Universidad de Investigación de Tecnología Experimental Yachay

Escuela de Ciencias Biológicas e Ingeniería

Development of a System to Detect Stress Using Physiological Signals

Trabajo de integración curricular presentado como requisito para la obtención del título de Ingeniero Biomédico.

Autor:

Vásquez Ucho, Paola Aracely

Tutor:

Almeida Galárraga, Diego Alfonso, Ph.D.

Urcuquí - November 10, 2022

Urcuquí, 26 de septiembre de 2022

SECRETARÍA GENERAL
ESCUELA DE CIENCIAS BIOLÓGICAS E INGENIERÍA
CARRERA DE BIOMEDICINA
ACTA DE DEFENSA No. UITEY-BIO-2022-00029-AD

En la ciudad de San Miguel de Urcuquí, Provincia de Imbabura, a los 26 días del mes de septiembre de 2022, a las 15:00 horas, en el Aula S_CAN de la Universidad de Investigación de Tecnología Experimental Yachay y ante el Tribunal Calificador, integrado por los docentes:

Presidente Tribunal de Defensa	Mgs. VILLALBA MENESES, GANDHI FERNANDO
Miembro No Tutor	Dr. MOROCHO CAYAMCELA, MANUEL EUGENIO , Ph.D.
Tutor	Dr. ALMEIDA GALARRAGA, DIEGO ALFONSO , Ph.D.

Se presenta el(la) señor(ita) estudiante **VASQUEZ UCHO, PAOLA ARACELY**, con cédula de identidad No. **0105820799**, de la **ESCUELA DE CIENCIAS BIOLÓGICAS E INGENIERÍA**, de la Carrera de **BIOMEDICINA**, aprobada por el Consejo de Educación Superior (CES), mediante Resolución **RPC-SO-43-No.496-2014**, con el objeto de rendir la sustentación de su trabajo de titulación denominado: **"Development of a System to Detect Stress Using Physiological Signals"** , previa a la obtención del título de **INGENIERO/A BIOMÉDICO/A**.

El citado trabajo de titulación, fue debidamente aprobado por el(los) docente(s):

Tutor	Dr. ALMEIDA GALARRAGA, DIEGO ALFONSO , Ph.D.
Co - Tutor	Dra. SALUM , GRACIELA MARISA , Ph.D.

Y recibió las observaciones de los otros miembros del Tribunal Calificador, las mismas que han sido incorporadas por el(la) estudiante.

Previamente cumplidos los requisitos legales y reglamentarios, el trabajo de titulación fue sustentado por el(la) estudiante y examinado por los miembros del Tribunal Calificador. Escuchada la sustentación del trabajo de titulación, que integró la exposición de el(la) estudiante sobre el contenido de la misma y las preguntas formuladas por los miembros del Tribunal, se califica la sustentación del trabajo de titulación con las siguientes calificaciones:

Tipo	Docente	Calificación
Miembro Tribunal De Defensa	Dr. MOROCHO CAYAMCELA, MANUEL EUGENIO , Ph.D.	10,0
Tutor	Dr. ALMEIDA GALARRAGA, DIEGO ALFONSO , Ph.D.	10,0
Presidente Tribunal De Defensa	Mgs. VILLALBA MENESES, GANDHI FERNANDO	10,0

Lo que da un promedio de: **10 (Diez punto Cero)**, sobre 10 (diez), equivalente a: **APROBADO**

Para constancia de lo actuado, firman los miembros del Tribunal Calificador, el/la estudiante y el/la secretario ad-hoc.

VASQUEZ UCHO, PAOLA ARACELY
Estudiante

Mgs. VILLALBA MENESES, GANDHI FERNANDO
Presidente Tribunal de Defensa

Dr. ALMEIDA GALARRAGA, DIEGO ALFONSO , Ph.D.
Tutor

Dr. MOROCHO CAYAMCELA, MANUEL EUGENIO , Ph.D.
Miembro No Tutor

Karla Alarcón F.

ALARCON FELIX, KARLA ESTEFANIA
Secretario Ad-hoc

Autoría

Yo, **Paola Aracely Vásquez Ucho**, con cédula de identidad **0105820799**, declaro que las ideas, juicios, valoraciones, interpretaciones, consultas bibliográficas, definiciones y conceptualizaciones expuestas en el presente trabajo; así cómo, los procedimientos y herramientas utilizadas en la investigación, son de absoluta responsabilidad de el autor del trabajo de integración curricular. Así mismo, me acojo a los reglamentos internos de la Universidad de Investigación de Tecnología Experimental Yachay.

Urcuquí, Octubre del 2022.

PaolaVz

Paola Aracely Vásquez Ucho
CI: 0105820799

Autorización de publicación

Yo, **Paola Aracely Vásquez Ucho**, con cédula de identidad **0105820799**, cedo a la Universidad de Tecnología Experimental Yachay, los derechos de publicación de la presente obra, sin que deba haber un reconocimiento económico por este concepto. Declaro además que el texto del presente trabajo de titulación no podrá ser cedido a ninguna empresa editorial para su publicación u otros fines, sin contar previamente con la autorización escrita de la Universidad.

Asimismo, autorizo a la Universidad que realice la digitalización y publicación de este trabajo de integración curricular en el repositorio virtual, de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior.

Urcuquí, Octubre del 2022.

Paola Vz

Paola Aracely Vásquez Ucho
CI: 0105820799

Acknowledgments

I want to express my sincere gratitude to all the people who have believed in me through this long journey. I want to acknowledge my advisor Prof. Diego Almeida for the opportunity to work with him, the valuable comments and remarks, and the encouragement through the learning process of this thesis project and in the university classes. At one of his lectures that became my favorite, I got interested in these computational medical topics, and without his guidance, this thesis would not have been possible.

Besides my advisor, I would also like to thank Prof. Fernando, Prof. Si Amar, Prof. Graciela, Prof. Paola, and all my university professors for their instruction, patience, and consideration. I have had the opportunity to get to know them outside the classrooms, and each one of them is a noble person. They have let me work on projects and learn with them; they have encouraged me and guided me through this university journey.

A special thank you goes to the friends I made at the university. I have shared so many great, and sometimes not so great, experiences with all of them. I learn so many lessons just from sharing time with each of them, and I have grown as a person with their company. Chicas, thank you so much for your friendship; you have been so special and kind since we met. We have shared several nights of study, gossip, party, and deep conversations that will always stay in my mind. You have always made me feel in-home besides being hours away from my hometown.

Most importantly, I want to thank God and my family for being the primary emotional support through this long journey. To my parents and my brothers, thank you for your unconditional love and support in me, my studies, and every activity I do in the greatest and most challenging times. You all have always believed in me, even when I don't do it myself. I am so blessed to have amazing parents that have raised me, trying to give me all of you in so many aspects that I could never pay you back; you have taught me the essential things in life that university will never do. Dad, thanks for driving more than 10 hours to go pick me up every semester. Mom, thanks for always having my back and supporting me despite my problems. Edwin, thanks for always being so proud of me in every minor achievement. Mathias, thanks for counting the days to see me and always putting a smile on my face when I come back home.

Finally, to all my grandparents, uncles, aunts and cousins, thank you for always being there for me and believing in my potential. It has always been gratifying to know all of you care about me and my journey.

Abstract

In the 21st century, mental health has become a significant concern worldwide due to the modern lifestyle, environment, and global events. It has given rise to people's awareness about emotions that can evoke not only psychological problems but also physiological diseases, such as stress. As a biological response to external agents that change the body's stability, stress is a problem that had not been addressed in the last few years, and even nowadays in communities without easy access to health care systems. However, it has been proved that long periods of stress exposition can lead to severe diseases regarding different human systems and mental health. Therefore, detecting and predicting stress is vital to alert people and take actions to deal with it and avoid related diseases. Previous research efforts have focused on exploiting the sympathetic nervous system information in response to stress, analyzing physiological signals that can serve as biomarkers to detect stress in daily life activities. Artificial intelligence and its applications in the medical area have supplied several advances for classifying physiological signals for different purposes, including stress detection.

This project aims to develop a monitoring protocol that can provide daily healthcare by detecting stress responses using physiological signals and machine learning techniques, considering the possible use of this system in real-life ambulatory conditions. This work focuses on two objectives:

First, we aim to identify the most suitable physiological signals for stress recognition and carry out a careful preprocessing method that allows the obtention of features that contain vital information concerning the presence of stress in an interval of time. For this stage, electrocardiogram signals are processed, and features are obtained based on the peaks detection. Additionally, galvanic skin response signals are used to determine the body's stress response and label the data. Subsequently, feature selection techniques are performed to eliminate information that does not affect the following steps positively.

Second, we aim to perform supervised classification using machine learning algorithms to detect stress successfully. Once the data is thoroughly processed and normalized, different classifiers are tested to analyze their performance in detecting stress using the electrocardiogram signals. The results of the models are measured using metrics such as accuracy, precision, recall, and F1 score. Several experiments are performed by varying hyperparameters and datasets to detect the most accurate model for this specific task.

It is concluded that the signals selected, the preprocessing methods, feature extraction, and feature selection techniques chosen establish an excellent protocol for the obtention of a dataset that feeds supervised machine learning classifiers. Moreover, after analyzing and comparing the algorithms' performance, it was concluded that the random forest classifier could serve as a robust machine learning model for the detection of stress using electrocardiogram signals that can be monitored in real-life ambulant environments. This work provides a baseline for personalized stress detection, precision medicine, and personalized healthcare using non-complex systems for communities with no easy access to a health care system.

Keywords: Stress, Electrocardiogram, Classification, Machine Learning, Random Forest, Support Vector Machine, Artificial Neural Network, Decision Tree, K-Nearest Neighbor.

Resumen

En el siglo 21, la salud mental se ha convertido en una preocupación significante mundialmente dado el estilo de vida moderno, el medio ambiente, y los eventos globales. Esto ha provocado conciencia en las personas sobre las emociones que pueden provocar no solamente problemas psicológicos, si no también enfermedades fisiológicas, como el estrés. El estrés, como una respuesta biológica a agentes externos que generan un cambio en la estabilidad del cuerpo, es un problema que no había sido considerado algunos años atrás, e incluso hoy en día en comunidades donde el acceso a los sistemas de salud no es fácil. Sin embargo, ha sido probado que periodos largos de exposición a estrés pueden conducir a severas enfermedades que conciernen algunos sistemas del cuerpo y la salud mental. Es por esto que detectar y predecir el estrés es de vital importancia para alertar a las personas, tomar acciones, tratar este problema y evitar enfermedades relacionadas. Los estudios se han enfocado en aprovechar la respuesta del sistema nervioso simpatético mediante el análisis de señales fisiológicas que puedan servir como biomarcadores para detectar el estrés en actividades diarias. La inteligencia artificial y sus aplicaciones en el área medica han proveído varias ventajas para la clasificación de señales fisiológicas con diferentes propósitos, incluyendo la detección de estrés.

Este proyecto busca desarrollar un protocolo de monitoreo que provea cuidado de la salud diario mediante la detección de las respuestas al estrés usando señales fisiológicas y técnicas de aprendizaje automático, considerando el posible uso de este sistema en condiciones ambulatorias de la vida real. Este trabajo se enfoca en dos objetivos:

Primero, se aspira identificar las señales fisiológicas mas aptas para el reconocimiento de estrés y llevar a cabo un cuidadoso método de pre procesamiento, que permita la obtención de características que contengan información vital respecto a la presencia de estrés en un intervalo de tiempo. Para esta etapa se procesan señales electrocardiográficas y se obtienen las características en base a la detección de picos de las señales. Adicionalmente, las señales de respuesta galvánica de la piel son usadas para determinar la respuesta del cuerpo al estrés, y con esto etiquetar los datos. Posterior a esto, se ejecutan técnicas de selección de características para eliminar la información que no afecta de manera positiva los siguientes pasos.

Segundo, se busca realizar clasificación supervisada usando algoritmos de aprendizaje automático para detectar exitosamente el estrés. Una vez que los datos son minuciosamente procesados y normalizados, se prueban diferentes clasificadores para analizar su rendimiento en la detección de estrés usando las señales electrocardiográficas. Los resultados de los modelos son medidos usando métricas como la exactitud, precisión, recuerdo, y el puntaje F1. Varios experimentos son llevados a cabo con variaciones en los hiperparámetros y conjunto de datos, para detectar el modelo más exacto en esta tarea en específico.

Se concluye que las señales elegidas, el método de pre procesamiento, las técnicas de extracción de características y selección de características escogidas establecen un excelente protocolo para la obtención de un conjunto de datos que puede satisfacer un clasificador supervisado de aprendizaje automático. Además, después de analizar y comparar el rendimiento de los algoritmos, se concluyó que el clasificador de nombre *bosque aleatorio* puede funcionar como un modelo de aprendizaje automático que detecte estrés mediante el uso de señales electrocardiográficas, que puedan ser monitoreadas en ambientes ambulantes de vida real. Este trabajo provee una base para la detección personalizada de estrés, para la medicina de precisión y para el cuidado de la salud personalizado usando sistemas no complejos para comunidades que no tienen acceso fácil al sistema de salud.

Palabras Clave: Estrés, Electrocardiograma, Clasificación, Aprendizaje Automático, Bosque Aleatorio, Máquina de Vector de Soport, Red Neuronal Artificial, Árbol de Decisión, K-Vecino más Próximo.

