEG-based brain-computer interfaces: A 10 year update. In *Journal of Neural Engineering*. https://doi.org/10.1088/1741-2552/aab2f2

Lybeck, N., Marble, S., & Morton, B. (2007). Validating Prognostic Algorithms: A Case Study Using Comprehensive Bearing Fault Data. *2007 IEEE Aerospace Conference*, 1–9. https://doi.org/10.1109/AERO.2007.352842

Maior, C. B. S., Araújo, L. M. M., Lins, I. D., Moura, M. D. C., & Droguett, E. L. (2023). Prognostics and Health Management of Rotating Machinery via Quantum Machine Learning. *IEEE Access*, *11*, 25132–25151. https://doi.org/10.1109/ACCESS.2023.3255417

Maior, C. B. S., Moura, M. das C., & Lins, I. D. (2019). Particle swarm-optimized support vector machines and pre-processing techniques for remaining useful life estimation of bearings. *Eksploatacja i Niezawodnosc - Maintenance and Reliability*, *21*(4), 610–619. https://doi.org/10.17531/ein.2019.4.10

Maior, C. B. S., Moura, M. J. das C., Santana, J. M. M., & Lins, I. D. (2020). Real-time classification for autonomous drowsiness detection using eye aspect ratio. *Expert Systems with Applications*, *158*. https://doi.org/10.1016/j.eswa.2020.113505

Maior, C., das Chagas Moura, M., Didier Lins, I., Lopez Droguett, E., & Henrique Lima Diniz, H. (2016). Remaining Useful Life Estimation by Empirical Mode Decomposition and Support Vector Machine. *IEEE Latin America Transactions*, *14*(11), 4603–4610. https://doi.org/10.1109/TLA.2016.7795836

Massoz, Q., Langohr, T., François, C., & Verly, J. G. (n.d.). *The ULg Multimodality Drowsiness Database (called DROZY) and Examples of Use*. http://www.drozy.ulg.ac.be

Massoz, Q., Langohr, T., Francois, C., & Verly, J. G. (2016). The ULg multimodality drowsiness database (called DROZY) and examples of use. *2016 IEEE Winter Conference on Applications of Computer Vision, WACV 2016*. https://doi.org/10.1109/WACV.2016.7477715

Mehmood, I., Li, H., Umer, W., Arsalan, A., Anwer, S., Mirza, M. A., Ma, J., & Antwi-Afari, M. F. (2023). Multimodal integration for data-driven classification of mental fatigue during construction equipment operations: Incorporating electroencephalography, electrodermal activity, and video signals. *Developments in the Built Environment*, *15*, 100198. https://doi.org/10.1016/j.dibe.2023.100198

Meng, T., Jing, X., Yan, Z., & Pedrycz, W. (2020). A survey on machine learning for data fusion. *Information Fusion*, *57*, 115–129. https://doi.org/10.1016/j.inffus.2019.12.001

MIURA. (2023). *STEAM BOILERS*. Https://Miuraboiler.Com/Steam-Boilers-101/.

Mokarram, M. J., Rashiditabar, R., Gitizadeh, M., & Aghaei, J. (2023). Net-load forecasting of renewable energy systems using multi-input LSTM fuzzy and discrete wavelet transform. *Energy*, *275*. https://doi.org/10.1016/j.energy.2023.127425

Morag, I., Chemweno, P., Pintelon, L., & Sheikhalishahi, M. (2018). Identifying the causes of human error in maintenance work in developing countries. *International Journal of Industrial Ergonomics*, *68*, 222–230. https://doi.org/10.1016/j.ergon.2018.08.014

M.R.Taha, S., & K. Nawar, A. (2014). A New Quantum Radial Wavelet Neural Network Model Applied to Analysis and Classification of EEG Signals. *International Journal of Computer Applications*, *85*(7), 7–11. https://doi.org/10.5120/14851-3216

Murthy, K. S. R., Siddineni, B., Kompella, V. K., Aashritha, K., Sri Sai, B. H., & Manikandan, V. M. (2022). An Efficient Drowsiness Detection Scheme using Video Analysis. *International Journal of Computing and Digital Systems*, *11*(1), 573–581. https://doi.org/10.12785/ijcds/110146

National Transportation Safety Board. (2023). *Contact of Offshore Supply Vessel Elliot Cheramie with Oil and Gas Production Platform EI-259A*. https://www.ntsb.gov/investigations/AccidentReports/Reports/MIR2301.pdf

Nayak, G. K., & Kim, E. (2021). Development of a fully automated RULA assessment system based on computer vision. *International Journal of Industrial Ergonomics*, *86*. https://doi.org/10.1016/j.ergon.2021.103218

Ngai, W. K., Xie, H., Zou, D., & Chou, K.-L. (2022). Emotion recognition based on convolutional neural networks and heterogeneous bio-signal data sources. *Information Fusion*, *77*, 107–117. https://doi.org/10.1016/j.inffus.2021.07.007

Nijaguna, G. S., Babu, J. A., Parameshachari, B. D., de Prado, R. P., & Frnda, J. (2023). Quantum Fruit Fly algorithm and ResNet50-VGG16 for medical diagnosis. *Applied Soft Computing*, *136*, 110055. https://doi.org/10.1016/j.asoc.2023.110055

Ogino, M., & Mitsukura, Y. (2018). Portable Drowsiness Detection through Use of a Prefrontal Single-Channel Electroencephalogram. *Sensors*, *18*(12), 4477. https://doi.org/10.3390/s18124477

Okello, E. J., Abadi, A. M., & Abadi, S. A. (2016). Effects of green and black tea consumption on brain wave activities in healthy volunteers as measured by a simplified Electroencephalogram (EEG): A feasibility study. *Nutritional Neuroscience*, *19*(5), 196–205. https://doi.org/10.1179/1476830515Y.0000000008

Omidi, L., Zakerian, S. A., Nasl Saraji, J., Hadavandi, E., & Yekaninejad, M. S. (2018). Safety performance assessment among control room operators based on feature extraction and genetic fuzzy system in the process industry. *Process Safety and Environmental Protection*, *116*, 590–602. https://doi.org/10.1016/j.psep.2018.03.014

Pandey, N. N., & Muppalaneni, N. B. (2023). Dumodds: Dual modeling approach for drowsiness detection based on spatial and spatio-temporal features. *Engineering Applications of Artificial Intelligence*, *119*. https://doi.org/10.1016/j.engappai.2022.105759

Paneru, S., & Jeelani, I. (2021). Computer vision applications in construction: Current state, opportunities & challenges. *Automation in Construction*, *132*, 103940. https://doi.org/10.1016/j.autcon.2021.103940

Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors*, *52*(3), 381–410. https://doi.org/10.1177/0018720810376055

Parkes, K. R. (2012). Shift schedules on North Sea oil/gas installations: A systematic review of their impact on performance, safety and health. *Safety Science*, *50*(7), 1636–1651. https://doi.org/10.1016/j.ssci.2012.01.010

Pedregosa FABIANPEDREGOSA, F., Michel, V., Grisel OLIVIERGRISEL, O., Blondel, M., Prettenhofer, P., Weiss, R., Vanderplas, J., Cournapeau, D., Pedregosa, F., Varoquaux, G., Gramfort, A., Thirion, B., Grisel, O., Dubourg, V., Passos, A., Brucher, M., Perrot andÉdouardand, M., Duchesnay, andÉdouard, & Duchesnay EDOUARDDUCHESNAY, Fré. (2011). Scikit-learn: Machine Learning in Python Gaël Varoquaux Bertrand Thirion Vincent Dubourg Alexandre Passos PEDREGOSA, VAROQUAUX, GRAMFORT ET AL. Matthieu Perrot. In *Journal of Machine Learning Research* (Vol. 12). http://scikit-learn.sourceforge.net.

Phan, A.-C., Trieu, T.-N., & Phan, T.-C. (2023). Driver drowsiness detection and smart alerting using deep learning and IoT. *Internet of Things*, *22*, 100705. https://doi.org/10.1016/j.iot.2023.100705

Picchioni, D., Fukunaga, M., Carr, W. S., Braun, A. R., Balkin, T. J., Duyn, J. H., & Horovitz, S. G. (2008). fMRI differences between early and late stage-1 sleep. *Neuroscience Letters*, *441*(1), 81–85. https://doi.org/10.1016/j.neulet.2008.06.010

Picot, A., Charbonnier, S., & Caplier, A. (2008). On-line automatic detection of driver drowsiness using a single electroencephalographic channel. *Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS'08 - "Personalized Healthcare through Technology."* https://doi.org/10.1109/iembs.2008.4650053

Pilanawithana, N. M., Feng, Y., London, K., & Zhang, P. (2023). Framework for measuring resilience of safety management systems in Australian building repair and maintenance companies. *Journal of Safety Research*, *85*, 405–418. https://doi.org/10.1016/j.jsr.2023.04.008

Prakash, K. B. . (2021). *Cognitive Engineering for Next Generation Computing* (K. B. Prakash, G. R. Kanagachidambaresan, V. Srikanth, & E. Vamsidhar, Eds.). Wiley. https://doi.org/10.1002/9781119711308

Qi, M.-S., Yang, W.-J., Xie, P., Liu, Z.-J., Zhang, Y.-Y., & Cheng, S.-C. (2018). Driver fatigue Assessment Based on the Feature Fusion and Transfer Learning of EEG and EMG. *2018 Chinese Automation Congress (CAC)*, 1314–1317. https://doi.org/10.1109/CAC.2018.8623087

Rahman, A., Hriday, M. B. H., & Khan, R. (2022). Computer vision-based approach to detect fatigue driving and face mask for edge computing device. *Heliyon*, *8*(10), e11204. https://doi.org/10.1016/j.heliyon.2022.e11204

Ramos, M. A., & Mosleh, A. (2021). Human Role in Failure of Autonomous Systems: A Human Reliability Perspective. *Proceedings - Annual Reliability and Maintainability Symposium*, *2021-May*. https://doi.org/10.1109/RAMS48097.2021.9605790

Ramos, P. M. S., Maior, C. B. S., Moura, M. C., & Lins, I. D. (2022). Automatic drowsiness detection for safety-critical operations using ensemble models and EEG signals. *Process Safety and Environmental Protection*, *164*, 566–581. https://doi.org/10.1016/j.psep.2022.06.039

Ramos, P. M. S., Ramos, M., Maiora, C. B. S., Moura, M. C., Lins, I. D., & Bispo, H. (2023). Experimental Set-Up for Evaluating Operator Performance through Operations Control Room Simulation in the Oil and Gas Industry. *Proceeding of the 33rd European Safety and Reliability Conference*, 764–771. https://doi.org/10.3850/978-981-18-8071-1_P239-cd

Rasmussen, S. E., & Zinner, N. T. (2022). Parameterized Two-Qubit Gates for Enhanced Variational Quantum Eigensolver. *Annalen Der Physik*, *534*(12). https://doi.org/10.1002/andp.202200338

Rieffel, E. G., & Polak, W. H. (2011). *Quantum Computing A Gentle Introduction* (MIT Press).

Sadeghi, R., & Goerlandt, F. (2023). Validation of system safety hazard analysis in safety-critical industries: An interview study with industry practitioners. *Safety Science*, *161*. https://doi.org/10.1016/j.ssci.2023.106084

Sadeghniiat-Haghighi, K., Zahabi, A., Najafi, A., Rahimi-Golkhandan, A., & Aminian, O. (2020). Evaluating the quality and duration of sleep using

actigraphy in petroleum industry shift workers. *Sleep Health*, *6*(3), 407–410. https://doi.org/10.1016/j.sleh.2020.04.010

Sandberg, D., Åkerstedt, T., Anund, A., Kecklund, G., & Wahde, M. (2011). Detecting driver sleepiness using optimized nonlinear combinations of sleepiness indicators. *IEEE Transactions on Intelligent Transportation Systems*. https://doi.org/10.1109/TITS.2010.2077281

Scherer, W. (2019). *Mathematics of Quantum Computing*. Springer International Publishing. https://doi.org/10.1007/978-3-030-12358-1

Selvik, J. T., & Bellamy, L. J. (2020). Addressing human error when collecting failure cause information in the oil and gas industry: A review of ISO 14224:2016. *Reliability Engineering & System Safety*, *194*, 106418. https://doi.org/10.1016/j.ress.2019.03.025

Shahbakhti, M., Beiramvand, M., Nasiri, E., Far, S. M., Chen, W., Solé-Casals, J., Wierzchon, M., Broniec-Wójcik, A., Augustyniak, P., & Marozas, V. (2023). Fusion of EEG and Eye Blink Analysis for Detection of Driver Fatigue. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *31*, 2037–2046. https://doi.org/10.1109/TNSRE.2023.3267114

Sharma, S., Khare, S. K., Bajaj, V., & Ansari, I. A. (2021). Improving the separability of drowsiness and alert EEG signals using analytic form of wavelet transform. *Applied Acoustics*, *181*, 108164. https://doi.org/10.1016/j.apacoust.2021.108164

Shepovalnikov, A. N., Tsitseroshin, M. N., Galperina, E. I., Rozhkov, V. P., Kruchinina, O. V., Zaitseva, L. G., & Panasevich, E. A. (2012). Characteristics of integrative brain activity during various stages of sleep and in transitional states. *Human Physiology*, *38*(3), 227–237. https://doi.org/10.1134/S0362119712030127

Shiwu, L., Linhong, W., Zhifa, Y., Bingkui, J., Feiyan, Q., & Zhongkai, Y. (2011). An active driver fatigue identification technique using multiple physiological features. *2011 International Conference on Mechatronic Science, Electric Engineering and Computer (MEC)*, 733–737. https://doi.org/10.1109/MEC.2011.6025569

Shor, P. W. (1997). Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer. *SIAM Journal on Computing*, *26*(5), 1484–1509. https://doi.org/10.1137/S0097539795293172

Soares, P. R. F. T. (2019). *REALIDADE VIRTUAL COMO FERRAMENTA DE TREINAMENTO PARA BRIGADA DE EMERGÊNCIA DE UMA REFINARIA*. https://repositorio.ufpe.br/handle/123456789/33772

Sors, A., Bonnet, S., Mirek, S., Vercueil, L., & Payen, J. F. (2018). A convolutional neural network for sleep stage scoring from raw single-channel EEG. *Biomedical Signal Processing and Control*, *42*, 107–114. https://doi.org/10.1016/j.bspc.2017.12.001

Soukupova, T., & Cech, J. (2016). Eye blink detection using facial landmarks. *21st Computer Vision Winter Workshop, Rimske Toplice, Slovenia*, 2.

Sriraam, N., Padma Shri, T. K., & Maheshwari, U. (2016). Recognition of wake-sleep stage 1 multichannel eeg patterns using spectral entropy features for drowsiness detection. *Australasian Physical and Engineering Sciences in Medicine*, *39*(3), 797–806. https://doi.org/10.1007/s13246-016-0472-8

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, *15*(1), 1929–1958.

Tajani, A. H. N., Bamshad, A., & Ghaffarzadeh, N. (2023). A novel differential protection scheme for AC microgrids based on discrete wavelet transform. *Electric Power Systems Research*, *220*, 109292. https://doi.org/10.1016/j.epsr.2023.109292

Technical Committee ISO/TC 67. Materials petrochemical and natural gas industries, equipment and offshore structures for petroleum. (2016). *Petroleum, Petrochemical and Natural Gas Industries: Collection and Exchange of Reliability and Maintenance Data for Equipment*. ISO.

Theophilus, S. C., Esenowo, V. N., Arewa, A. O., Ifelebuegu, A. O., Nnadi, E. O., & Mbanaso, F. U. (2017). Human factors analysis and classification system for the oil and gas industry (HFACS-OGI). *Reliability Engineering and System Safety*, *167*, 168–176. https://doi.org/10.1016/j.ress.2017.05.036

Tian, T., Wang, L., Luo, M., Sun, Y., & Liu, X. (2022). ResNet-50 based technique for EEG image characterization due to varying environmental stimuli. *Computer Methods and Programs in Biomedicine*, *225*, 107092. https://doi.org/10.1016/j.cmpb.2022.107092

Tinga, A. M., van Zeumeren, I. M., Christoph, M., van Grondelle, E., Cleij, D., Aldea, A., & van Nes, N. (2023). Development and evaluation of a human machine interface to support mode awareness in different automated driving modes. *Transportation Research Part F: Traffic Psychology and Behaviour*, *92*, 238–254. https://doi.org/10.1016/j.trf.2022.10.023

Uflaz, E., Akyuz, E., Arslan, O., & Celik, E. (2022). A quantitative effectiveness analysis to improve the safety management system (SMS) implementation on-board ship. *Safety Science*, *156*, 105913. https://doi.org/10.1016/j.ssci.2022.105913

van Leeuwen, W. M. A., Pekcan, C., Barnett, M., & Kecklund, G. (2021). Mathematical modelling of sleep and sleepiness under various watch keeping schedules in the maritime industry. *Marine Policy*, *130*, 104277. https://doi.org/10.1016/j.marpol.2020.104277

Vanhollebeke, G., De Smet, S., De Raedt, R., Baeken, C., van Mierlo, P., & Vanderhasselt, M.-A. (2022). The neural correlates of psychosocial stress: A systematic review and meta-analysis of spectral analysis EEG studies. *Neurobiology of Stress*, *18*, 100452. https://doi.org/10.1016/j.ynstr.2022.100452

Vicente, J., Laguna, P., Bartra, A., & Bailón, R. (2016). Drowsiness detection using heart rate variability. *Medical and Biological Engineering and Computing*, *54*, 927–937. https://doi.org/10.1007/s11517-015-1448-7

Waage, S., Harris, A., Pallesen, S., Saksvik, I. B., Moen, B. E., & Bjorvatn, B. (2012). Subjective and objective sleepiness among oil rig workers during three different shift schedules. *Sleep Medicine*, *13*(1), 64–72. https://doi.org/10.1016/j.sleep.2011.04.009

Wang, P., Min, J., & Hu, J. (2018). Ensemble classifier for driver's fatigue detection based on a single EEG channel. *IET Intelligent Transport Systems*, *12*(10), 1322–1328. https://doi.org/10.1049/iet-its.2018.5290

Wang, X., Wang, J., Zhang, K., Lin, F., & Chang, Q. (2020). Convergence and objective functions of noise-injected multilayer perceptrons with hidden multipliers. *Neurocomputing*. https://doi.org/10.1016/j.neucom.2020.03.119

Wang, Y., & Liu, H. (2022). Quantum Computing in a Statistical Context. *Annual Review of Statistics and Its Application*, *9*(1), 479–504. https://doi.org/10.1146/annurev-statistics-042720-024040

Wei, L., Mukhopadhyay, S. C., Jidin, R., & Chen, C.-P. (2013). Multi-source information fusion for drowsy driving detection based on wireless sensor networks. *2013 Seventh International Conference on Sensing Technology (ICST)*, 850–857. https://doi.org/10.1109/ICSensT.2013.6727771

Widodo, A., & Yang, B. S. (2007). Support vector machine in machine condition monitoring and fault diagnosis. In *Mechanical Systems and Signal Processing*. https://doi.org/10.1016/j.ymssp.2006.12.007

Xu, W., Fu, Y.-L., & Zhu, D. (2023). ResNet and its application to medical image processing: Research progress and challenges. *Computer Methods and Programs in Biomedicine*, *240*, 107660. https://doi.org/10.1016/j.cmpb.2023.107660

Yang, E., & Yi, O. (2024). Enhancing Road Safety: Deep Learning-Based Intelligent Driver Drowsiness Detection for Advanced Driver-Assistance Systems. *Electronics*, *13*(4), 708. https://doi.org/10.3390/electronics13040708

Yıldırım, Ö., Baloglu, U. B., & Acharya, U. R. (2020). A deep convolutional neural network model for automated identification of abnormal EEG signals. *Neural Computing and Applications*, *32*(20), 15857–15868. https://doi.org/10.1007/s00521-018-3889-z

You, F., Li, X., Gong, Y., Wang, H., & Li, H. (2019). A Real-time Driving Drowsiness Detection Algorithm With Individual Differences Consideration. *IEEE Access*, *7*, 179396–179408. https://doi.org/10.1109/ACCESS.2019.2958667

Yu, S., Li, P., Lin, H., Rohani, E., Choi, G., Shao, B., & Wang, Q. (2013). Support Vector Machine Based Detection of Drowsiness Using Minimum EEG Features. *2013 International Conference on Social Computing*, 827–835. https://doi.org/10.1109/SocialCom.2013.124

Zarei, E., Khan, F., & Abbassi, R. (2023). How to account artificial intelligence in human factor analysis of complex systems? In *Process Safety and Environmental Protection* (Vol. 171, pp. 736–750). Institution of Chemical Engineers. https://doi.org/10.1016/j.psep.2023.01.067