Nomenclature

ACTH Adreno-Corticotropic Hormone

ADC Analog-to-Digital Conversion

AI Artificial Intelligence

ANN Artificial Neural Network

ANS Autonomic Nervous System

AV Atrio Ventricular

AWGN Additive White Gaussian Noise

BN Bayesian Network

BP Blood Pressure

bpm beats per minute

BW Baseline Wander

CNN Convolutional Neural Network

CV Cross-Validation

CVD Cardiovascular Disease

DAC Digital-to-Analog Conversion

DASS Depression, Anxiety and Stress Scale questionnaire

DL Deep Learning

DNN Deep Neural Network

DSP Digital Signal Processing

DT Decision Tree

DWT Discrete Wavelet Transform

$E-nose$ Electrical Nose System

ECG Electrocardiogram

EDA Electrodermal Activity

EDL Electrodermal Level

EDR Electrodermal Response

EEG Electroencephalogram

EMG Electromyogram

$FCNN$ Fully Connected Neural Network

FFT Fast Fourier Transform

FIR Finite Impulse Response

FN False Negative

FP False Positive

GAS General Adaptation Syndrome

GSR Galvanic Skin Response

HPA Hypothalamic-Pituitary-Adrenocortical

HRV Heart Rate Variation

IBI Inter-Beat Interval

IIR Infinite Impulse Response

KNN K-Nearest Neighbor

LA Left Atrium

LDA Linear Discriminant Analysis

$LMIC$ Low-and-Middle-Income Country

LR Logistic Regression

LV Left Ventricle

MA Muscle Artifact

ML Machine Learning

MLP Multlayer Perceptron

NB Naive Bayes

NN Normal to Normal

OSI Occupational Stress Inventory

PC Pearson's Correlation

PCG Phonocardiogram

PLI Powerline Interference

PPG Photoplethysmogram

PPV Positive Predictive Value

PSD Power Spectral Density

PSS Perceived Stress Scale

RA Right Atrium

$RESP$ Respiration

RF Random Forest

RFE Recursive Feature Elimination

RNN Recurrent Neural Network

RV Right Ventricle

SA Sino Atrial

SAM Sympathetic-Adrenimedullary

SCL Skin Conductance Level

SCR Skin Conductance Response

SGD Stochastic Decent Gradient

SMO Sequential Minimal Optimization

SNS Sympathetic Nervous System

SP Skin Potential

SPR Skin Potential Response

$STAI$ State-Trait Anxiety Inventory

SVM Support Vector Machine

T Body Temperature

TN True Negative

TP True Positive

$TSST$ Trier Social Stress Test

ZSN Z-Score Normalization

Contents

List of Tables

List of Figures

List of Equations

Chapter 1

Introduction

1.1 Problem statement

In the modern world, due to the current lifestyle, more people are suffering from stress, often unaware. Furthermore, since the COVID19 pandemic started, stress level assessments have increased for millennials, adults, and older adults, while health, money, and decision-making are the main stressors [1]. Stress is a term mainly associated with high levels of psychological tension and is often faced over an extended period causing adverse effects on heart rate (HR) and blood pressure (BP) that consequently harm the body and have serious long-term consequences. If stress can be detected early, people can be warned and take appropriate policies to cope with it. Some years ago, many people believed stress had little or no impact on physical or mental health [2]. Nonetheless, stress has proven to be a precursor of major illnesses such as depression, obesity, and even cardiovascular diseases, known for being the first cause of morbidity and mortality in America and Europe [3].

On the other side, the relationship between poverty and health care has been addressed. Stress is a global health problem, especially in developing countries [4]. Health services are less accessible in low- and middle-income countries (LMICs), such as many countries in Latin America, Ecuador not being an exception as a middle-income country [5]. This problem has been tried to approach by many researchers, governments, and commercial organizations through health equity funds and regulation of health services. However, the challenge remains since these approaches focus directly on treating diagnosed fatal and chronic diseases that can increase mortality levels, such as diabetes or hypertension, but not on preventing by early detection of 'minor' problems that can lead to these diseases, such as stress.

Therefore, people have recognized that stress, especially long-term exposure, negatively impacts health; hence, interest has increased in managing it before it becomes chronic by accurately recognizing when and where it occurs. Many detection systems have been developed based on questionnaires, tests, and physiological measures such as hormone levels or biological signals. Biosignals have taken the lead among the most used stress detection methodologies. The focus has been placed on studying different physiological signals to classify stress levels in people. Many systems proposed in the literature contain a hardware-based system that records relevant features by multi-physiological sensors and a software-based architecture that predicts a person's current stress level. Sensors to measure physiological signals for stress detection are based on electrocardiogram (ECG), electroencephalogram (EEG), and electromyogram (EMG) signals; still, the medical devices and biosensors used for these purposes are relatively expensive and not easy to access in some developing countries. Developing multiparameter hardware for stress detection requires more expenses and time, while physically, a higher number of sensors will result in more discomfort for the subject provoking bad contact between sensors, non-accurate measures, or stress induction during experiments. The software becomes computationally intensive with more parameters to process, noises to eliminate, and details to take into account, resulting in non-accurate measures.

For stress classification, novel computational techniques, such as Artificial Intelligence, are being studied. Classification technologies require accurate data, labels, and ground-truth information as the primary source for the correct stress classification and to avoid false diagnoses in people who do not suffer stress or for people who suffer it and need to be treated. The experimental scenario is not the same as an actual situation, so assessing stress in real-life situations can be challenging, especially in developing countries where technology and experts can be limited. For this reason, it is indispensable to research and develop accurate systems using the appropriate tools that can provide reliable data for detecting and classifying stress in a real-life scenario.

1.2 Contribution

The present work aims to find an optimal set of techniques and methods of medical signal preprocessing, extraction, and recognition to classify stress using electrocardiograph data. The process involves using ECG signals for proper cleaning by signal filtration methods, extracting HR data from the signals, and its time and frequency domain features. Features obtained are filtered using features selection techniques to select relevant features that provide relevant information related to stress and do not compromise the following classification step. The final stage uses different machine learning techniques for binary stress classification. All steps performed are based on the literature and experimental results and compared to find the best sequence of techniques that provide accurate detection. Based on the literature, this work is done under the hypothesis that physiological signals provide reliable biomarkers for stress detection.

1.3 Objectives

1.3.1 General Objective

Develop a method to diagnose psychological or physical stress using ECG signals that serve as input to a machine learning algorithm that predicts if the subject is or is not under stress to avoid chronic stress episodes and the diseases that follow it.

1.3.2 Specific Objectives

- Process ECG signals using filters that eliminate their noises and enable the proper recognition of peaks indispensable for the obtention of HR data.
- Extract features from the HR processed data using time and frequency domain information.
- Carry out feature selection process to remove data correlated or feature with unpredictive power
- Use galvanic skin response (GSR) signals to label the data as stress or no stress and serve as a dataset to enable supervised machine learning classification.
- Perform binary supervised classification using different ML models by changing features and tuning to select the most accurate classifier with its optimal hyperparameters that can detect the presence of stress in subjects.

1.4 Chapter-by-Chapter Overview

This undergraduate thesis is structured as explained below:

Chapter 2 presents the definition of stress from its beginnings and how it evolved over the years. It explains this stimulus's psychological and physiological effects on some human systems. The human body's response to stress is detailed, and the diseases or pathologies resulting from a person being exposed to prolonged stress. The conventional methods used to detect stress are briefly described, including examples. Nowadays, physiological stress indicators used to detect stress, such as physiological signals, are exposed, explaining their relationship to stress and their extraction methods. Further, it introduces different supervised machine learning techniques and their advantages and disadvantages for this classification task.

Chapter 3 exhibits the current information on stress and its detection during the last five years. It exposes the existing research regarding the physiological signals used for stress analysis. The signal processing techniques used to filter ECG signals are investigated, considering the different noises they can have. Feature extraction methods for detecting the peaks and intervals essential for the HRV analysis are exhibited based on past works. Finally, a literature review of the latest stress detection systems is tabulated, evidencing the different ML models employed, the type of data used as input, the aim of the study, and the metrics obtained in the results of each work.

Chapter 4 explains the programming language, libraries, and other software tools used to develop this work. It explains each step of the process for the classification of stress. The physiological data, signal preprocessing method, feature extraction, feature selection techniques, and machine learning classification models are detailed with a strong argumentation of why each process is selected for this stress classification task. Moreover, it introduces the stress detection models proposed based on the literature and how these models' performances are measured using different metrics to compare and determine the most suitable one.

Chapter 5 presents the results of each methodology stage described in Chapter 4. It starts by illustrating the ECG signals to notice the filters' successful effect on the signals. The features extracted from the signals are listed, and the dataset obtained from them and the labels acquired from the GSR signals. The set of features selected with each feature selection technique is displayed. Each classifier tested in this study is exposed, detailing their ideal hyperparameters after tuning and explaining how the parameters affect each model. The results obtained by each ML classifier are listed in terms of accuracy, recall, precision, and F1 score. The highest scores obtained from each model are highlighted and compared to select the one that outperforms the others. Techniques variations show different experiment results listed.

Chapter 6 analyzes the results obtained and exposed in 5. Each step of the methodology performed for the stress classification is discussed and compared to previous works to find similarities, differences, and contrast results. Machine learning classifiers are deeply analyzed, and their performances are compared based on the metrics and state-of-the-art works, datasets, methodologies, and results.

Chapter 7 presents the conclusions of this thesis work and proposes relevant future work.

Chapter 2

Theoretical Framework

2.1 Stress Definition

Hans Selye, M.D., Ph.D., D.Sc., F.R.S., often called the "Einstein of the Medical Research," authored the short landmark letter published in Nature over 85 years ago about the general adaptation syndrome (GAS) or as we know it nowadays: biologic stress [6]. The scientific publication describes his experiments in rats exposed to severe stressors, resulting in the most stereotypical manifestations of the 'general alarm reaction of the organism'. Although the experiments were proven in animal models, he got this idea by observing sick patients as a medical student. He noticed that besides being sick because of an organ malfunction, there was a "common look" in all of the patients; they all had "non-specific" common symptoms such as loss of appetite, decreased muscular strength, elevated BP, and a loss of ambition [7]. Therefore, stress was first defined as "the non-specific neuroendocrine response of the body". However, since almost all other organ systems are involved in stress, especially cardiovascular, pulmonary and renal systems, the 'neuroendocrine' part was later removed. However, there was always an emphasis on non-specificity as the main characteristic of these agents that trigger the stress response.

Many years before, in 1676, stress was already defined in physics the Hooke's law which described the effect of external stresses, or loads, that produced various degrees of "strain," or distortion, on different materials. Therefore, confusion began about whether stress is referred to as a "stimulus" in physics or a "response," as mentioned by Selye. Finally, the term stress was used to denote a response; in fact, the stimuli that engender the stress response was named 'stressors' and can be physical, chemical, or psychologic in nature [8]. These stressors can come in different manners, from a physical insult such as trauma or injury, or physical exertion, when the body has to operate beyond its capacity, to other physical stressors like noise, overcrowding, excessive heat or cold, and psychological experiences as time-pressured tasks, interpersonal conflict, unexpected events, frustration, isolation, and traumatic life events [9]. However, they differ in the extent to which they can be controlled. For example, people have been shown to feel less pain and anxiety during dentistry procedures when they are told they have the control to stop it whenever they provide a characteristically signal, a phenomenon known as cognitive control [10].

This Stress concept was evolving over the years, and it took almost four decades to recognize that based on each subject's perception and emotional reaction, not all stress reactions are equal; there was a distinction between "positive" and "negative" stress [11]. Stress can often be defined as a 'threat' to homeostasis, which is the stability of some physiological systems that maintain life, such as pH, body temperature (T), glucose levels, and oxygen tension essential for life. The term "stress" is full of ambiguities; it is used to refer to an adverse event that causes a response (distress) or to refer to a "good" challenge that leads to feeling exhilaration. However, considering the organism's homeostasis, stress can describe threatening events that elicit behavioral and physiological responses as part of allostasis, which refers to the process that supports homeostasis by achieving stability through change [12]. In the same way, it has been defined based on stressors, such as psychological or physiological. Psychological stress is a reaction to an aversive stimulus in the external environment, while physiological stress has been defined as the disturbance of an individual's internal milieu, leading to the activation of regulatory mechanisms that restore homeostasis [13].

2.2 Physiological Stress Response

A stress response is a coordinated pattern of changes formed by natural selection to allow organisms to face specific situations that can cause damage and require action or defense. Many years before Selye, Walter Cannon highlighted changes in fight or flight situations that are useful for the body [14]. The changes included increased HR to speed circulation, increased rate and depth of breathing to speed gas exchange, sweating to cool the body and make it slippery, increased glucose synthesis to provide energy, shunting of blood from gut and skin to muscles, increased muscle tension to increase strength, and increased blood clotting in preparation for possible tissue damage [15]. In general, every type of stressor may produce behavioral responses, and physiological reactions in the body [9]. Behavioral responses are designed to get the individual out of risk and lessen the probability of death by promoting healthy activities (good diet, regular exercise). These responses prepare the body to survive physical threats by mobilizing stored energy, increasing cardiac output, and suppressing nonessential digestive, immune, and reproductive functions.

On the other side, the physiological effects of stress include alterations in the neuroendocrine, autonomic nervous system, and immune function, so these are of great interest for researchers due to their implications in many diseases [16]. Stress responses primarily activate two nervous system pathways: the hypothalamic-pituitary-adrenocortical (HPA) axis and the sympathetic–adrenomedullary (SAM) system. The immediate response mediated by the sympathetic nervous system (SNS) and SAM release catecholamines into the bloodstream, such as epinephrine and norepinephrine, also known as adrenaline and noradrenaline, from the adrenal medulla. It also results in increased cardiovascular arousal, and the halting of nonessential parasympathetic functions [17]. A more delayed stress response mediated by the HPA axis, also known as the adrenocortical response: adrenocorticotropic hormone (ACTH), is characterized mainly by the production and release of glucocorticoids such as cortisol, corticosterone, and other corticosteroids from the adrenal cortex [18, 19].

These physiological responses are characterized by having protective and damaging effects. The catecholamines of the SNS and the glucocorticoids from the adrenal cortex initiate cellular events that promote adaptive changes in cells and tissues throughout the body, protecting the organism and promoting survival. However, excessive or prolonged exposure to cortisol is associated with accelerated aging, and increased risk of cognitive impairments, cardiovascular disease, infectious diseases, and other illnesses [20] and may significantly impact the progression of chronic illness. Efficient and flexible physiological responses to stress are adaptive in the short run. However, pronounced or repeated and delayed stress responses are thought to contribute, over time, to the etiology of hypertension, heart disease, infectious diseases, and other illnesses [21]. Clinical observations and experiments have asserted that prolonged exposure to stress results in the overproduction of chemicals and hormones, causing gastroduodenal ulcers and high BP (diseases of adaptation) [22]. Namely, too much stress or inefficient operation of the stress acute responses can cause and exacerbate disease processes.

2.3 Stress and Diseases

Stress plays a vital role in individual survival and the development and aging of people. Therefore, the intensity and duration of stress are essential factors determining the effects produced on the organism. Nowadays, stress is a daily experience since many aspects or situations in today's lifestyles may not qualify as stressors but can adversely affect people. Additionally, the body systems operation that promotes adaptations and homeostasis and the damage it can cause under extreme circumstances is not fully considered [23].

Stress can also be classified as acute and chronic stress. Acute stress is the transient exposure to various distressing or challenging tasks and short-duration naturalistic events; it has been associated with a rapid release of chemical mediators such as catecholamines that increase HR and BP: "fight or flight" response. The acute stress responses help the individual cope with the situation and promote adaptation and survival via neural, cardiovascular, autonomic, immune, and metabolic systems [24]. On the other hand, it becomes chronic when stress is intense in amplitude. Chronic stress is present in different forms, from traumatic single-life events to the accumulation of daily hassles. More than 40 years ago, The Surgeon General's Report declared that when stress reaches excessive proportions, psychological changes can be so dramatic as to have profound implications for mental and physical health [25].

A chronic elevation of the mediators, HR, and BP produces regular wear and tear of the organ systems and results in dysfunctions or target organ diseases that can lead to physical disorders [26]. Chronic stress has been associated with the accumulation of adipose tissue [27]and insulin resistance [28], which contributes to the development of obesity [29], diabetes [30] and development and progression of cardiovascular disease (CVD) [31]. Moreover, obesity can be stressful due to the high prevalence of weight stigma, and it induces a low-grade inflammation mediated by proinflammatory adipokines that activate the acute phase reaction and act as an additional chronic stimulus to the stress system activation. Indeed, it results in a vicious cycle whereby chronic activation of the stress system contributes to obesity-related inflammation and insulin resistance, and vice versa

[32]. Furthermore, chronic stress results in prolonged HPA axis activation, leading to the hypersecretion of glucocorticoids that can have detrimental effects on neural structure and function [33].