Zhao, C., Zhao, M., Yang, Y., Gao, J., Rao, N., & Lin, P. (2017). The Reorganization of Human Brain Networks Modulated by Driving Mental Fatigue. *IEEE Journal of Biomedical and Health Informatics*, *21*(3), 743–755. https://doi.org/10.1109/JBHI.2016.2544061

Zhao, L., Li, M., He, Z., Ye, S., Qin, H., Zhu, X., & Dai, Z. (2022). Data-driven learning fatigue detection system: A multimodal fusion approach of ECG (electrocardiogram) and video signals. *Measurement*, *201*, 111648. https://doi.org/10.1016/j.measurement.2022.111648

Zhou, A., Wang, K., & Zhang, H. (2017). Human factor risk control for oil and gas drilling industry. *Journal of Petroleum Science and Engineering*, *159*, 581–587. https://doi.org/10.1016/j.petrol.2017.09.034

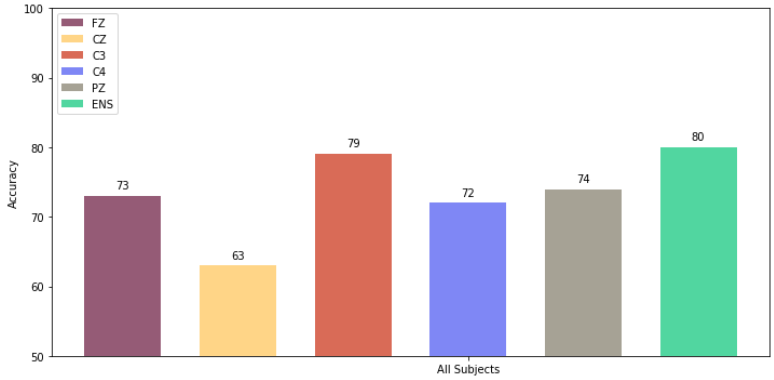
Zhuang, Q., Gan, S., & Zhang, L. (2022). Human-computer interaction based health diagnostics using ResNet34 for tongue image classification. *Computer Methods and Programs in Biomedicine*, *226*, 107096. https://doi.org/10.1016/j.cmpb.2022.107096

APPENDIX A – Case 1: All subjects

- Raw data

For the raw data, results from bagging with five channels present good accuracy (80%), as seen in Figure 49. The ensemble models could take advantage of the good predictions, surpassing all other models for the bagging approach.

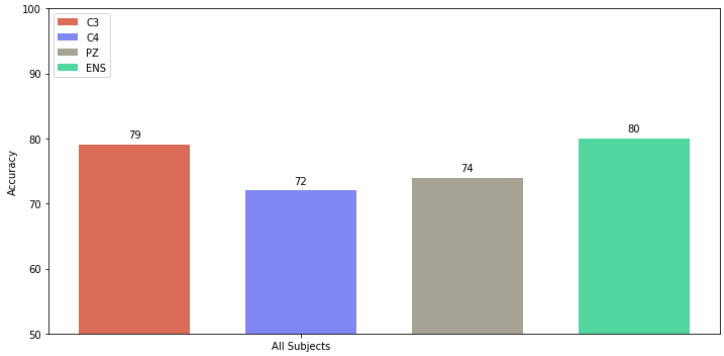
Figure 49 - Bagging-based model for all subjects from the MLP technique with raw data and five-channels



Source: The author (2024).

Then, we also investigate the performance in the bagging-based model reducing the number of channels to three (i.e., C3, C4, Pz), which is less intrusive. Once again, majority voting is used for the prediction and the result is shown in Figure 50 indicates the best performance from the ensemble model.

Figure 50 - Bagging-based model for all subjects from the MLP technique with raw data and three-channels



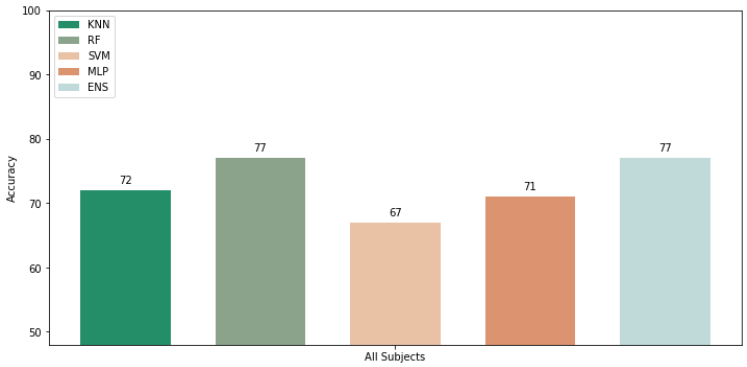
Source: The author (2024).

Compared to the result for the five-channels bagging models, the three-channels ensemble model obtained the same performance of accuracy, once again surpassing all single-channel models, however using less of them. Thus, it is noted that although it is important to have a significant number of channels to provide greater robustness to the

model's prediction, it is equally important that these channels are able to provide good predictions so as not to harm the output of the ensemble model.

For the voting model for all subjects, the results can be seen in Figure 51. Here, the result of the ensemble model was at least equal to or greater than the results of the individual classifiers. Comparing with the results obtained previously, despite having considered only one channel (C4), the result was slightly inferior to those of the bagging-based model.

Figure 51 - Voting models for all subjects from the C4 channel with raw data

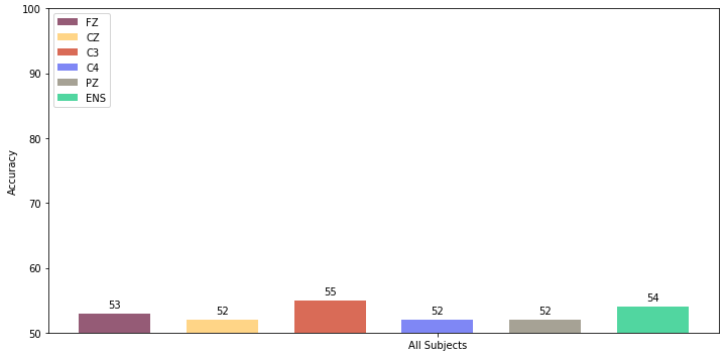


Source: The author (2024).

- EEG features

For the three features, the performance of the bagging-based approach considering the five-channels brought poor results and with all models presenting similar performances (Figure 52).

Figure 52 - Bagging-based model for all subjects from the MLP technique with features extraction and five channels

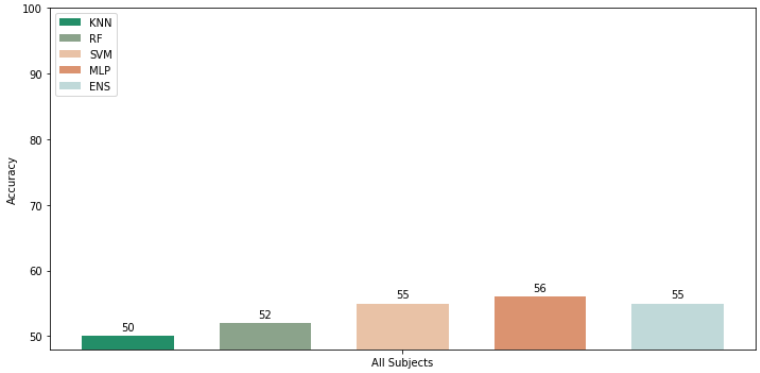


Source: The author (2024).

Considering the reduction of the bagging-based model to three-channels, the performance of the ensemble model is competitive to the best result of the classifier (C3) for this approach.

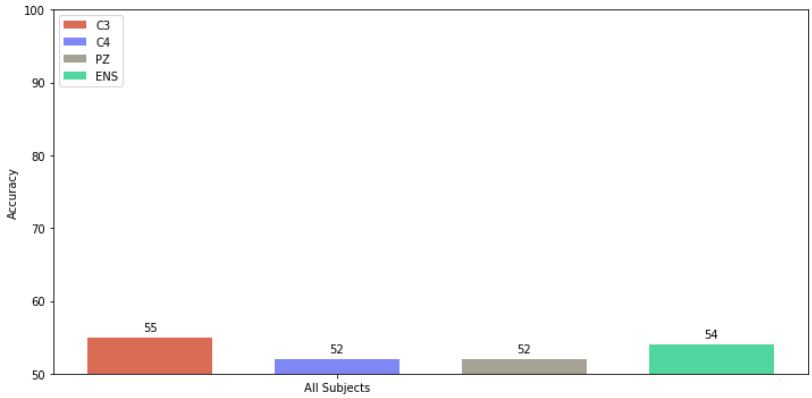
A similar pattern is seen in the voting approach, with all models presenting a close, but poor result, around 55% (Figure 53). This suggests that feature extractions considering coupled information from all subjects for training and testing, makes performances not only similar in all approaches, but also poor in detecting drowsiness in random subjects.

Figure 53 - Voting model for all subjects from the C4 channel with features extraction



Source: The author (2024).

Thus, considering Case 1, the performance of ensemble models, whether bagging-based or voting, was better for raw data. The advantages of these models can be seen when tested on different subjects in the next subsection.



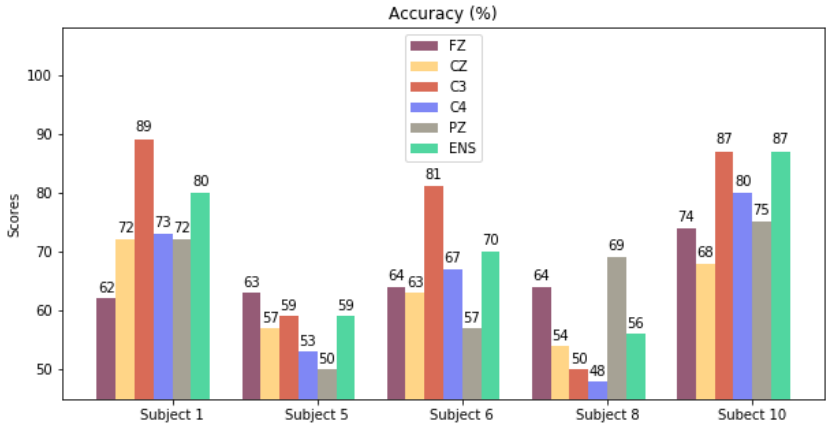
Bagging-based model for all subjects from the MLP technique with features extraction and three-channels

APPENDIX B - Case 2: Specifics subjects with a general model

- Raw data

In this case, the model is trained with data from all participants but tested in data from the five specific subjects individually. Analogously to the previous case, we analyzed raw data, applying the bagging-based model with five-channels and, then, for three-channels considering a specific ML model (i.e., MLP). Figure 54 shows the results for bagging-based model and five-channels.