In the same way, many pathological conditions are associated with altered activity of the HPA axis, such as severe chronic disease, panic disorder, and malnutrition, among others. Additionally, stress causes a psychological reaction that impairs neuronal structure and function and leads to emotional disorders or psychiatric diseases such as anxiety, depression, somatization, and hostility [34]. Therefore, there is a great interest in researching and developing methods for stress detection to diagnose extensive episodes of stress that can lead to chronic responses, thereby avoiding the problems or diseases that can be caused in the short or long term.

2.4 Methods for Stress Detection

As mentioned before, stress is a common problem since stressors can take different forms, such as loud noise, extreme temperatures, physical interventions, pathogens, social situations, and emotional arousal. Since continuous stress can lead to problems and pathologies, stress responses and reaction symptoms such as anxiety, depression, and physiological problems have been widely studied for their early and accurate detection. There is a demand to obtain stress assessments of people in real life to offer solutions and avoid the problems related to late detection of stress. Therefore, numerous metrics to measure stress levels have been studied over the years.

2.4.1 Conventional Stress Evaluation

Self-report questionnaires are the most common measure of stress focused on behaviors while performing specific tasks. Several questionnaires have been designed to assess an individual's degree of chronic stress. Different questionnaires focus on assessments, such as primary and minor life events, real-time experiences, moods, and environmental self-reports. During the past decades, many tests for human acute stress detection have been developed.

The State-Trait Anxiety Inventory (STAI) is a commonly used measure of trait and state anxiety [35]. STAI is a self-report for adults designed to measure feelings of immediate anxiety that an individual feels at the current moment (state anxiety) and dispositional anxiety (trait anxiety). It is commonly used in clinical settings to diagnose anxiety and distinguish it from depressive syndromes. It also is often used in research as an indicator of distress. It consists of 40 items total: a 20-item state anxiety scale where participants report the intensity of their anxious feelings "right now, at this moment", and a 20-item trait anxiety scale where they endorse how they generally feel regarding anxious thoughts and feelings.

The Depression, Anxiety, and Stress Scale questionnaire (DASS 21) is a 21 questions survey used to screen the symptoms of these mental illnesses. It is a suitable tool for measuring stress for research and clinical purposes and is a validated tool among various ethnicity and population groups. There is an extended version with 42 questions (DASS 42); however, the short form has been broadly studied and used for stress detection [36]. The DASS 21 has three scales intended to assess depression, anxiety, and stress, each containing seven questions, and the final scores are obtained through the sum of the question scores. Then the final score of each subscale must be doubled since this questionnaire is a shortened major scale form.

The Perceived Stress Scale (PSS) is one of the most widely used tools for measuring psychological stress in clinical and non-clinical situations. It measures stress by asking participants to report whether their lives seem unpredictable, uncontrollable, or overloaded [37]. PSS evaluates the thoughts and feelings of subjects about stressful events that occurred in the month before the detection. The original English version has 14 items (PSS-14), but PSS is also available in two shortened versions of 10 items (PSS-10) and four items (PSS-4). It is considered a brief measure of perceived stress that can be administered in a few minutes.

The Occupational Stress Inventory (OSI) helps assess the same stress-related variables across different occupational groups [38]. Its revised edition (OSI-R) model is theory-based and assesses the effects on the individual of three factors (occupational roles, psychological strain, and coping resources). The OSI-R results from one such model of stress that incorporates the significant variables impacting stress or the outcomes of stressful situations. It measures domains of occupational adjustment that include occupational stress and stress associated with subject's inability to manage stressors effectively in the environment, each assessed through a comprehensive questionnaire.

Many others focus on detecting other psychological illnesses such as depression, anxiety, trauma, or thought of death. On the other side, there are also tests to induce stress on subjects, enabling the accurate study of stress in subjects, such as the Trier Social Stress Test. The Trier Social Stress Test (TSST) is one of the most widely used psychosocial stressor tasks during the past two decades [39]. It is a robust tool to induce acute psychobiological stress and a reliable tool for examining the effects of acute stress psychologically and physiologically in humans. TSST consists of an anticipation phase followed by a 5-min mock job interview and a 5-min mental arithmetic task in front of a non-responsive jury of two or three people. However, there are significant variations in the timing of events, the number and method of biological sampling, the set-up of equipment and rooms, panel composition, and panel interaction with participants. It leads to a significantly stronger stress reaction than other cognitive stressor tasks such as simple arithmetic or Stroop tasks.

The TSST has been shown to affect several psychobiological measures, such as psychological measures ex.: anxiety, negative mood, and perceived stress, and autonomic measures, ex.: BP, HR, T, and electrodermal activity (EDA) [40].

2.4.2 Physiological Stress Indicators

In contrast to the traditional retrospective questionnaire approach, real-time assessment involves multiple prospective assessments of the respondent's current experience. As mentioned above, stress activates different physiological systems, such as the HPA or the ANS, producing a high presence of stress hormones. Therefore, the physiological indicator of emotion or stress can be measured using physiological parameters such as cortisol. Cortisol is a well-established hormonal mediator of the stress response, which can act as an output, meaning levels decrease in relaxation phases or as an input as a cause leading to behavioral inhibition in people [41]. This stress indicator is relatively accessible to researchers and does not require invasive or stressful collection methods such as plasma or urine. The most common way of measuring cortisol is through blood tests, making it tedious to assess during experiments to study stress.

Stressful situations can provoke changes in other systems, like autonomic cardiovascular activity, leading to changes in heart rhythm. For example, high heart activity and low HR regulation have been associated with high-stress levels, and depression [42]. Stress induces changes in these physiological signals that can be monitored with technological advances such as mHealth, which refers to the use of mobile information and communication technology, such as mobile computers, medical sensors, and wearable devices in healthcare [43]. Biosensors are tools used to detect the humans' physiological changes, such as brain activation, EDA, HR, muscle tension, and respiration rate, often with a particular focus on the SNS, which, as mentioned before, is in charge of the body's response to a threat. Additionally, sensors can record data from subjects performing tasks to provide helpful information about the body's condition during the activity, such as working, sleeping, or driving. Therefore, organs controlled by the SNS are used to monitor this system's activity and stress response. However, each physiological sensor has its advantages and limitations, especially when the signals are measured during experiments where the subjects perform a specific task. These limitations must be overcome with other mathematical or statistical techniques dependent on the signal studied.

2.5 Sensors and Physiological Signals

2.5.1 Electrocardiogram

The Electrocardiogram (ECG) measures the electrical activity that passes through the heart, providing time to voltage information of the heartbeats. It is a meaningful, helpful, accurate, and non-invasive method to detect normal or abnormal heart rhythms that can indicate electrical abnormalities and heart problems. Commonly, cardiac biopotentials originate in the pacemaker cells of the sinoatrial (SA) node, located in the right atrium (RA). This stimulus is conducted over the RA and left atria (LA) working myocardium from the SA node, initiating atrial contraction. The excitation is then briefly delayed at the atrioventricular (AV) node, allowing the atrial blood to enter the ventricles. Next, the stimulus quickly sweeps into the left and right ventricles (LV and RV), spreading to the left and right ventricular muscle cells through the Purkinje fibers. Consequently, the atrial electrical activation and then ventricle activation lead to the contraction of these chambers, respectively, and the excitation can be observed by measuring the cardiac biopotentials as electrocardiograms (Figure 2.1) [44].

The electrical activation or stimulation, named above stimulus, is technically known as *depolarization*. A term derived since normal resting myocardial cells are polarized, which means they carry electrical charges on their surface. After depolarization, the return of the muscle cells to the resting state is called *repolarization*. The primary cellular processes

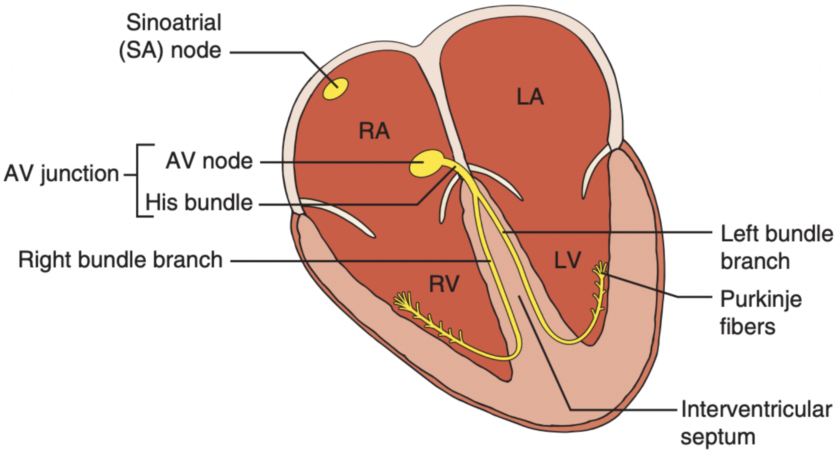

Figure 2.1: Cardiac Electrophysiology.

of depolarization and repolarization are responsible for the ECG waveforms, segments, and intervals Fig.2.2. The ECG records atrial depolarization as a P wave, ventricular depolarization as QRS complex, and ventricular repolarization as the ST segment, T wave, and U wave. Due to their low amplitudes, atrial repolarization segments and waves are not observed on a routine ECG. Thus, this P/QRS/ST-T/U sequence represents the cycle of the heart's electrical activity [45].

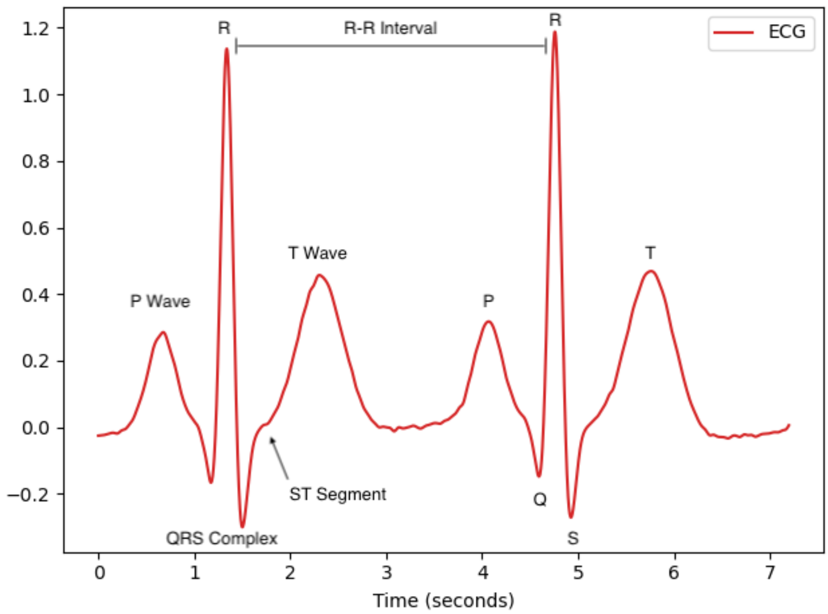

Figure 2.2: Electrocardiogram sample.

An electrocardiograph is a digital instrument that allows measuring, displaying, and analyzing these cardiac biopotentials and obtaining the ECG, measured as voltages on the body's skin surface through surface sensors called electrodes. Since voltage is the potential difference between two points, these cardiac biopotentials are measured between positive and reference electrodes.

2.5.2 Heart Rate and Heart rate variability

Usually, the ECG is a periodic signal where the time interval between two heartbeats is called inter-beat interval (IBI), and it can be calculated by observing the time interval between two consecutive R peaks by detecting the QRS complex, which is used to measure the heart rate (HR) and determine the heart rate variability (HRV). The HR is computed with the Formula 2.1, with beats per minute (bpm) as the unit.

$$HR(bpm) = \frac{1}{IBI(s)} \times 60 \tag{2.1}$$

HR is widely accepted as a noninvasive measure of the ANS regulation of the heart and is a standard for assessing stress and related psychological processes. Individuals experiencing stress typically report a rapid, pounding heartbeat and other physiological signs of stress, such as sweaty palms and rapid and shallow breathing. When a person is under stress, the time between each heartbeat is irregular, so HRV provides a vital tool to measure this irregularity for stress recognition [46]. Many different methods can evaluate HR variations; being time domain measures the simplest one. In time domain analysis of a continuous ECG record, QRS complex and the normal-to-normal (NN) intervals or the HR can be determined at any time. Time domain variables can be calculated from a series of instantaneous cycle intervals over long periods; however, the total variance of HRV increases with the length of the analyzed recording. The most commonly derived measures from interval differences are shown in Table 2.1

On the other side, Power spectral density (PSD) analysis provides the basic information on how power (variance) distributes as a function of frequency. Independent of the method used, only an estimate of the actual PSD of the signal can be obtained by proper mathematical algorithms. PSD calculation methods have advantages like simplicity of the algorithm used such as Fast Fourier Transform (FTT) in most cases, high processing speed, smooth spectral components that can be distinguished independently of preselected frequency bands, easy postprocessing with automatic calculation of low and high-frequency power components, and an accurate estimation of PSD even on a small number of samples [47]. From the frequency domain, the features that can be derived are shown in Table 2.1

2.5.3 Electrodermal Activity

Since the 1880s, psychological factors have been related to electrodermal phenomena; however, electrodermal activity (EDA) was first introduced by Johnson and Lubin in 1966 as a common term for the active and passive electrical phenomena traced back to the skin and its appendages [48]. Additionally, obtaining a distinct EDA response is relatively easy since its intensity is closely related to the stimulus intensity; as a result, EDA recordings have become one of the most frequent biosignals used in neurology and physiology. In the literature, the terms electrodermal response(EDR), electrodermal level (EDL), skin conductance activity (SCA), skin conductance response (SCR), and a lot more are also used to

Table 2.1: HRV Features

Symbol	Feature Description
Time Domain Features	
MeanNN	Mean of the RR intervals or mean heart rate
SDNN	Standard deviation of the RR intervals
RMSSD	Square root of the mean of the sum of successive differences between adjacent RR intervals
SDSD	Standard deviation of the successive differences between RR intervals
CVNN	SDNN divided by the MeanNN
CVSD	RMSSD divided by the MeanNN
MedianNN	Median of the absolute values of the successive differences between RR intervals
MadNN	Median absolute deviation of the RR intervals
MCVNN	MadNN divided by the MedianNN
IQRNN	The interquartile range (IQR) of the RR intervals
pNN50	Proportion of RR intervals greater than 50ms, out of the total number of RR intervals
pNN20	Proportion of RR intervals greater than 20ms, out of the total number of RR intervals
TINN	Baseline width of the RR intervals distribution obtained by triangular interpolation. An approximation of the RR interval distribution
HTI	The HRV triangular index, measuring the total number of RR intervals divided by the height of the RR intervals histogram
Frequency Domain Features	
ULF	Spectral power density of ultra-low frequency band (0-0.0033Hz)
LF	Spectral power density of low frequency band (0.04-0.15Hz)
HF	Spectral power density of high frequency band (0.15-0.4Hz)
VHF	Variability, or signal power, in very high frequency (0.4-0.5Hz)
LFn	Normalized LF, obtained by dividing the low frequency power by the total power
HFn	Normalized HF, obtained by dividing the low frequency power by the total power
LnHF	The log transformed HF

describe this phenomenon [49]. However, galvanic skin response (GSR) has more than one property since it can be described in terms of conductance, resistance, and electrophysiological potential. The sweat glands generate this biosignal, and the sweat is probably the origin of the variation in resistance and conductivity, although vasodilatation and constriction may also play an important role. This study will mention the term GSR to refer to these electrodermal recordings.