Figure 54 - Bagging-based model for five specific subjects from the MLP technique with raw data and five-channels



Source: The author (2024).

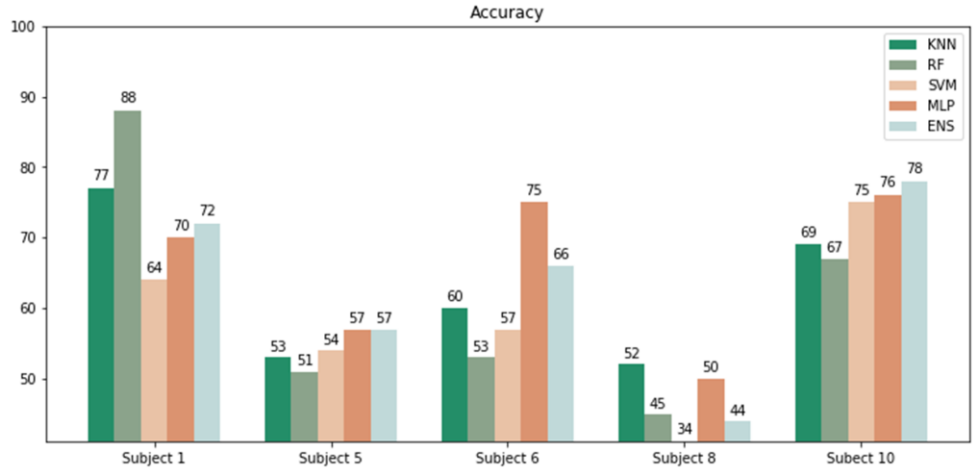
Once again, the ensemble model is useful. Comparing the best results, for a few subjects, channel C3 present brought a more significant performance (i.e., subjects 1 and 6), for other subjects (i.e., 5 and 10) the results are equivalent to the proposed ensemble model. But in subject 8 the performance of the C3 channel is better compared to the ensemble model. Moreover, the mean performance of the model based on C3 and on bagging are 73.2% and 70.4%, respectively, but the standard deviation using C3 is 15.75% compared to 11.88% for the bagging approach. These variabilities may be explained by the interpersonal characteristics of each participant, endorsing the use of the ensemble model as a positive and generalist contribution.

When considering only three-channels for the bagging-based model, the ensemble model remains with the second-highest accuracy, but no specific channel has a dominant characteristic over the proposed ensemble model, leaving it in a competitive position.

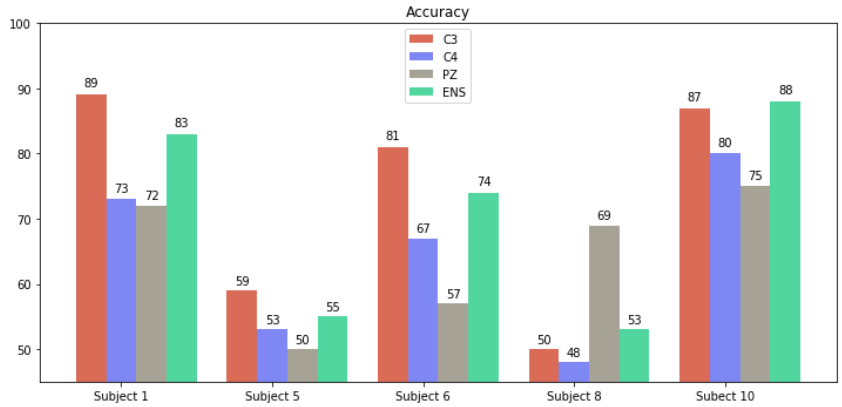
For the voting model, results are similar to those of the previous models, and can be seen in Figure 55 . The performances of the ensemble model are slightly lower than

the best accuracy of the different ML techniques tested, except for subject 8. Furthermore, no specific ML technique presents dominant results over the ensemble.

Figure 55 - Voting model for five specific subjects from the C4 channel with raw data



Source: The author (2024).

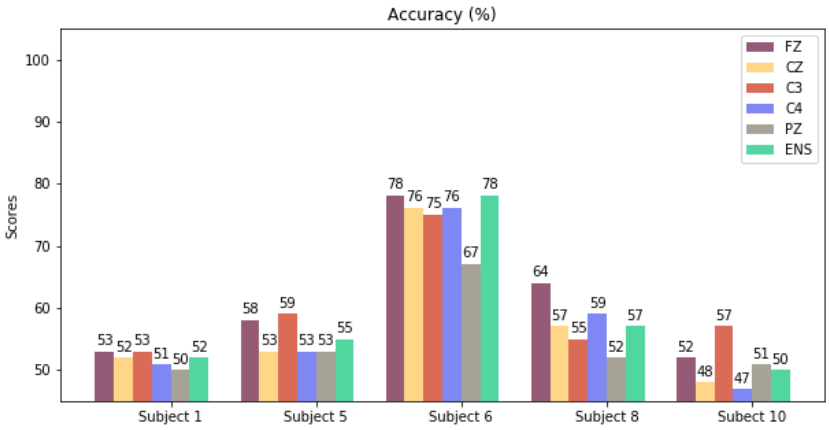


Bagging-based model for five specific subjects from the MLP technique with raw data and three-channels

- EEG features

Here, the relatively good performance of using the ensemble model is presented again. For the bagging-based model, Figure 56 bring the results with five-channels. From it, compared to the ensemble model, the single-channel C3 presents superior results in subjects 1, 5, and 10 and inferior in subjects 6 and 8. Particularly in this model, considering the EEG feature, the single-channel Fz seems to leverage the results of the ensemble, while the single-channel Pz does not seem to contribute significantly to the results in any subject.

Figure 56 - Bagging-based model for five specifics subjects from the MLP technique with features extraction and five-channels

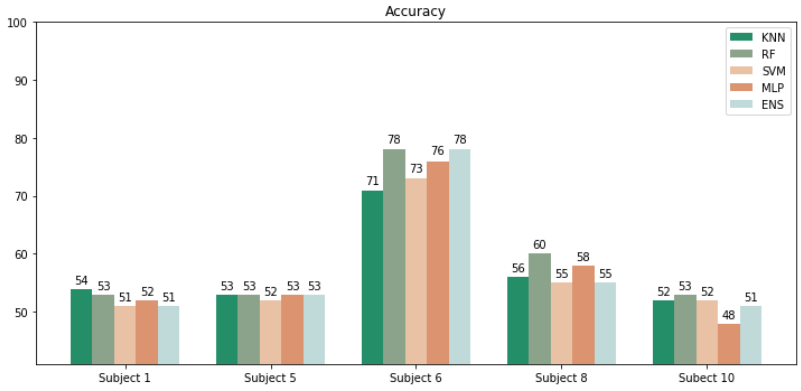


Source: The author (2024).

For the three-channel configuration, the ensemble model presents good results when comparing the accuracies of the best classifiers and, once again, no single-channels presented a result strictly dominant to the ensemble model. The performance can be seen in supplementary material.

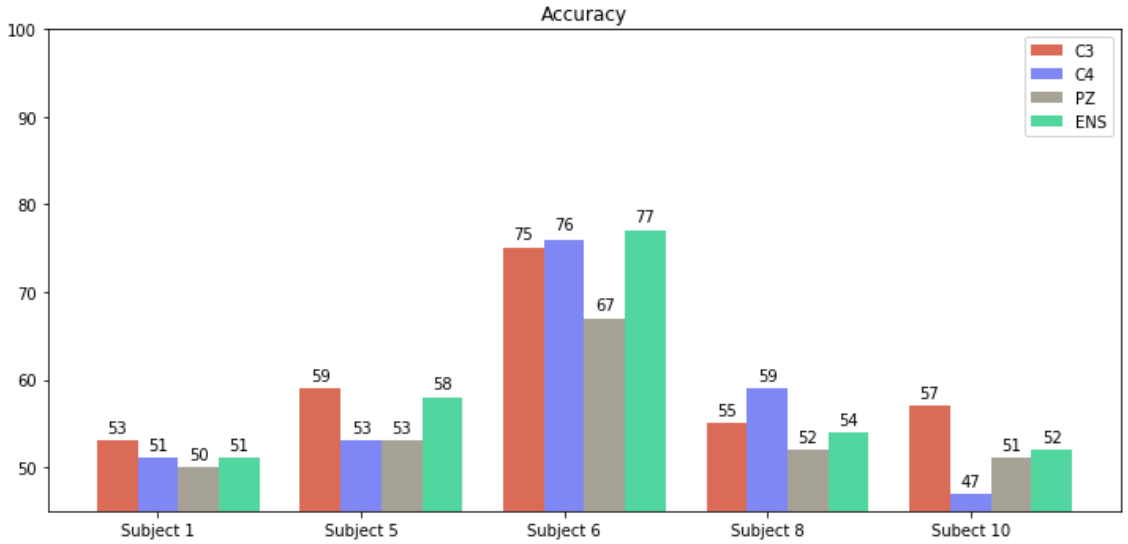
For the voting model (Figure 57), learners behave very similarly, except for subject 6, which shows a greater difference between the performance of the ensemble and the KNN technique. Furthermore, for that specific model, the ensemble model remains in a competitive position (except for subject 8) with the performance of the RF technique.

Figure 57 - Voting model for five specific subjects from the C4 channel with features extraction



Source: The author (2024).

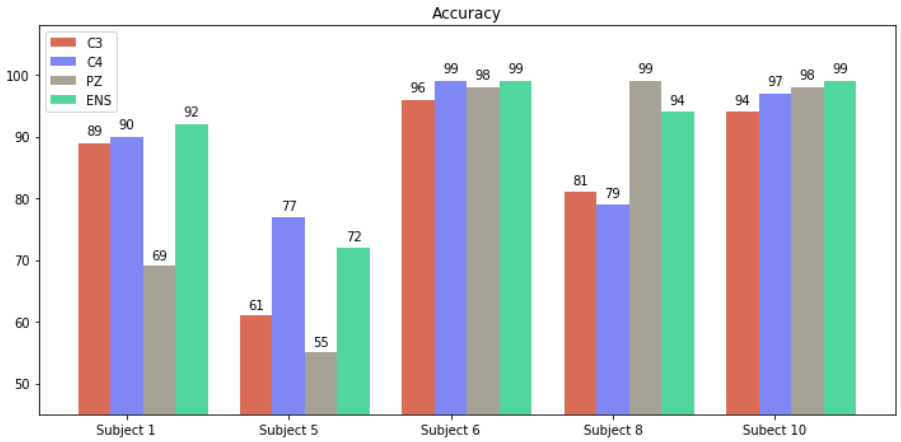
Likewise, raw data results continue to show better achievement when contrasted with features extraction. The next subsection aims to present the behavior of these same subjects tested, from now on, with classifiers trained with data from the subjects themselves.



Bagging based model for five specific subjects from the MLP technique with features extraction and three channels

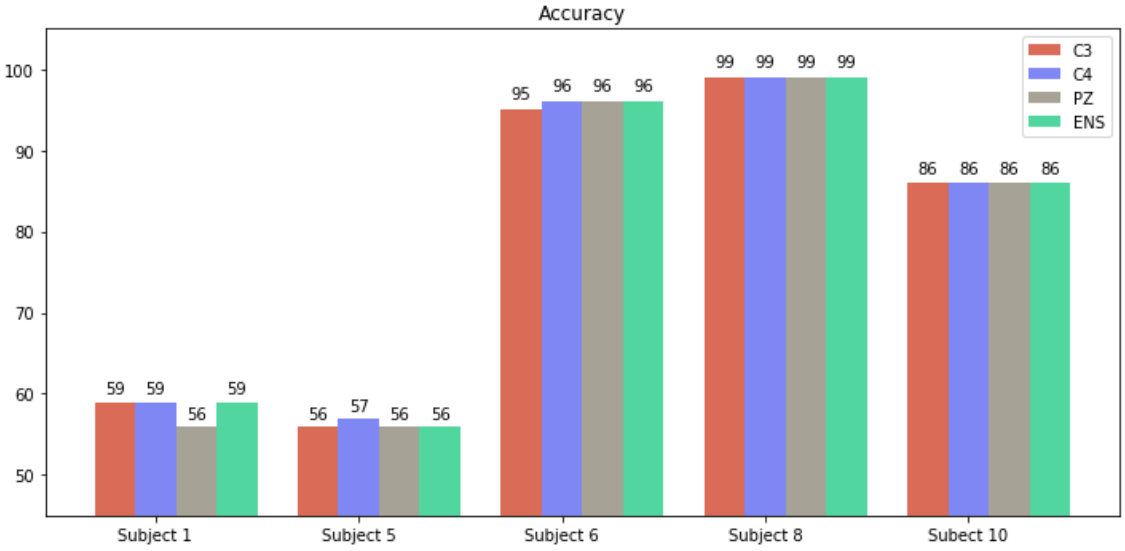
APPENDIX C - Case 3: Specifics subjects with dedicated model

- Raw data



Bagging-based model for five specifics subjects in a dedicated model from the MLP technique with raw data and three-channels

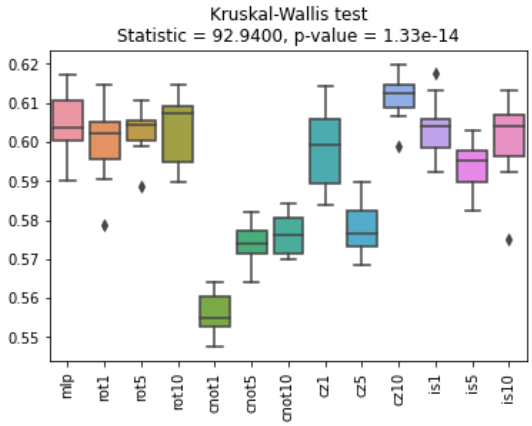
- EEG features



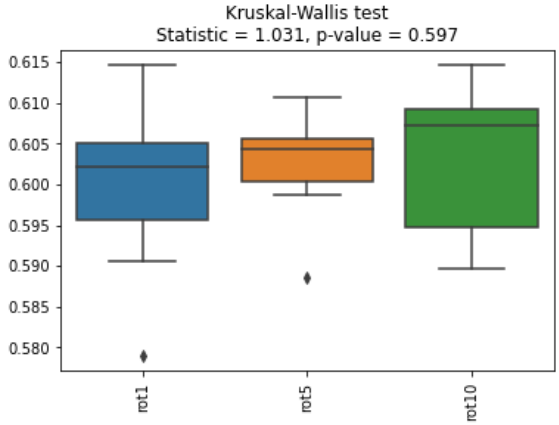
Bagging-based model for five specifics subjects in a dedicated model from the MLP technique with features extraction and three-channels

APPENDIX D - Kruskal-Wallis tests

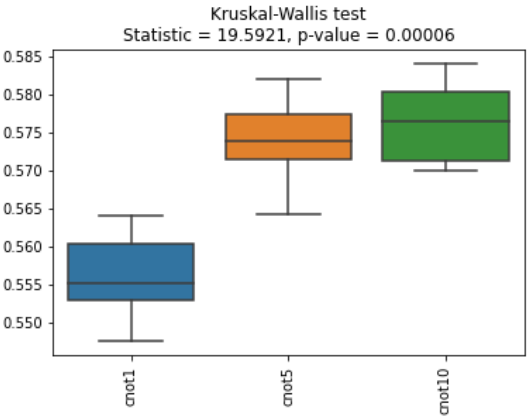
Subject 5



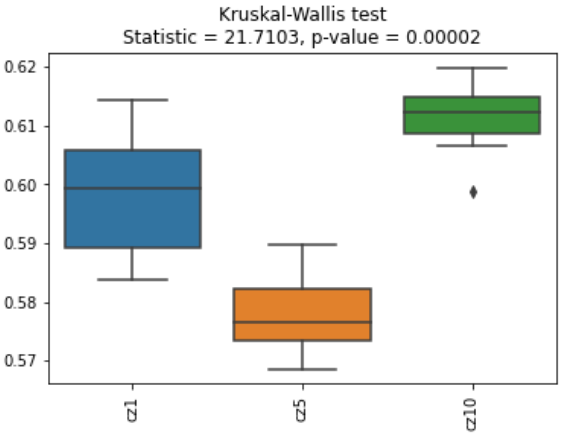
Subject 5 - all architecture models.



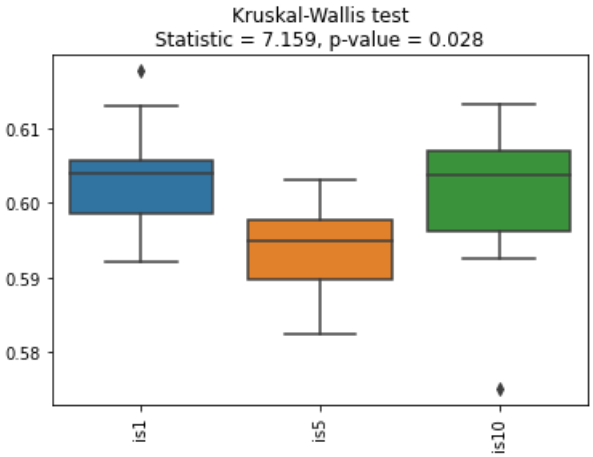
Subject 5 - Ry Rz Ry .



Subject 5 - Ry Rz Ry + CNOT

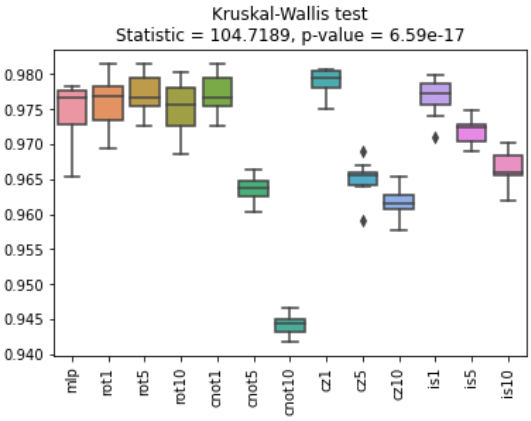


Subject 5 - Ry Rz Ry + CZ

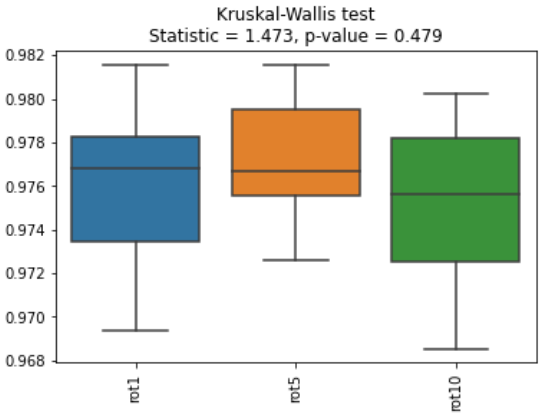


Subject 5 in Ry Rz Ry + iSWAP.

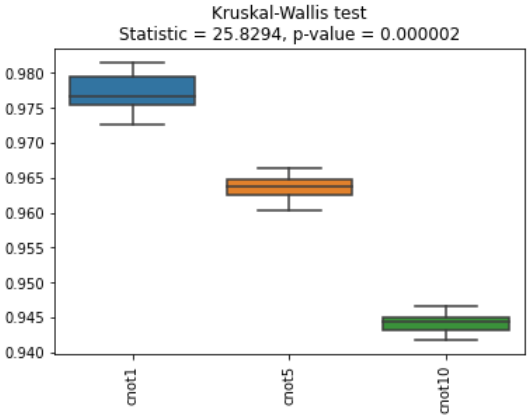
Subject 6



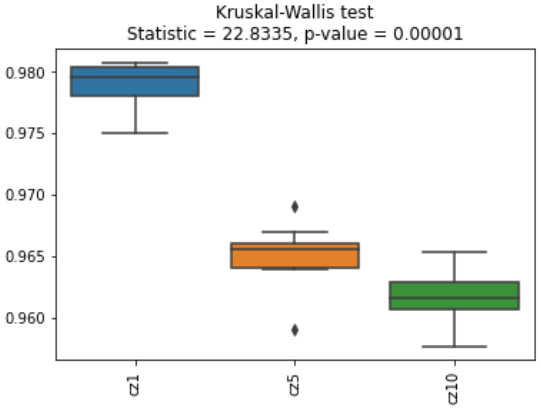
Subject 6 - all architecture models.