GSR sensors measure the electrical characteristics of the skin (eccrine sweat gland activity) using methods such as skin conductance response (SCR), skin potential (SP), skin conductance level (SCL), and skin potential response (SPR) [50]. It can capture the autonomic nerve responses as a parameter of the sweat gland function. The measurement is relatively simple and has good repeatability; the biosignal is obtained by attaching two leads to the skin and acquiring a base measure by passing a small current via electrodes positioned on a skin surface. As the activity is performed, recordings are made from the leads, and conductivity is measured between the two points. An example of the GSR signal is shown in Fig. 2.3. GSR must be measured in a part of the skin having many sweat glands, so the most common sites for GSR recordings are the fingers and palms, the bottom of the feet and the forehead, and a part of the skin with less or no sweat glands as the reference. GSR measures have been associated with physiological changes that accompany psychological processes underlying attention, emotion, and stress and are frequently used within mental stress testing paradigms [51]. Studies have linked GSR to arousal [52], emotions [53], frustration [54], and especially stress [55]. For that reason, many studies have used GSR recordings as an accurate method to investigate the stress response during specific activities [56].

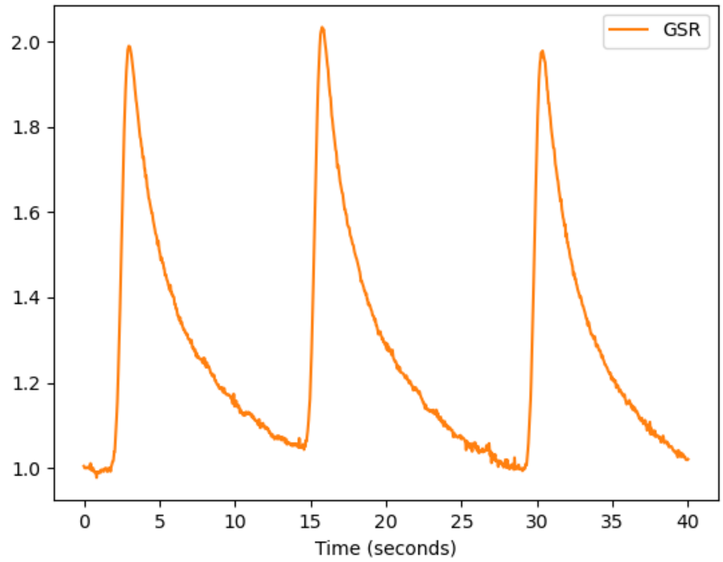

Figure 2.3: Galvanic Skin Response signal sample.

2.6 Machine Learning Classification Models

Medical data is intrinsically complex due to multiple and diverse parameters, including but not limited to quantitative test results such as HR, T, and respiratory rate, or analog outputs such as medical imaging, handwritten notes from the physician, and other diagnostics-related information. The growth of biomedical engineering and sciences, along with the rising medical challenges, has pushed the development of artificial intelligence (AI) in the medical field.

Machine Learning (ML) is a type of AI that make machines capable of decision making and actuation without being explicitly designed to do so [57]. As stated by Tom Mitchell in 1997 *"A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E"* [58]. Algorithms employed for a particular task look forward to improving the performance on similar tasks, gaining experience by repeating such tasks, and fine-tuning their parameters to improve their performance, improving the accuracy of output prediction. ML aims to develop programs that help machines access data and use it for learning a specific task on their own. This learning process is broadly categorized into four types: supervised, unsupervised, semi-supervised, and reinforcement. It is based on the training dataset and how it is inferred for the learning process.

Supervised Learning: when data is presented with labels. It is like giving standard answers to computers; based on this, computers will reply.

Unsupervised Learning: when the training data is non-classified and not labeled. Computers deduce a function to explain the hidden patterns from the unlabeled data

Semi-Supervised Learning: the combination of supervised and unsupervised learning. It is applied to data that is partially labeled.

Reinforcement Learning: the system interacts with its surroundings by action, and errors or rewards will be calculated. Trial and error search and delayed rewards are standard features. Computers and software programs can identify the ideal behavior in a specific context to increase the performance [59].

2.6.1 K-Nearest Neighbor

K-Nearest Neighbor (KNN) Algorithm can be used for classification and regression, where the input can be the same for both, while the output will depend on the task. KNN does not assume any underlying data distribution, which is called non-parametric. Classification identifies the class of an unknown instance based on the majority voting of its nearest neighbors.

Advantages: straightforward technique that is easily implemented. Building the model is cheap. It is a highly flexible classification scheme and well suited for multi-modal classes. Records are with multiple class labels. It can sometimes be the best method.

Disadvantages: classifying unknown records is relatively expensive. It requires distance computation of k-nearest neighbors. If the size of the training set increases, the algorithm gets computationally intensive. Noisy or irrelevant features will result in degradation of accuracy. It does not generalize the training data and keeps all of them. It handles large data sets and hence expensive calculations.

KNN can be used to diagnose multiple diseases with similar symptoms, handwriting detection, financial analysis, video recognition, and image recognition.

2.6.2 Support Vector Machine

Support Vector Machines (SVM) aim to classify objects correctly based on examples in the training data set. It classifies data by building an n-dimension hyperplane that maximizes the margin between two classes. SVM is highly correlated to ANNs. Generally, an SVM model with a sigmoid kernel function is the same as a 2-layer Perceptron NN.

Advantages: it can handle both semi-structured and structured data. It can handle complex functions if the appropriate kernel function can be derived. Not overly influenced by noisy data and not very prone to overfitting. It can scale up with high-dimensional data. It does not get stuck in local optima.

Disadvantages: performance goes down with extensive data set due to the increase in the training time. Finding the best model requires testing various combinations of kernels and model parameters. Does not work well with noisy datasets. Does not provide probability estimates. Results in a complex black box model difficult, if not impossible, to interpret.

SVM finds its practical application in cancer diagnosis, credit card fraud detection, handwriting recognition, face detection, and text classification. This model has gained popularity due to its high accuracy and high-profile wins in data mining competitions.

2.6.3 Decision Trees and Random Forest

A decision tree (DT) is a Supervised Machine Learning approach that continuously splits data based on a particular parameter. The decisions are in the leaves, and the data is split into the nodes. The data space division is performed iteratively until the leaf nodes hold a particular number of fewer records which can be utilized for classification purposes. In the Classification Tree, the decision variable is categorical, the outcome in the form of Yes/No.

Advantages: ease in interpretation, ease of handling categorical and quantitative values, the capability of filling missing values in attributes with the most probable value, and high performance due to the efficiency of the tree traversal algorithm.

Disadvantages: it can be unstable. It may not be easy to control the size of the tree. It may be prone to sampling error and gives a locally optimal solution instead of a globally optimal solution.

DTs can be used in applications like predicting tumor prognosis problems. However, this classifier might encounter the problem of over-fitting. Therefore, DTs' classification performance can significantly improve by growing an ensemble of trees and letting them vote for the most popular class, an ensemble known as Random Forest (RF). Different techniques can be used to grow the ensemble, including bootstrap aggregating or bagging, where each tree is built using a random selection of data and features, boosting, similar to bagging but in a sequential approach where data points that were misclassified in the previous tree have a higher chance to be selected to build the next tree, ex: AdaBoost, and other DT models using subsets. It selects a subset of features from an individual tree node, avoiding correlation in the bootstrapped set. For a classification task, RF is a forest of k trees and is computed as in Formula 2.2

$$RF = DTj, \quad where \quad j = 1, k \tag{2.2}$$

2.6.4 Bayesian Networks and Naïve Bayes

Bayesian Networks (BN) is based on the idea that the estimated likelihood of an event should be based on the available evidence across multiple trials, so a prior probability distribution is selected and then updated to obtain a posterior distribution. Later on, with the availability of new observations, the previous posterior distribution can be used as a prior. Therefore, the relationship between different dependent variables, ex. Using Bayes' theorem, stress and HR can be described (Formula 2.3).

$$P(A|B) = \frac{P(A) \times P(B|A)}{P(B)} \tag{2.3}$$

This theorem states that the probability of event A to occur given that event B has occurred equals the proportion of trials in which A occurred together with B out of all trials in which B occurred. For the stress example, this means that the probability of having stress (A) while having a high HR (B) equals the proportion of trials in which subjects had high stress while having a high HR out of all trials in which subjects had a high HR.

Advantages: incomplete datasets can be handled. It can prevent over-fitting of data. There is no need to remove contradictions from the data.

Disadvantages: selection of prior is difficult. The posterior distribution can be influenced by prior to a great extent. If the prior selected is incorrect, it will lead to wrong predictions. It can be computationally intensive.

The Naïve Bayes (NB) algorithm is simple, based on conditional probability, and the most common application of the Bayes' theorem. It uses a probability table updated through training data. The table is based on its feature values, where it is needed to look up the class probabilities for predicting a new observation. The basic assumption is conditional independence and features equally essential, which is not the case in most real-world problems, so it is called *naïve*.

Advantages: easy implementation, works with less training data. Handles continuous and discrete data, can handle binary and multi-class classification problems, and make probabilistic predictions.

Disadvantages: adequately trained and tuned models often outperform NB models as they are too simple if there is a need to have one feature as "continuous variable" ex. Time, it is challenging to apply NB directly. There is no true online variant for NB, so all data must be kept to retrain the model. It will not scale when the number of classes is too high. It takes more runtime memory compared to SVM or simple logistic regression. Computationally intensive, especially for models involving many variables.

2.6.5 Artificial Neural Networks

The Artificial Neural Network (ANN) is a computational model inspired by the structure of the human brain. The human brain is composed of several nerve cells called neurons. ANN contains numerous neurons, and each neuron is a fundamental unit. ANNs use a divide-and-conquer strategy to learn a function: each neuron in the network learns a simple function, and the overall (more complex) function, defined by the network, is created by combining these more specific functions. Deep learning networks are ANNs that have many hidden layers of neurons. The minimum number of hidden layers necessary to be considered deep is two. However, most deep learning networks have many more than two hidden layers, and the depth of a network is measured in terms of the number of hidden layers plus the output layer. The DL networks generally comprise three main layers: input layer, hidden layer/layers, and output layer (Figure 2.4). As the number of layers increases, the network becomes more complex or more profound.

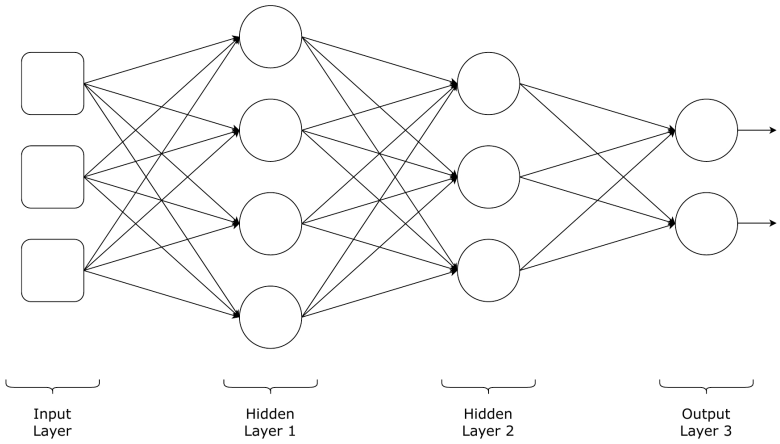

Figure 2.4: Illustration of a simple Neural Network architecture.

The depicted network has four layers: one input layer, two hidden layers, and one output layer. A hidden layer is just a layer that is neither the input nor the output layer. The squares in the input layer represent locations in memory that are used to present inputs to the network. These locations can be thought of as sensing neurons. There is no processing of information in these sensing neurons; each output is simply the value of the data stored at the memory location.

The circles represent the processing neurons in the network. Each neuron takes numeric values and maps them to a single output value. Every input to a processing neuron is either the output of a sensing neuron or the output of another processing neuron. Arrows illustrate how information flows through the network from the output of one neuron to the input of another neuron. Each connection in a network connects two neurons, and each connection is directed, which means that information carried along a connection only flows in one direction. Each of the connections in a network has a weight associated with it. A connection weight is simply a number, but these weights are critical. The weight of a connection affects how a neuron processes the information it receives along the connection. Training an artificial neural network essentially involves searching for the optimal set of weights [60].

Advantages: among the most accurate modeling approaches. Makes few assumptions about the data's underlying relationships

Disadvantages: computationally intensive and slow to train. Easy to overfit or underfit training data. Results in a complex black-box model.

Chapter 3

State of the Art

3.1 Stress Recognition Signals

Different physiological signals have been used in the classification of stress, including galvanic skin response (GSR), electromyogram (EMG), electrocardiogram (ECG), photoplethysmogram (PPG), phonocardiogram (PCG), among others. As shown in Table 3.1 many studies have focused on detecting stress using different measurements, specially biosignals.

Table 3.1: Physiological Signals used for stress detection

	Signal	Study	Year
[61]	ECG	Cardiac stress detection using a portable ECG	2020
[62]	PPG	Detection of five levels of stress using a PPG sensor	2020
[63]	PCG	Psychological stress detection using time duration of cardiac cycles of PCG	2019
[64]	EEG	Assessment of dispersion patterns for distress detection	2021
[65]	GSR	Automatic detection of car drivers' stress levels	2018
[66]	EMG	Acute stress detection measuring gastrointestinal activity	2021
[67]	HRV	Psychological and physical stress detection using a wearable device	2021

However, based on a search through the state-of-the-art studies focused on stress detection, whether psychological or physical, and physiological signals, it has been noticed that the most used biosignals for this purpose are ECG, GSR, and EMG. Additionally, many studies use a multivariable methodology where authors use more than one signal simultaneously to detect stress in different environments accurately [68, 69]. However, it leads to a more complex methodology and hardware to obtain all of them simultaneously.

3.2 Signal Processing and Feature Extraction

Digital signal processing (DSP) is an advancing technology that has dramatically impacted today's lives with digital audio, video, the internet, medical devices, and instruments.