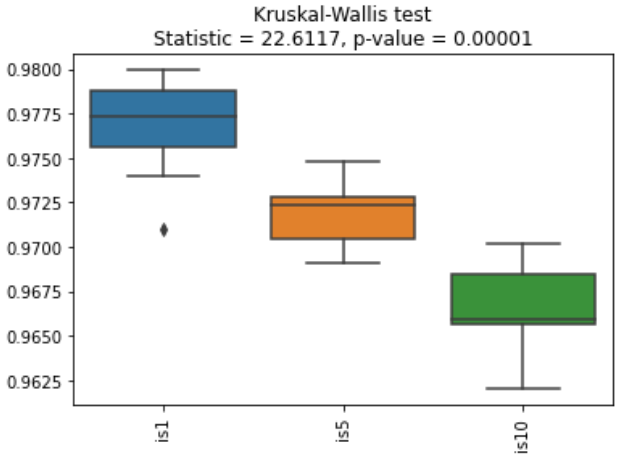
Subject 6 - Ry Rz Ry .



Subject 6 - Ry Rz Ry + CNOT

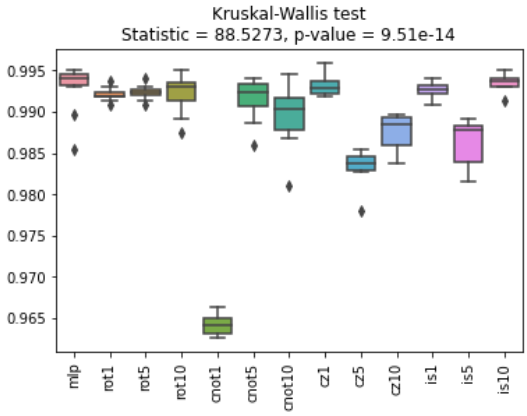


Subject 6 - Ry Rz Ry + CZ

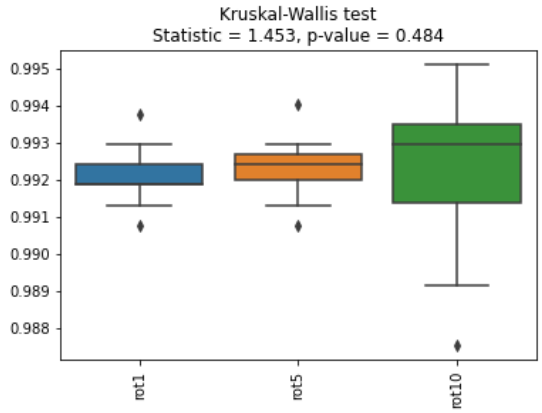


Subject 6 in Ry Rz Ry + iSWAP.

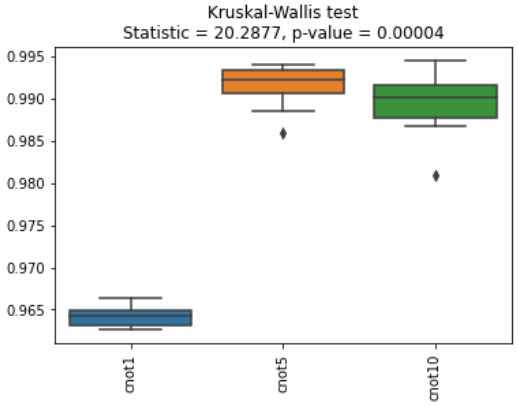
Subject 8



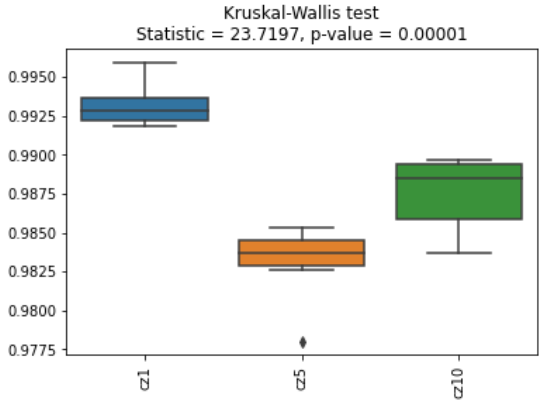
Subject 8 - all architecture models.



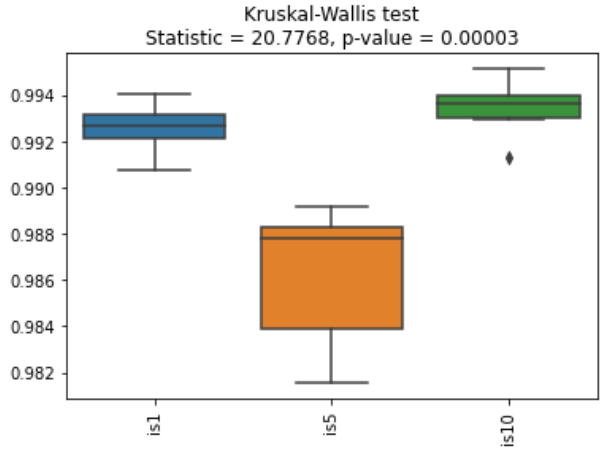
Subject 8 - Ry Rz Ry .



Subject 8 - Ry Rz Ry + CNOT

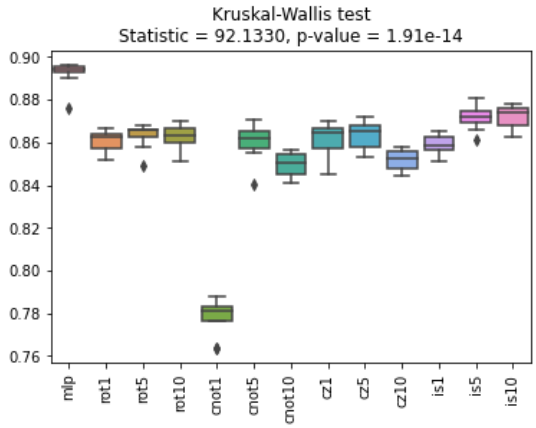


Subject 8 - Ry Rz Ry + CZ

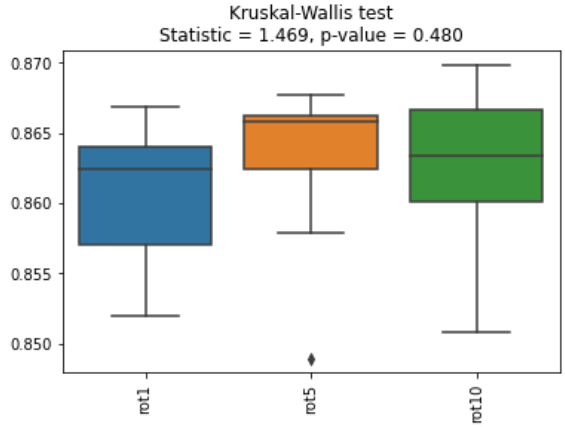


Subject 8 in Ry Rz Ry + iSWAP.

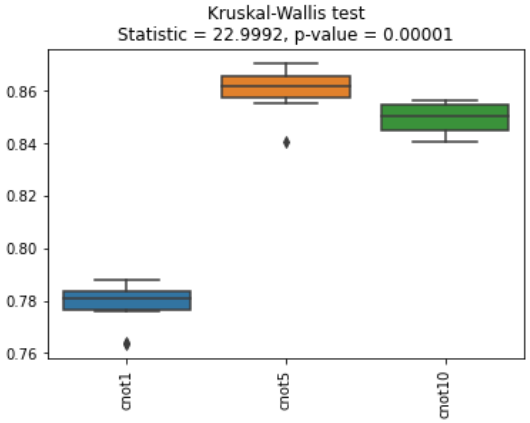
Subject 10



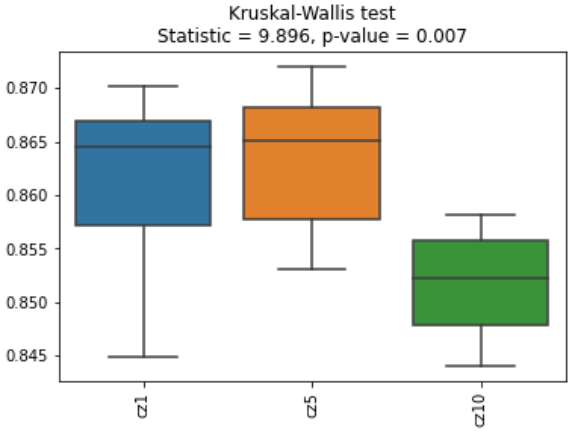
Subject 10 - all architecture models.



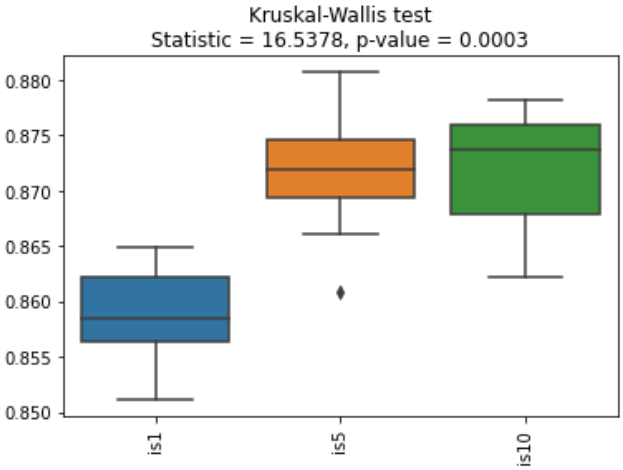
Subject 10 - Ry Rz Ry .



Subject 10 - Ry Rz Ry + CNOT

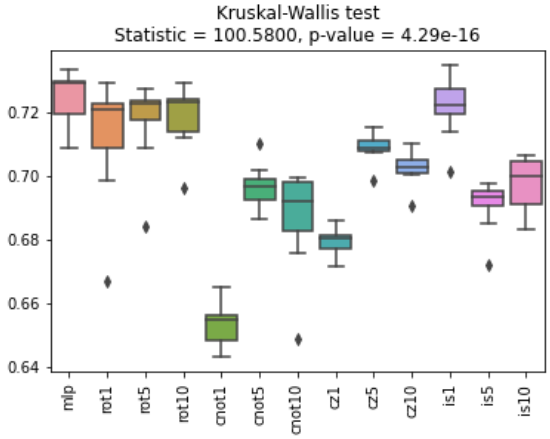


Subject 10 - Ry Rz Ry + CZ

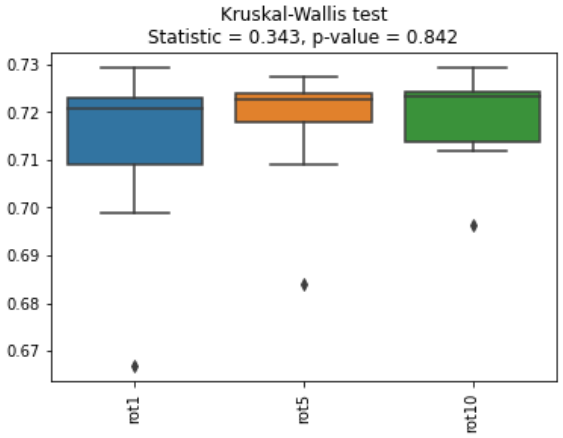


Subject 10 in Ry Rz Ry + iSWAP.

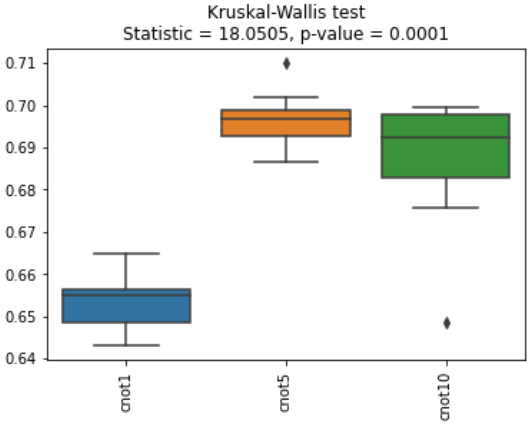
All Subjects



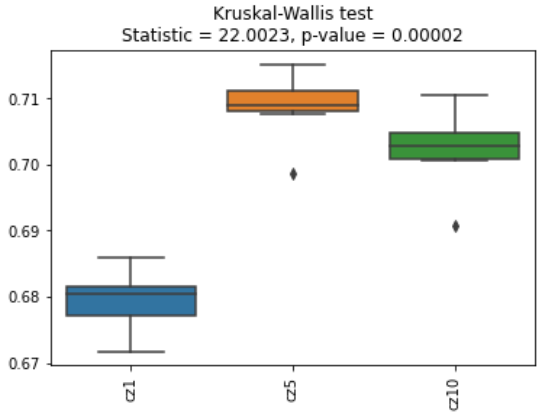
All Subjects - all architecture models.



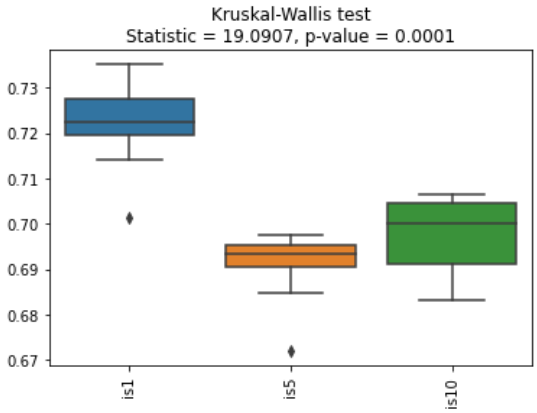
All Subjects - Ry Rz Ry .



All Subjects - Ry Rz Ry + CNOT



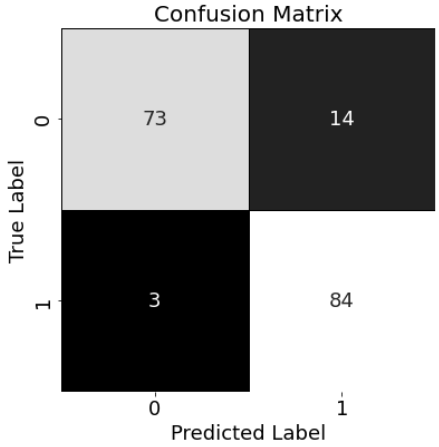
All Subjects - Ry Rz Ry +CZ



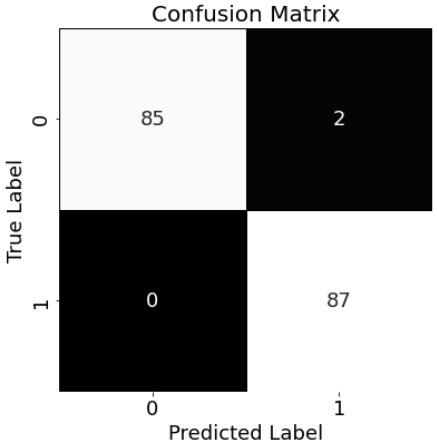
All Subjects in Ry Rz Ry + iSWAP.

APPENDIX E - Extended performance of feature-level models

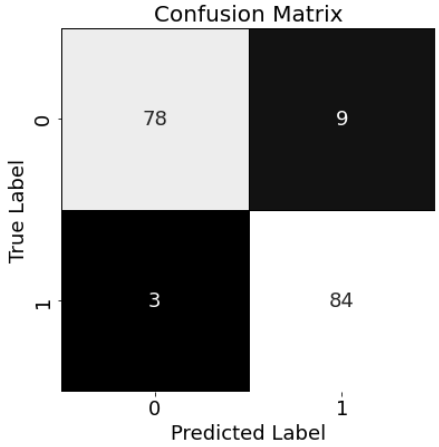
Subject 1



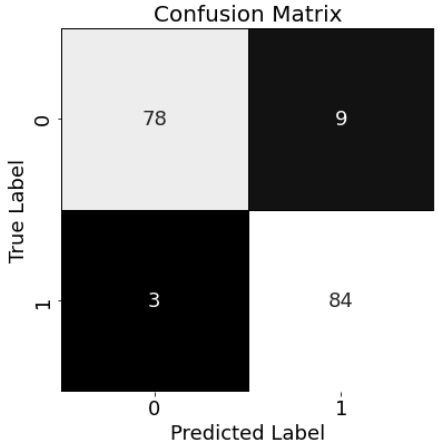
Confusion Matrix for Subject 1, Channel FZ (1-Second Window)



Confusion Matrix for Subject 1, Channel C3 (1-Second Window)

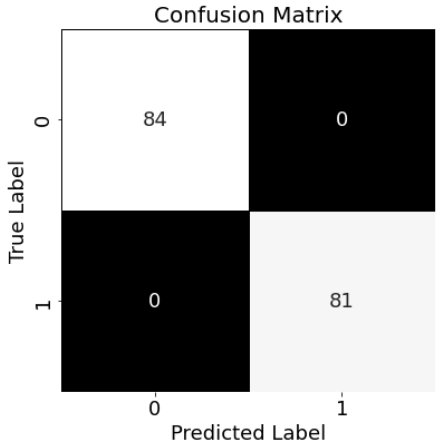


Confusion Matrix for Subject 1, Channel CZ (1-Second Window)

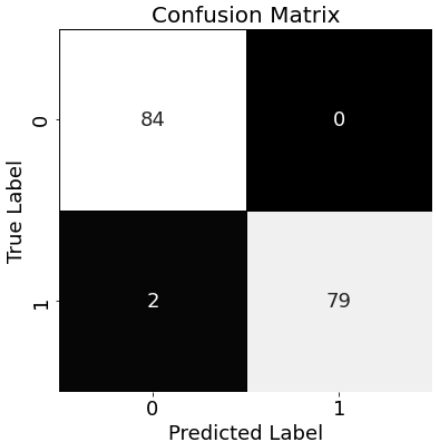


Confusion Matrix for Subject 1, Channel PZ (1-Second Window)

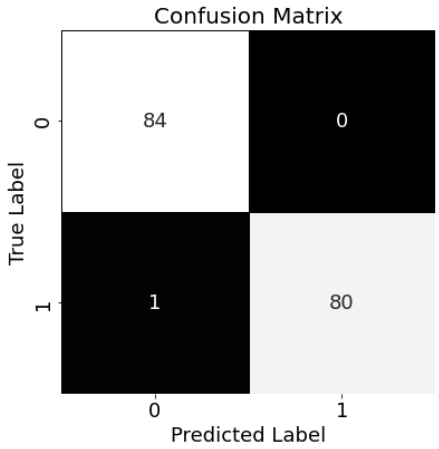
Subject 5



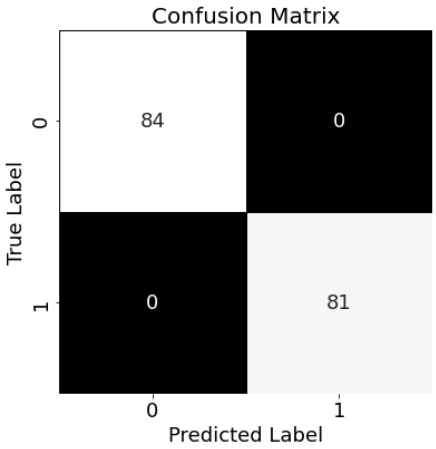
Confusion Matrix for Subject 5, Channel FZ (1-Second Window)



Confusion Matrix for Subject 5, Channel C3 (1-Second Window)

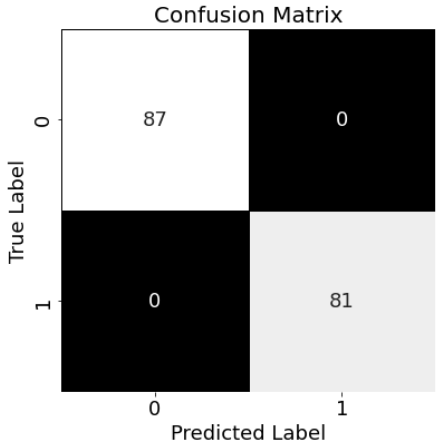


Confusion Matrix for Subject 5, Channel CZ (1-Second Window)

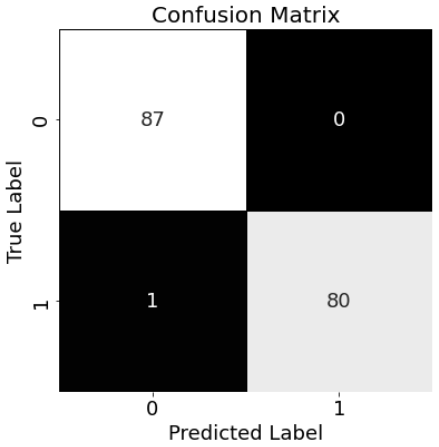


Confusion Matrix for Subject 5, Channel PZ (1-Second Window)