Digital ECG has made it possible to provide precise heart diagnoses and powerful tools for scientists and engineers to analyze and visualize this data. DSP consists of an analog filter, an analog-to-digital conversion (ADC) unit, a digital signal (DS) processor, a digital-to-analog conversion (DAC) unit, and a reconstruction (anti-image) filter as shown in Figure 3.1. While analog signal processing does not require software, algorithm, ADC, and DAC, DSP systems use software, digital processing, and algorithms. Therefore, the latter systems present more flexibility, less noise interference, and no signal distortion in various applications.

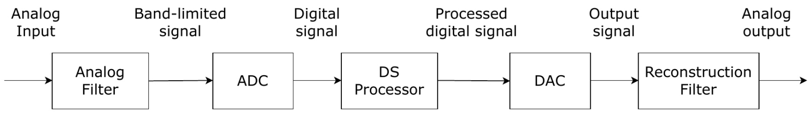

Figure 3.1: Digital Signal Processing Scheme.

In the time domain, the representation of digital signals describes the signal amplitude vs. the sampling time instant or the sample number. However, for some applications, such as signal filtering, it is required to analyze time domain information and the signal's frequency content, known as signal spectrum. ECG signals acquire different kinds of high and low-frequency noises generated mainly at recording or transmission. Digitized noisy signals can be enhanced using digital filtering. Since the valuable signal contains low-frequency components, all high-frequency components above the cut-off frequency of the signal are considered noise that can be removed using digital filters. Therefore, one of the most researched topics in ECG analysis is the elimination of these noises using different digital filtering techniques. A variety of methods have been proposed in the literature for biological signal preprocessing, including finite impulse response (FIR) filter, infinite impulse response (IIR) filter [70], discrete wavelet transform (DWT) [71], among others, in order to eliminate specific noises that can deteriorate the quality of the signal, such as baseline wander (BW) [72], powerline interference (PLI) [73], muscle artifact (MA) [74] or additive white gaussian noise (AWGN) [75].

The importance of this noise elimination relies on its facilitation for accurate diagnosis of patients when analyzing the ECG signals. For frequency ECG filtering, the FIR filter is stable, and the signals are not distorted, but it requires substantial computation that must be taken into account based on the purpose of the study. On the other hand, the IIR filter has a much smaller order number that is more computationally efficient than the FIR filter [76]. However, the IIR filter has a non-linear phase response that may distort the ECG signal, although it can be avoided using a non-causal forward-backward approach (bidirectional filter). A min-max normalization is usually applied to the ECG signal to remove subject-specific physiological signal baseline and lifestyle factors influencing it. The normalized signals are also passed to a Butterworth band-pass filter (5–15 Hz) to reduce muscle noise, and baseline wanders [77], a suitable preprocessing method for later classification.

The Pan-Tompkins peak detection algorithm developed in 1985 marks the characteristic points of the R peak on the preprocessed ECG signal to obtain the RR interval. It is based on digital filtering, and Fourier Transform, which is resistant enough to noise with a commendable beat-detection accuracy of 99.37% in determining peaks of a highly noisy ECG signal [78]. This algorithm is considered a pioneer in the field; however, it is still used for peak localization in ECG studies due to its high accuracy [79]. The accurate detection of Q, R, S, and T waves is necessary to extract features dependent on them in the time and frequency domain. Therefore, this algorithm is helpful for accurately extracting statistical features such as max, min, median, standard deviation, mean of HRV, and frequency domain features [80].

3.3 Stress Classification Methods

Several stress recognition techniques have been explored in the literature, where classification techniques and ML models are widely employed to obtain good results in different metrics, including accuracy. However, the metrics of stress detection, obtained with varying classification techniques, may be difficult to compare since protocols or physiological signals used might differ across the literature. Table 3.2 presents an overview of the past studies; it is based on an online search taking into account the last five years and including research studies focused on stress detection using ML models. As exposed, different ML algorithms are employed, including sequential minimal optimization (SMO), Multilayer perceptron (MLP), stochastic gradient descent (SGD), linear discriminant analysis (LDA) and logistic regression (LR).

Table 3.2: ML techniques used for the classfication of stress

	ML Model	Data	Study	Best Accuracy
[81]	LDA, KNN, SVM	E-nose, GSR	Academic Stress detection in university students	SVM with 96%
[82]	DT, KNN, SVM	T, BP, Pulse	Stress in cardiac patients to predict the severity of the cardiac disease	KNN with 89.06%
[83]	LR, NB, RF, SVM	PSS	Mental stress detection in university students	SVM with 85.71%
[84]	ANN	ECG, GSR, T, BP	Psychosocial stress in patients with metabolic syndrome	Average 92%
[85]	SVM	GSR, T, BVP, HR, ACC	Psychological and physical stress using wristband biosignals	During physical state: 99.19%
[86]	DT, RF, SVM, NB	DASS21	Predicting Anxiety, Depression and Stress in Modern Life	NB 85,5%
[87]	KNN, DT, RF, SVM, ANN	ECG, EMG, GSR, RESP	Stress prediction in automobile drivers during actual driving	RF with 98.92%
[88]	DT, LDA, LR, NB, SVM, KNN	ACC, GSR, T, HR	Detection of stress in patients during treatment for substance use disorder	SVM with 81,3%
[89]	DT	Images	Thermal analysis of face and fingers to detect academic stress	Accuracy 91%
[63]	LS-SVM	PCG	Psychological stress detection methodology suitable for telemedicine	93.14%
[90]	KNN, GDA, SVM, DT, LR	EEG, Cortisol	Workers' stress recognition at construction sites	SVM with 80.32%
[91]	SVM	EEG, ECG	Early mental-stress detection and predictions for treatment efficacy	79.54%
[77]	CNN, LSTM	ECG	Automatic driver stress level classification	Average: 92.8%
[92]	KNN, RF, SVM, DT, MLP	EEG, GSR, PPG	Stress classification during public speaking	SVM with 96.25%
[93]	SVM, KNN, RF	ECG	Assesses the effect of psychological stress on HRV features	SVM with 97%
[94]	SMO, SGD, LR, MLP	EEG	Stress classification in response to music tracks	LR with 98.76%
[95]	LR, SVM, RF, KNN, MLP	EEG	Multi-level stress classification with and without smoothing filter	MLP with 81%
[96]	CNN, DNN, SVM, RF	HR	Stress prediction using functional near-infrared spectroscopy fNIRS	CNN with 98.69%
[97]	KNN, NB, RF, AB	HR	Stress recognition focused on desk jobs in its initial stages	RF with 83%
[98]	DNN	ECG	Mental Stress monitoring using ECG Signals	87.39%
[99]	SVM, KNN, RF, ANN	EEG	Real time stress analysis using wearable EEG	SVM with 97.5%
[90]	KNN, DA, SVM	EEG	Workers' stress recognition at construction sites	SVM with 80.32%
[100]	DNN	EEG	Stress recognition on workers at construction sites	86.62%
[101]	KNN, ANN, NB, SVM	PPG, GSR	Stress monitoring using wearable sensors	KNN with 85.3%

As shown in Table 3.2, the most used signal is ECG, which is present in 6 studies. Some studies that do not employ ECG may use other signal sources such as PPG, PCG, or

HR to perform HRV analysis and extract their features for stress detection. However, other physiological signals such as BP, T, EEG, and GSR are also widely used for this classification task. Many studies use more than one physiological signal or stress detection method. For example, the studies [84, 85, 87, 88, 92] use at least three physiological signals, including motion acceleration in some of them. Data from these studies were obtained using multi-parameter systems created with different sensors.

Other stress detection methods, such as questionnaires, were employed. The PSS [83] questionnaire to analyze academic stress, and the [86] study used the DASS21 questionnaire to predict anxiety, depression, and stress in modern life; both studies made use of the questionnaires data to feed ML models. The physiological stress indicator cortisol was also used with an EEG signal to detect stress on workers using different ML models. Other particular measures have been employed, such as in the pilot study [81], where the authors measured GSR signals and used an electronic nose system to measure volatile organic compounds emitted from the skin. Besides being a methodology not extensively used, the data obtained with the proposed system was successfully used to detect academic stress in engineering university students using three ML algorithms.

In the same way, Resendiz et al. [89] use another novel system based on infrared thermography and thermal analysis of facial skin and fingertips. The images served as input for a DT to diagnose human stress in undergraduate university students. Moreover, fNIRS measures have also been used to derive HR and predict stress using various ML models [96].

Moreover, it is observed that most studies aim to detect stress in healthy people. Usually, students, workers, or healthy people in their daily lives are the subjects of study for the obtention of the data. Some experiments have obtained the data in real-life conditions or subjects' normal activities such as studying, working, or driving. In contrast, other investigations have induced stress through images, music, or the mentioned TSST for a stress measure. Therefore, psychological stress has been addressed in studies focused on subjects performing analytical tasks like public speaking and physical stress in studies where people perform physical activities like working.

On the other side, among the most used ML models, we can analyze SVM, KNN, ANN, RF, and DT are the most used for this stress detection purposes. Figure 3.2 illustrates the number of times each model was used based on the search described in Table 3.2, where ANN refers to deep, recurrent, and convolutional networks. Several ML algorithms are tested in each study, obtaining at least one model outperforming the others with a high accuracy level. Since different studies conclude that a specific ML technique provides the best accuracy, it is unclear which one would be the most appropriate for this study's stress detection objective, considering the difference in data and features. Some studies use data from public databases, while others get signals from laboratory experiments or real-life tasks. Therefore, these parameters can be considered for the outcome of the model classification. On the other side, as mentioned before, techniques such as SVMs and ANNs are black-box models and do not provide insight into how the link between physiological signals and stress is established.

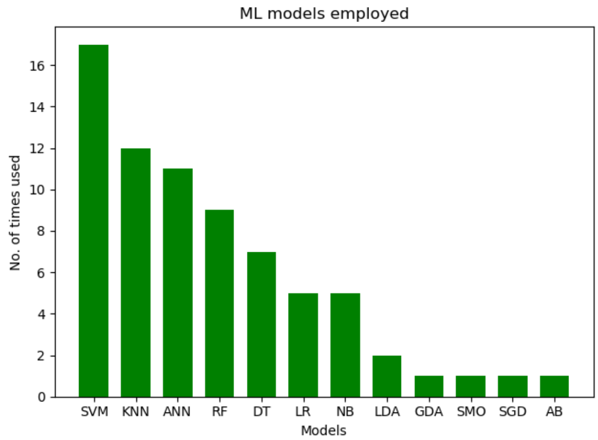

Figure 3.2: ML models used in the literature.

Year	Signals	ML model	Study	Results
2018	EEG, Cortisol	KNN, GDA, SVM, DT, LR	Workers' stress recognition at construction sites	SVM with 80.32%
2018	EEG, ECG	SVM	Early mental-stress detection and predictions for treatment efficacy	79.54%
2018	HR	KNN, NB, RF, AB	Stress recognition focused on desk jobs in its initial stages	RF with 83%
2019	ECG, EMG, GSR, RESP	KNN, DT, RF, SVM, ANN	Stress prediction in automobile drivers during actual driving	RF with 98.92%
2020	ACC, GSR, T, HR	DT, LDA, LR, NB, SVM, KNN	Detection of stress in patients during treatment for substance use disorder	SVM with 81,3%
2021	GSR, T, BVP, HR, ACC	SVM	Psychological and physical stress using wristband biosignals	During physical state: 99.19%
2022	ECG	DT, KNN, RF, SVM, MLP	Stress detection using physiological signals	RF with 82%

Chapter 4

Methodology

The programming language used for this study is Python, a high-level, general-purpose programming language widely used recently. Python has language constructs and an object-oriented approach to help programmers write clear, logical code for small- and large-scale projects. It is a simple programing language with a design philosophy emphasizing code readability, and its syntax allows programmers to express concepts in fewer lines of code than would be possible in languages such as C [102]. It has many libraries, such as Numpy, Pandas, Scipy, Scikilearn, Matplotlib, and Tensorflow, which have different functions and will be helpful for different techniques and steps in this experiment.

The methodology of this study is summarized in Figure 4.1, where signals preprocessing and feature extraction have been developed on Spyder IDE (Integrated Development Environment), a free and open-source scientific environment written in Python for Python, and designed by and for scientists, engineers and data analysts, while the other steps have been developed in Google Colaboratory (colab), a hosted version of Jupyter Notebooks with access to Google hardware, where Notebooks are run in Linux-based virtual machines (VMs) provided and maintained by Google where computation can be performed with central processing units (CPUs) or accelerated through specialized graphical processing units (GPUs), and tensor processing units (TPUs) that allows to run and share modern AI and ML techniques [103].

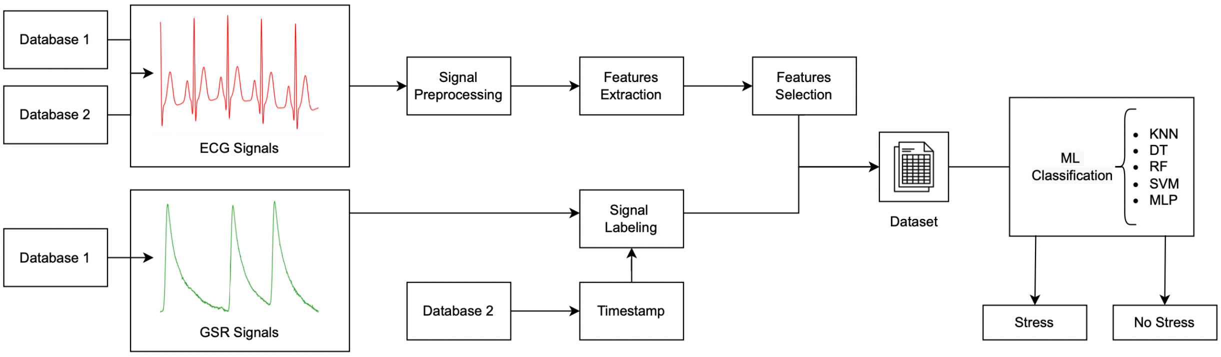

Figure 4.1: Block diagram of the proposed stress classification methodology.

4.1 Physiological Data

The biosignals used for this research are pre-recorded ECG signals. The publicly available datasets chosen for this study are the "Stress Recognition in Automobile Drivers" (SRAD) database from the MIT Dept. of Electrical Engineering and Computer Science [104], and WESAD, a Multimodal Dataset for Wearable Stress and Affect Detection [105]. Databases' details are listed in Table 4.1, where acc refers to acceleration in each axis.

Table 4.1: Databases Details

Database	No. of subjects	No. of signals	Signals	Recordings duration	Stress Phases
SRAD	17	6	ECG, EMG, footGSR handGSR, HR, RESP	65 to 93 minutes	Common driving tracts Stressful driving tracts
WESAD	15	8	ECG, EDA, EMG RESP, BT Xacc, Yacc, Zacc	36 minutes	Neutral Stress Amusement

Each database has been studied for the detection of stress using different ML models [105, 104]. However, in this thesis project, the ECG signals of both databases have been processed and joined to acquire more data that serves the classification purpose.