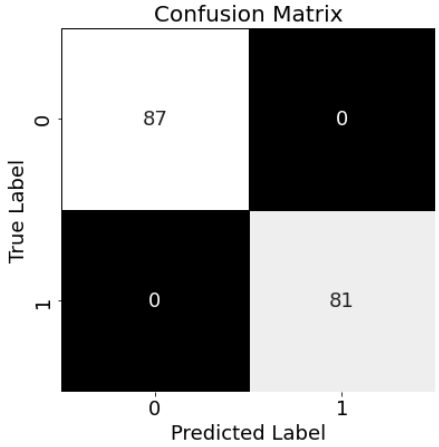
Subject 6



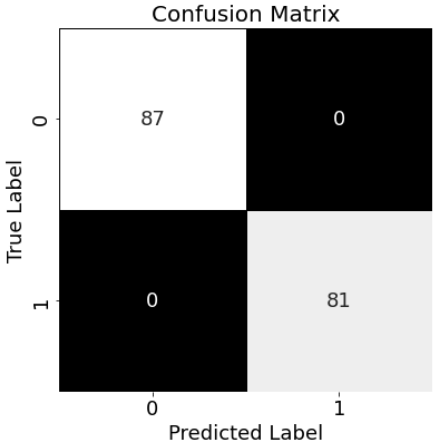
Confusion Matrix for Subject 6, Channel FZ (1-Second Window)



Confusion Matrix for Subject 6, Channel C3 (1-Second Window)

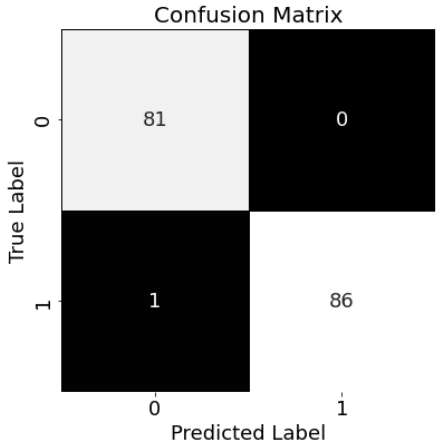


Confusion Matrix for Subject 6, Channel CZ (1-Second Window)

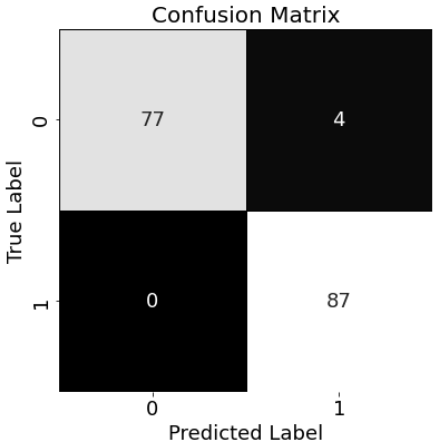


Confusion Matrix for Subject 6, Channel PZ (1-Second Window)

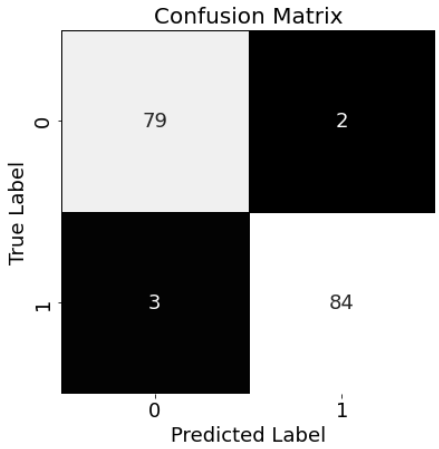
Subject 8



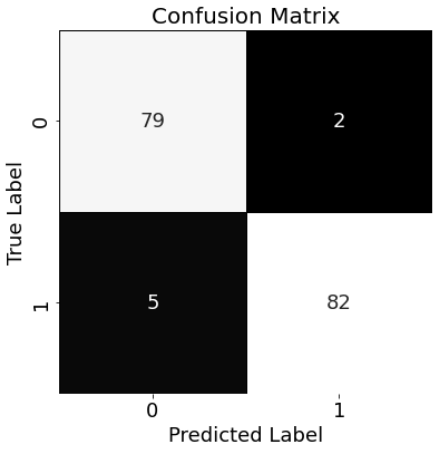
Confusion Matrix for Subject 8, Channel FZ (1-Second Window)



Confusion Matrix for Subject 8, Channel C3 (1-Second Window)

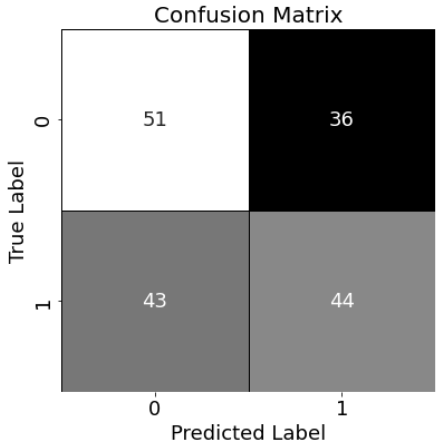


Confusion Matrix for Subject 8, Channel CZ (1-Second Window)

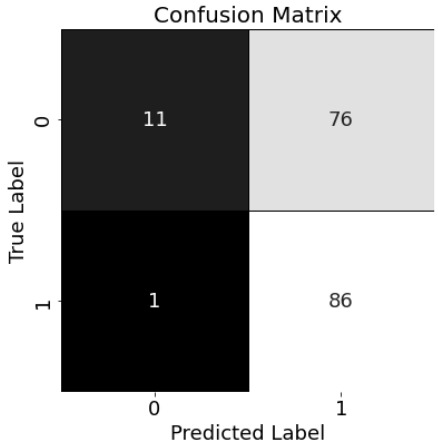


Confusion Matrix for Subject 8, Channel PZ (1-Second Window)

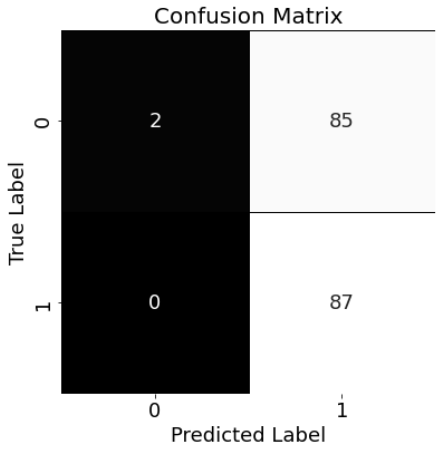
Subject 10



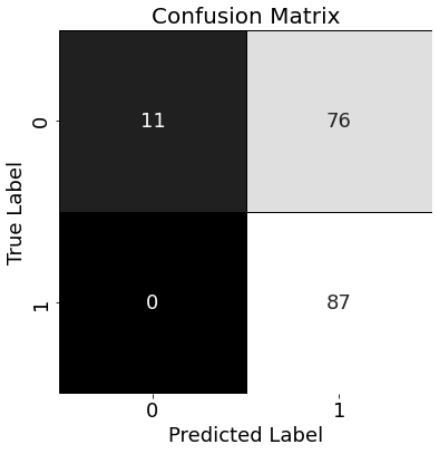
Confusion Matrix for Subject 10, Channel FZ (1-Second Window)



Confusion Matrix for Subject 10, Channel C3 (1-Second Window)

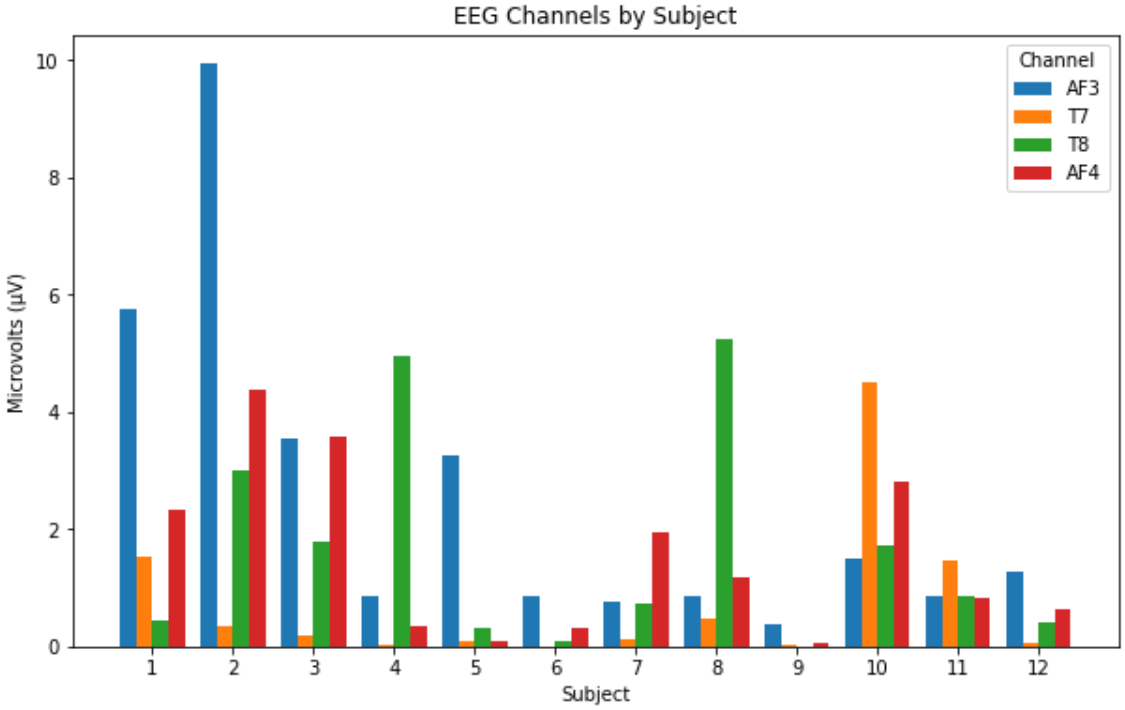


Confusion Matrix for Subject 10, Channel CZ (1-Second Window)



Confusion Matrix for Subject 10, Channel PZ (1-Second Window)

APPENDIX F - Bar Plot Representation of EEG Channels (AF3, T7, T8, AF4)



Average changes in EEG signals for subjects across channels (AF3, T7, T8, AF4)