4.1.1 "Stress Recognition in Automobile Drivers" Database

The SRAD is a database obtained in the work of Healey JA and Picard RW for detecting stress during real-world driving tasks [104]. The study collected physiological signals from healthy subjects during a prescribed driving route, including streets and highways around Boston, Massachusetts. The original database generated is a large dataset of physiological signals containing many day-to-day variations covering over 36 hours of driving. However, a smaller version of this database is publicly available on physionet.org, a repository of freely-available medical research data managed by the MIT Laboratory for Computational Physiology and supported by the National Institute of Biomedical Imaging and Bioengineering (NIBIB) [106].

The public database employed in this thesis project contains 17 recordings with a duration of 65 to 93 minutes. The driving protocol and data collection can be found in the original study. An example of the signals obtained can be seen in Figure 4.2, which includes the marker to differentiate the driving tracks.

The Healey and Picard study indicated that GSR and HR signals were significantly correlated to drivers' stress. Additionally, the literature has closely related GSR signals to stress at different levels [50]. Therefore, GSR is used to provide the stress labels for the signals of the SRAD database, as high GSR values have been related to the presence of stress while low values represent no stress.

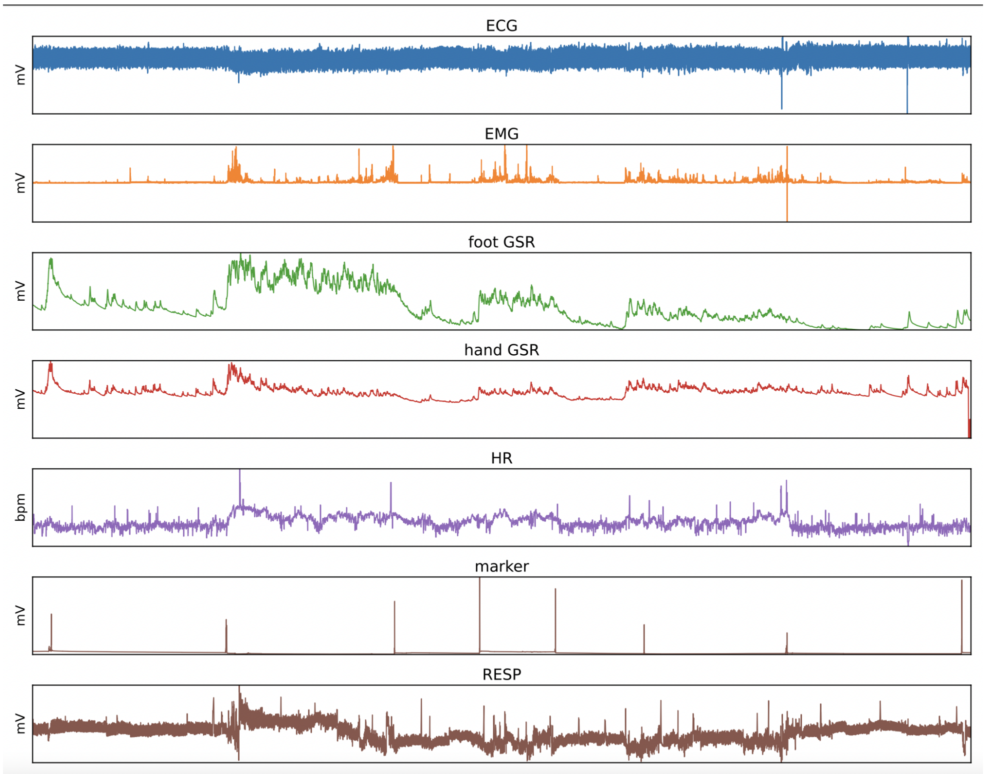

Figure 4.2: Signals obtained from the "Stress Recognition in Automobile Drivers" database

4.1.2 WESAD Dataset

The WESAD is a multimodal, publicly available dataset that includes physiological and motion modalities (Table 4.1). The authors state that this dataset bridges the gap between previous lab studies on stress and emotions by containing three different affective states (neutral, stress, amusement) obtained by working with a study protocol, including a baseline, amusement, and stress condition. Additionally, guided meditations were employed to de-excite subjects after stress and amusement. Data were collected from 15 graduate students using a chest-worn device. Figure 4.3 illustrates a sample of the eight signals.

In this case, WESAD ECG signals were labeled based on the timestamp information provided by the authors, where two stress levels were detected. However, the stress and amusement phases are set as stress periods for this binary classification.

4.2 Signal Preprocessing

The ECG signals from both databases are raw signals containing noises and corruptions. Therefore, the first step is to clean them. For reading the signals from the Physionet databank, the WFDB library for python has been used, allowing reading, annotating, and performing other signal techniques. Each experiment comes with a 'fields' property where the vital information about each signal is saved, such as signal name, units, and the sampling frequency. On the other side, the WESAD data come in a .txt format listing the eight signals sampled at 700Hz. The NeuroKit2 library has been used for signal cleaning, an

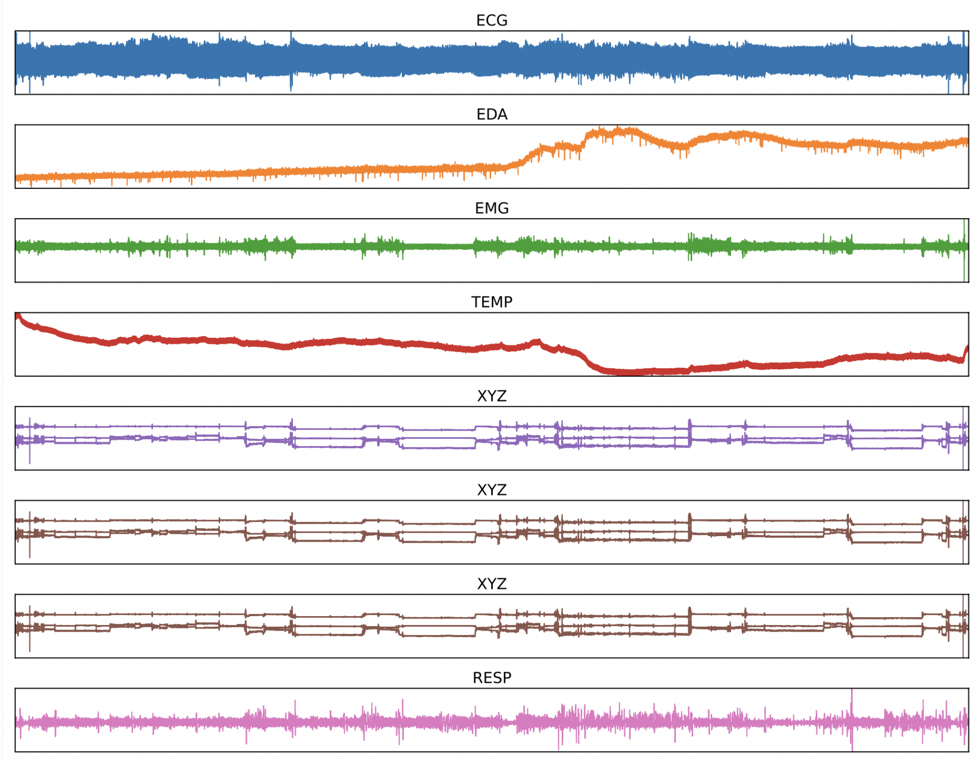

Figure 4.3: Signals obtained from the WESAD dataset

open-source, community-driven, and user-centered Python package for neurophysiological signal processing. It provides comprehensive processing routines for various physiological signals such as ECG, PPG, EDA, EMG, and RSP. These processing routines include high-level functions that enable data processing using validated pipelines in a few lines of code.

ECG signals are checked for missing data points; if any, the missing values are filled using the forward-filling method. Then the signal is cleaned with two filters. A high-pass Butterworth filter removes the baseline drift with a cut-off frequency of 0.5Hz. Conversely, a notch filter of 50Hz cut-off frequency is also applied to remove the powerline interference. Once the signals are cleaned, these can be used to extract features correctly.

4.3 Feature Extraction

Most of the clinically useful information of an ECG is found in the different intervals and amplitudes of the waves since abnormal variations can be an indicator of heart problems. Feature extraction reduces the available information but maintains the ECG morphology. In order to identify the stress level, firstly, many features must be extracted from the related physiological signals. Therefore, accurate ECG feature extraction is of significant importance.

ECG signals were segmented using a window sliding technique with a window shift of 1 second. All features were computed using a cleaned ECG signal window of 30 seconds, considering each frequency sample. For each ECG segment, R-peaks are found using the

Pan Tompkins algorithm. The HRV features are obtained in the time and the frequency domain (Figure 4.4).

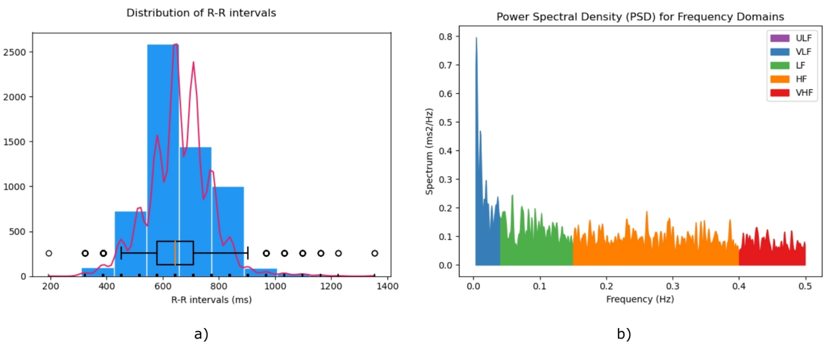

Figure 4.4: HRV features in the a) Time and b) Frequency domain.

For the SRAD database, GSR signals are segmented similarly to ECG signals, considering each sampling frequency. The mean value of each GSR signal is compared to the windows and labeled as stress (1) when the value is higher than the mean and no stress (0) when it is less than the mean. A schema of this 30s window sliding process is seen in Figure 4.5. The WESAD database does not require this since labels are obtained from the timestamp.

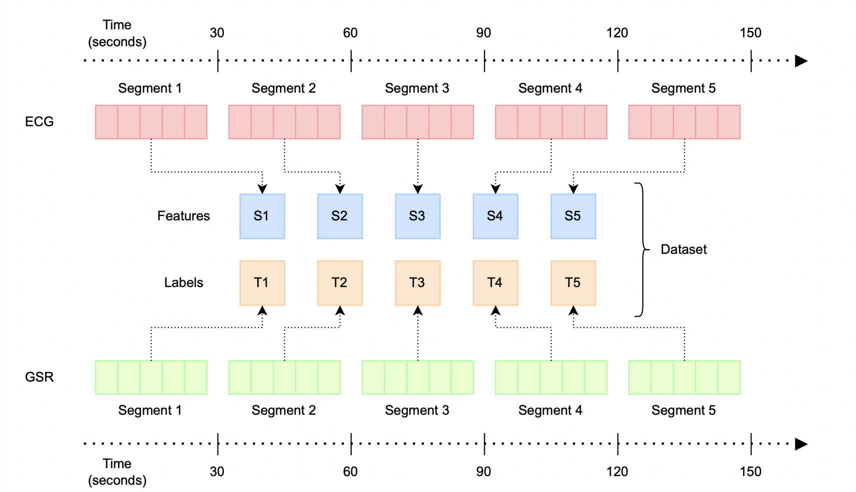

Figure 4.5: Signals segmentation for obtaining the labeled dataset.

Once the feature extraction is performed, normalization is another critical step before continuing the proposed methodology. *Data normalization* is a fundamental approach that transforms features in a standard range to avoid greater numeric feature values dominating smaller numeric feature values. ML models' success depends on the data quality, and normalization is a critical way to improve this quality and the performance of machine learning algorithms. Z-score Normalization (ZSN) is a mean and standard deviation-based normalization method where raw data's statistical mean and standard deviation are used to normalize the data. Each instance (x) of the data is transformed into (x') by the

Formula 4.1 where μ denotes the mean and σ the standard deviation of the ith feature. This normalization method has proven to be more suitable than other methods [107].

$$x'_{i,n} = \frac{x_{i,n} - \mu_i}{\sigma_i} \tag{4.1}$$

4.4 Feature Selection

We need to obtain specific features that work as predictors to predict an outcome adequately, given the underlying information they provide. However, this set of features may contain non-informative variables that can impact the model's performance. Feature selection methods are intended to reduce the number of input variables by removing non-informative or redundant predictors from the model and leaving those believed to be most beneficial for a model to predict the target variable, in this case, stress. These techniques have become a necessity since the more the features are, the more computation cost should be expensed, the more amount of system memory, the more sensors users have to wear, which brings uncomfortably to them, and the performance of some models can be degraded when including input variables that are not relevant to the target variable. Moreover, feature selection is essential to avoid overfitting, improve the model performance, provide faster and more cost-effective models, and gain a deeper insight into the underlying processes that generated the data. Therefore, the objective of making feature selection in this study is to reduce and select as least as possible the number of predictors (features) as far as possible without compromising the predictive performance of stress.

The Recursive Feature Elimination (RFE) is a backward selection of the predictors. It builds a model on the entire set of predictors and computes an importance score for each predictor. The least essential predictors are removed, the model is rebuilt, and importance scores are computed again. The number of predictors to evaluate must be specified, so it is a tuning parameter for RFE. A chosen 'estimator' or chosen ML model must be employed to define the method to use this technique. RFE can be used with different LR, LDA, or RF models. The latter is frequently used because it is a model that does not exclude variables from the prediction equation, and this model has a well-known internal method for measuring feature importance. This RF-RFE approach has been evaluated with great results [108]. Therefore, in this study, the essential scores calculated through an RF model will explain the number of predictors to use in the RFE. Therefore, tests are performed with different tuning parameters to analyze the number of predictors needed to provide similar accuracy to the model when employing all of them.

We may also be interested in the correlation between input variables with the output variable in order to provide insight into which variables may or may not be relevant as input for developing a model. Therefore, we can use other correlation methods. Pearson's Correlation (PC) is a statistical method that calculates the correlation coefficient between two variables. In this case, it determines the correlation between each feature and stress. PC values range between -1 and 1. A correlation can be positive, meaning both variables move in the same direction, or negative, meaning that when one variable's value increases, the other variables' values decrease. Correlation can also be neutral or zero, meaning the variables are unrelated. Therefore, values close to 0 mean a low correlation, close to 1 mean a high positive correlation, and close to -1 mean a high negative correlation. This study removes features that maintain a correlation higher than 0.9.

PC is performed using SciPy, an open-source scientific computing library for the Python programming language. In contrast, RFE is performed using Scikit-learn, an open-source machine learning library that supports supervised and unsupervised learning and provides various tools for model fitting, data preprocessing, model selection, evaluation, and many other utilities.

4.5 Stress Classification

Given the literature on ECG signals classification for stress detection, we use the most used models to perform binary classification. KNN, RF, DT, SVM, and ANN models are used as classifiers, trained, and tested with the HRV features selected.

The KKN model classifies the data based on the k distinct but closest neighbor type by exploring them and classifying the majority class of k neighbors. This classifier can detect linear or non-linear distributed data and performs very well, even with many data points, since it is automatically non-linear. It has been used for the classification of stress accomplishing high accuracy levels [82], where the three signals T, BP, and pulse obtained from cardiac patients were employed to feed a DT, a KNN, and an SVM model. The KNN outperformed the other two, concluding it is an accurate classifier to detect stress.

The RF model has been evaluated before for stress detection [83], using the results of a PSS questionnaire as the input data for four ML models: LR, NB, RF, and SVM. Besides the RF's acceptable accuracy, the SVM outperformed the other algorithms in this academic stress analysis. In the study of Sanchez et al. [97], ECG, EMG, GSR, and RESP signals comprise the dataset. Five ML algorithms were employed for employee stress classification, and an RF model achieved the highest accuracy compared to other classification models.

The SVM model is instrumental in solving high dimensionality feature space problems. The classification of the data is based on the hyper-plane. This classifier is suited for both linear and non-linear data classifications. Based on Table 3.2, It has been the mostly employed model for different studies [93, 92, 81],and sometimes outperforming other classification algorithms.

Finally, Artificial Neural Networks (ANN) have been widely studied for different classification tasks, including stress classification. ANN [84], DNN [98], CNN [77], and MLP [92] are examples of networks that have successfully classified stress.

Consequently, it is enough information to choose the mentioned ML algorithms as classifiers for this stress detection task using HRV features. Each classifier possesses hyperparameters tuned based on the literature and experiments to obtain an accurate classification. The python libraries Scikit-learn and Keras are used for the classification model development. Keras is a deep learning API running on top of TensorFlow's machine learning platform.

It was developed to enable fast experimentation, which allows going from idea to result as fast as possible, the key to doing good research.

4.5.1 Model Performance

Four metrics are implemented to measure the model performance: accuracy, recall, precision, and F1 Score.

The most widely-used and straightforward measure is classification accuracy, mathematically defined by the ratio of the total number of correct predictions and the total number of predictions, as shown in Formula 4.2. True positive (TP) is an outcome where the model correctly predicts the positive class. True negative (TN) is an outcome where the model correctly predicts the negative class. On the other side, false positive (FP) is an outcome where the model incorrectly predicts the positive class, and false negative (FN) is an outcome where the model incorrectly predicts the negative class.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{4.2}$$

Accuracy is used to evaluate and fine-tune the objective function of each model. However, accuracy is sometimes not enough to represent the overall performance because accuracy can be high while other metrics differ.

Precision is a metric that quantifies the number of cases the model correctly identified as positive predictions (TP), out of all the cases classified as positive. It is calculated as the ratio of correctly predicted positive examples divided by the total number of predicted positive examples (Formula 4.3). It is also called the Positive Predictive Value (PPV). High precision is the aim when the objective is to minimize false positives.

$$Precision = \frac{TP}{TP + FP} \tag{4.3}$$

Recall is another performance metric that measures the number of correct positive predictions (TP) the model correctly identified out of all positive predictions that could have been made. For binary classification, it is calculated as the number of true positives divided by the total number of true positives and false negatives (Formula 4.4). It is also called Sensitivity or the True Positive Rate. Unlike precision, which only comments on the correct positive predictions out of all positive predictions, recall indicates missed positive predictions. In this way, recall provides some notion of the coverage of the positive class. A high recall is aimed when the objective is to minimize false negatives.

$$Recall = \frac{TP}{TP + FN} \tag{4.4}$$

Using accuracy as the defining metric is intuitive, but it is always advisable to use precision and recall. There can be situations where accuracy is very high, but precision or recall is low. Avoiding situations where the subject presents stress but the model classifies it as no stress (FN) is better. On the other hand, when the subject is not suffering stress and the model predicts the opposite (FP), the objective is to avoid treating the subject with no stress. Maximizing precision will minimize the number of FP, whereas maximizing the recall will minimize the number of FN. Although high precision and high recall values are aimed, achieving both simultaneously is hard. For example, if the model is tuned to give a high recall, all the subjects who actually present stress will be detected, but many subjects who do not suffer from it will treat a nonpresent problem. Similarly, high precision is aimed to avoid giving any wrong and unrequired treatment, but many subjects who actually have stress problems end up without any treatment.

Both metrics are important in medical diagnosis because subjects incorrectly classified as suffering from stress are equally important since they could indicate some other problems, so the aim is not only for a high recall but also for high precision. Therefore, we need to balance these two metrics, conveyed with an F1 Score. F1-score is the harmonic mean of the precision and recall (Formula 4.5). This metric is easier to work with since the F1 Score can be improved to indicate a good precision and a good recall value instead of improving precision or recall.

$$F1score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4.5}$$

4.6 Experimental Setup

One aim of this study is to provide a methodology for stress detection that can be replied in a real-life scenario and outside laboratories in an environment not necessarily strictly controlled, such as a patient home. Therefore, an experiment was set to obtain raw ECG signals that can be classified for stress detection by following the proposed protocol.

The Single Lead Heart Rate Monitor AD8232, a cost-effective board used to measure the heart's electrical activity, is used for signal obtention. Besides, this module can extract, amplify, and filter small biopotential signals. The Bio Protech ECG electrodes called "telectrodes" are employed for this test. It uses an Ag/AgCl sensing element and hydrogel for adhesion which are excellent components for sensitive monitoring. For the signals obtained with the module, another preprocessing step was applied to these ECG signals. The module was powered using an Arduino Uno, a versatile microcontroller equipped with the well-known ATmega328P and the ATMega 16U2 Processor (Figure 4.6. Signals were recorded using the open-source Arduino software and saved into .csv files for their posterior cleaning.

The experiment was performed on a healthy 25-year-old male subject seated in a relaxed phase focused on deep breathing. No distractions were performed in order to achieve a non-stress phase.

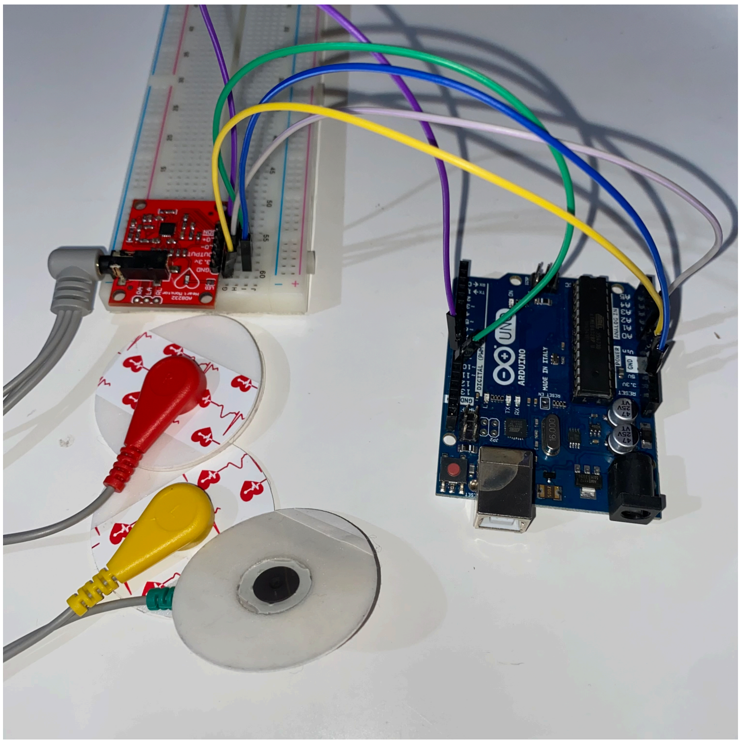

Figure 4.6: ECG experimental setup.

In this case, since the raw signals contain noises, different filters such as high pass, low pass, notch, and combinations of them have been applied to the signal to analyze their behavior and select the filter that cleans the signals the most, considering 0.4Hz as the high cutoff frequency, 60Hz as the low cutoff, and placing the powerline at 60Hz. After successfully cleaning the signals, the following steps of the methodology proposed in Figure 4.1 have been applied to obtain the HRV features and to test the classification protocol using the selected ML models.

Chapter 5

Results

5.1 Signals Preprocessing

The ECG signals were cleaned following the process mentioned in chapter 4, with the two filters. Figure 5.1 illustrates the ECG signal before and after cleaning. A decrease in the amplitude of the signal can be observed, especially in some noisy peaks after filters are applied.

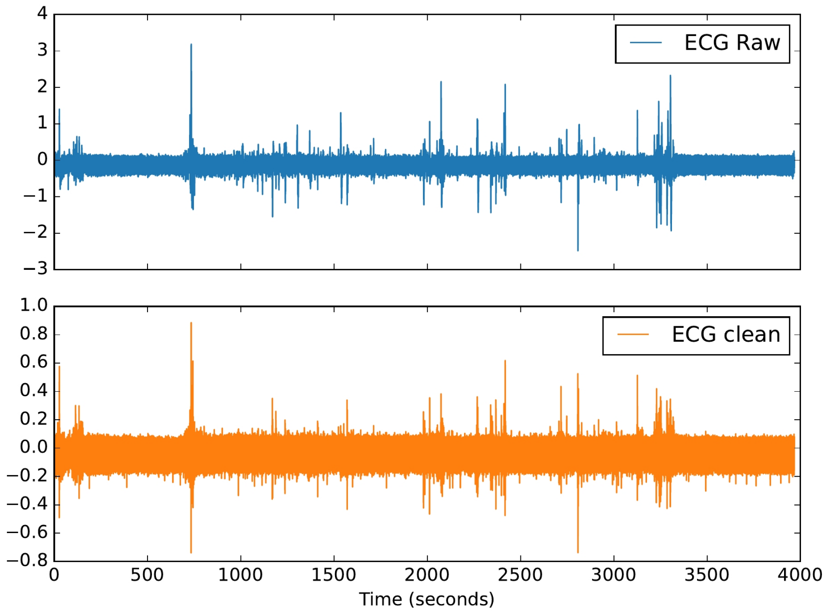

Figure 5.1: ECG signal before and after filters applied.

To analyze the filtration performance, Figure 5.2 zooms into the signals to observe their behaviors. After the filtration, it could be noticed how the baseline wander was almost completely removed, the R peaks were more pronounced, and some other corruptions in the signals were improved significantly.

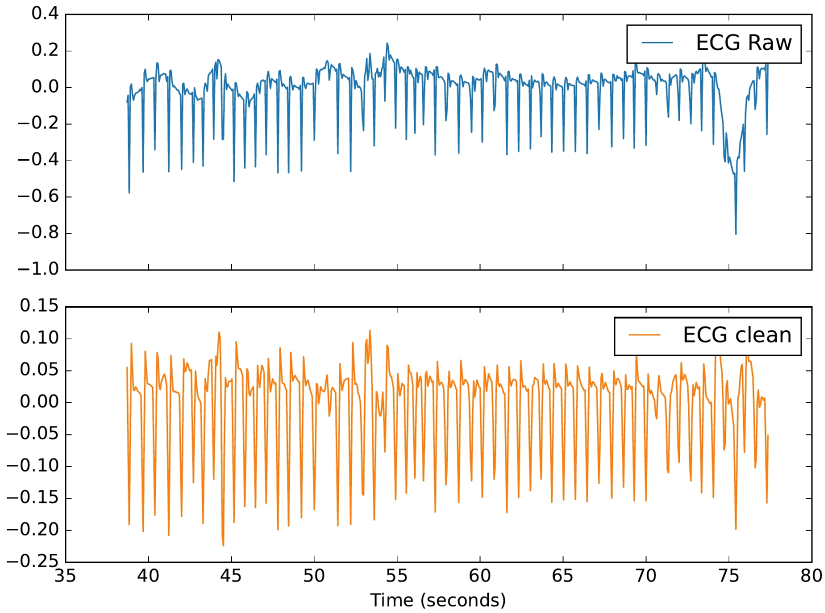

Figure 5.2: Zoom into raw and clean ECG signals.

5.2 Feature Extraction

After cleaning the signals, these were split into the 30s windows, and a split window of 1s was applied to each signal to obtain the intervals. Each interval was processed for the extraction of the features in the mentioned domains, providing a total of 14 useful features for the time domain: 'MeanNN', 'SDNN', 'RMSSD', 'SDSD', 'CVNN', 'CVSD', 'MedianNN', 'MadNN', 'MCVNN', 'IQRNN', 'pNN50', 'pNN20', 'HTI', 'TINN', while for the frequency domain there are 4 useful features: 'HF', 'VHF', 'HFn', 'LnHF'. Therefore, the data consisting of the HRV values and the 18 features were normalized using the Formula 4.1 for better data quality. The GSR signals were also split, obtaining the same number of values. GSR data was labeled as proposed, providing the labels for the supervised classification, while for the second database, the labels were obtained from the timestamp. Hence, the normalized and labeled dataset forms a data frame with the last column as the labeled one (No Stress:0, or Stress:1) Figure 5.3.

	HRV_MeanNN	HRV_SDNN	HRV_RMSSD	HRV_SDSD	HRV_CVNN	HRV_CVSD	HRV_MedianNN	HRV_MadNN	HRV_MCVNN
0	-1.175499631	-0.406853003	0.353805559	0.346654248	0.195770926	1.152293985	-1.190748203	-0.497954235	-0.063125097
1	-0.82545909	0.354649848	0.700955503	0.696364009	0.96717447	1.203220169	-0.832041471	0.623307896	1.096722008
2	-0.932265728	0.132702766	1.396296968	1.388066035	0.763684102	2.146047628	-0.832041471	0.623307896	1.096722008
3	-0.452333147	0.770106562	0.782998239	0.779573085	1.148613002	0.946727219	-0.473334739	1.183938962	1.39742311
4	-1.029261199	-0.18945681	0.946748305	0.938848947	0.393459644	1.726207651	-0.832041471	0.623307896	1.096722008
5	-0.959573267	0.23766655	1.532961254	1.521542275	0.937448195	2.347447058	-0.832041471	0.623307896	1.096722008
6	-0.658519554	0.736530455	1.453057356	1.446060686	1.306029417	1.871058471	-0.473334739	1.744570028	1.998825313
7	-0.597797913	0.652439354	1.571032478	1.56404293	1.138534244	1.929988606	-0.473334739	1.744570028	1.998825313
8	-0.796158956	0.310347746	1.296915449	1.289121023	0.879576266	1.857444963	-0.832041471	0.623307896	1.096722008
9	-0.551101779	0.665845152	1.169287241	1.164717062	1.111321084	1.446544217	-0.473334739	1.744570028	1.998825313
10	-0.60584353	0.669621104	1.815246771	1.808484377	1.168087956	2.204790094	-0.473334739	1.744570028	1.998825313
11	-0.392224174	0.868270372	0.967349868	0.965321663	1.21207028	1.085774704	-0.473334739	1.744570028	1.998825313
12	-0.857556422	-0.067416932	0.68540002	0.679943141	0.421633372	1.218004765	-0.832041471	0.623307896	1.096722008
13	-0.834192784	0.092073618	0.937483164	0.929964376	0.619209939	1.486540619	-0.832041471	0.623307896	1.096722008
14	-0.68220995	0.408574066	1.084747904	1.079665213	0.904145559	1.489604186	-0.832041471	0.623307896	1.096722008
15	-1.019557653	-0.312492352	0.554695219	0.547426894	0.208581809	1.233085512	-0.832041471	0.623307896	1.096722008

	HRV_IQRNN	HRV_pNN50	HRV_pNN20	HRV_HTI	HRV_TINN	HRV_HF	HRV_VHF	HRV_HFn	HRV_LnHF	Label
0	-0.075316873	1.058643203	1.2033396	0.478052725	-0.681495871	-0.833860411	-0.8184361	0.486528202	-0.564944181	0
1	0.427781424	1.240375217	1.451658706	0.313396059	-0.681495871	-0.124191226	-0.483532555	0.480270619	0.204449743	0
2	0.427781424	1.499768572	1.806094474	0.188791015	-0.681495871	-1.406622283	-1.038066465	0.181700762	-1.954675182	0
3	0.930879721	1.347409107	1.597910096	0.064185971	-0.982612831	-1.538213865	-1.008511881	-0.847220347	-2.716371587	0
4	-0.075316873	1.15614758	1.336569839	-0.06431298	-0.681495871	-1.575941779	0.296659263	-4.366772001	-3.055198962	0
5	0.427781424	1.385371866	1.649782511	1.219614561	-0.681495871	-0.459426538	0.696420483	-1.122433857	-0.103111632	0
6	0.930879721	1.103337218	1.264409621	0.851463295	-0.681495871	0.085478076	-0.932315468	1.16780403	0.36509435	0
7	0.930879721	1.357994876	1.612374519	0.458768611	-0.681495871	-0.773653006	-0.885531618	0.701191033	-0.478574991	0
8	0.427781424	1.372430016	1.632098733	-0.266796177	-0.982612831	-1.130580183	0.427222056	-2.234429562	-1.103788562	0
9	0.930879721	1.352805773	1.605284116	-0.100470694	-0.681495871	1.051113074	-1.005159552	1.382564676	0.921153957	0
10	1.056654296	0.961028527	1.069958648	1.160710358	-0.982612831	-0.671828971	0.910053952	-1.634497549	-0.344529216	0
11	1.433978019	1.341792168	1.590235096	0.917730522	-0.681495871	-0.540403976	-0.671203342	0.467825652	-0.189894867	0
12	0.427781424	1.254027499	1.47031322	0.438001103	-0.982612831	-0.029128649	0.391281941	-0.380756413	0.279777763	0
13	0.427781424	0.997116377	1.119269182	-0.266796177	-0.681495871	1.473041073	0.116968472	0.575346661	1.104704255	0
14	0.930879721	1.233162691	1.441803491	0.251093537	-0.681495871	0.589568801	0.587109651	-0.106376258	0.684933273	0
15	-0.075316873	1.139779456	1.314204391	-0.165554578	-0.681495871	0.050162283	2.280952942	-1.572318877	0.339401986	0

Figure 5.3: All columns and first fifteen rows of the original dataset.

5.3 Feature Selection

The feature selection methods were performed to determine the features in the dataset that most contribute to the classification task. The PC heat map plot obtained is shown in Figure 5.4, where colors can show how much each feature correlates to the other. Additionally, each square contains its value of correlation that ranges between -1 and 1. If two features have a correlation value of 1 or - 1, it implies that one set of the data is redundant, and we can assume the two features have the same information content. Therefore, based on the correlation value, we remove features that have a value above 0.9, assuming they provide redundant information. This process results in the elimination of 4 features.

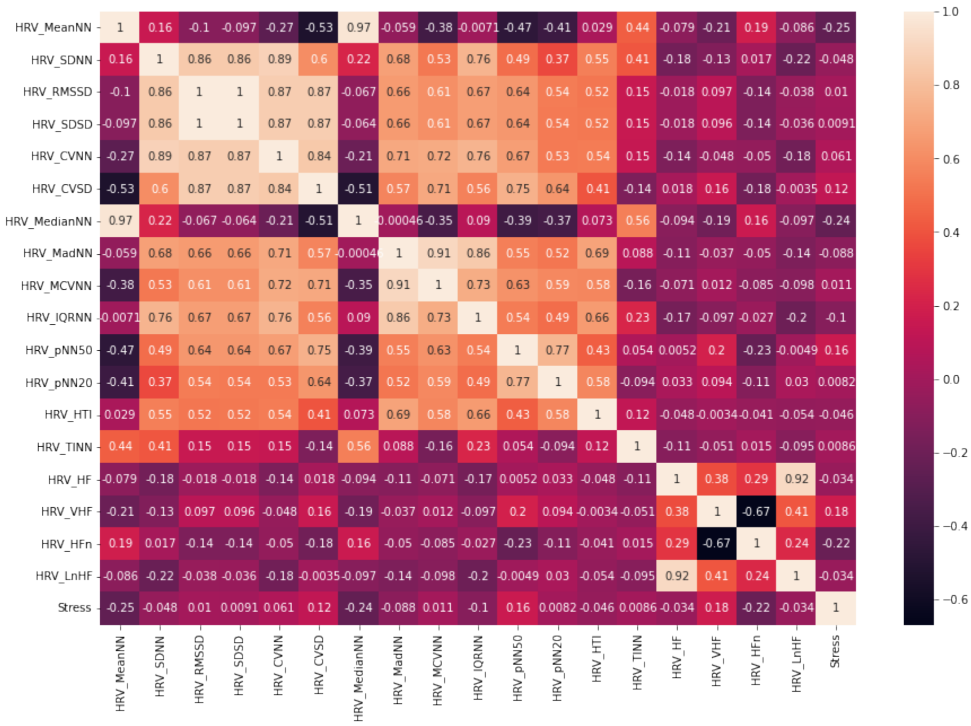

Figure 5.4: Heat map plot of Pearson's Correlation Feature Selection.

A two-tailed p-value indicates the likelihood that two uncorrelated objects might produce a Pearson's correlation value as extreme as the calculated value. For that reason, we use the p-value to eliminate features almost perfectly correlated. For this purpose, we set the null hypothesis to “The selected combination of dependent variables (features) do not have any effect on the independent variable (Stress)”. As the statistical analysis states: if the p-value is higher than the threshold, we discard that combination of features. Consequently, we obtain a new dataset containing nine features shown in Figure 5.5, contributing to an accurate classification based on the PC methodology.

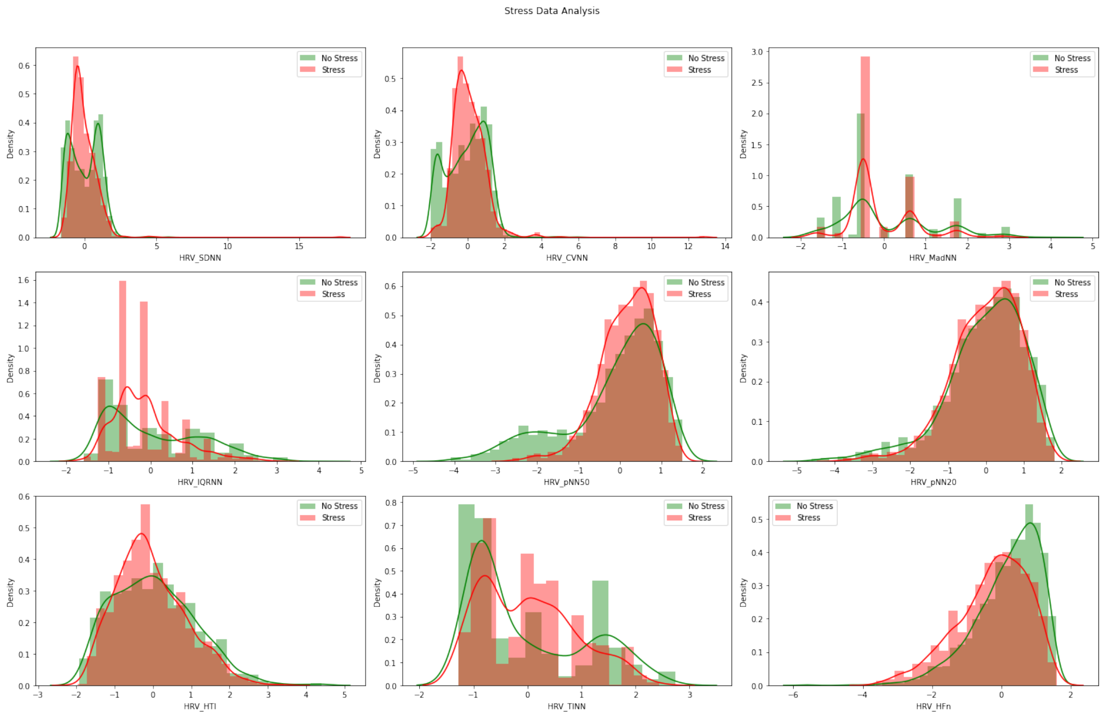

Figure 5.5: Distribution plot of selected features.

Following the proposed methodology, RFE was performed as another feature selection method. The estimator is one of the essential hyperparameters; an RF classifier is employed as the estimator. The other important hyperparameter is the number of features to select, which must be chosen for the algorithm; however, it is also possible to automatically select the number of features chosen by RFE with a Cross-Validation (RFECV) that automatically finds the optimal number of features to keep. In RFECV, feature importance is calculated based on the estimator selected, RF classifiers, and a few features are dropped in each iteration. A cross-validation generator is used, which splits the dataset into a sequence of train and test portions. We employ the Repeated Stratified K-Fold cross validator that repeats Stratified K-Fold (k smaller data sets) n times with different randomization in each repetition. This process is computationally expensive, but the great advantage is that it does not waste too much data, which is an important factor given the amount of data. The cross-validation is scored using accuracy as the metric. The RFECV performed a mean and standard deviation accuracy of 75.7% and 0.024, respectively. The RFE used an RF classifier, automatically selected a specific number of features, and fitted the RF using the selected features. Exploring the model, we can see 16 features shown in Table 5.1, used for the obtention of the results. Therefore, we obtain two datasets with a different number of features selected based on two reliable feature selection techniques.

Table 5.1: Feature Selection Results using two techniques

FS Technique	No. of features	Features Selected
PC	9	SDNN, CVNN, MadNN, IQRNN, pNN50, pNN20, HTI, TINN, HFn
RFE	16	MeanNN, SDNN, RMSSD, SDSD, CVNN, CVSD, MedianNN, IQRNN, pNN50, pNN20, HTI, TINN, HF, VHF, HFn, LnHF

5.4 Stress Classification

Stress classification with KNN and DT was performed using the two datasets to analyze metrics scores and determine which features or, in this case, a set of features provide vital information about stress. Since the number of stress cases does not differ significantly from nonstress cases, we can affirm that the data is balanced. Therefore, both datasets are ready for classification.

5.4.1 KNN Classifier

The KNN model as a classifier has one hyperparameter: the number of nearest neighbors, and the choice of k has a crucial effect on its classification performance. The optimal choice of the value k is highly data-dependent: generally, a larger k suppresses the effects of noise but makes the classification boundaries less distinct. Figure 5.6 shows the differences in boundaries as the k number of neighbors changes: a) 3, b) 6, and c) 9 neighbors. Based on the 3-Class dataset, it is observed how the boundaries for each class change as the k number increases.

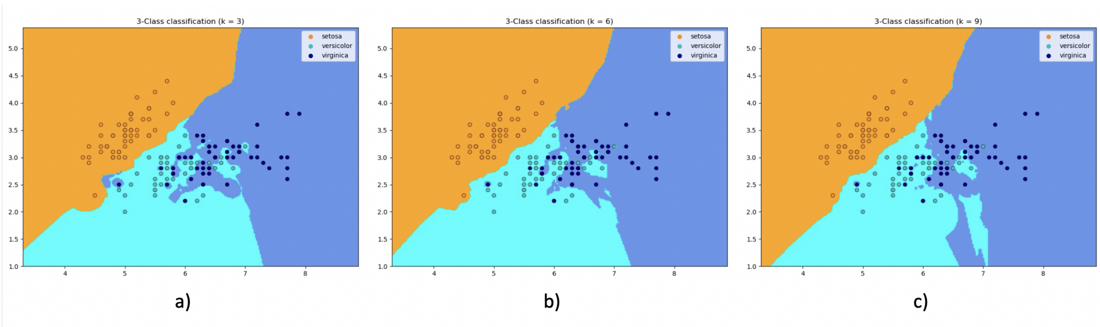

Figure 5.6: K number influence on the model.

This hyperparameter was changed from 3 to 10 k neighbors to analyze the model's behavior and see which number provides the best accuracy. Once again, cross-validation is added to the model to avoid overfitting, data leakage, and reduction of samples. Table 5.2 shows the accuracy, precision, recall, and F-1 score obtained with the k neighbors. Each metric shows its mean value and its standard deviation. *Experiment 1* references the dataset obtained with PC feature selection, while *Experiment 2* shows the results using the dataset obtained with RFECV feature selection.

Table 5.2: KNN stress detection performance

Neighbors	Accuracy	Recall	Precision	F-1 Score
Experiment 1 (PC results dataset)				
3	0.683 (0.024)	0.609 (0.038)	0.635 (0.031)	0.674 (0.024)
4	0.685 (0.022)	0.487 (0.035)	0.686 (0.041)	0.660 (0.024)
5	0.700 (0.024)	0.637 (0.041)	0.654 (0.033)	0.692 (0.025)
6	0.696 (0.026)	0.536 (0.040)	0.684 (0.041)	0.677 (0.028)
7	0.704 (0.024)	0.642 (0.041)	0.658 (0.032)	0.696 (0.025)
8	0.704 (0.022)	0.572 (0.044)	0.686 (0.033)	0.690 (0.024)
9	0.713 (0.027)	0.653 (0.045)	0.670 (0.035)	0.706 (0.027)
10	0.710 (0.025)	0.590 (0.049)	0.688 (0.035)	0.697 (0.028)
Experiment 2 (RFE results dataset)				
3	0.692 (0.032)	0.614 (0.045)	0.648 (0.042)	0.683 (0.032)
4	0.698 (0.022)	0.485 (0.042)	0.717 (0.037)	0.671 (0.026)
5	0.705 (0.024)	0.611 (0.044)	0.671 (0.030)	0.695 (0.025)
6	0.701 (0.022)	0.521 (0.040)	0.704 (0.035)	0.680 (0.025)
7	0.713 (0.026)	0.630 (0.042)	0.678 (0.035)	0.704 (0.027)
8	0.716 (0.022)	0.562 (0.040)	0.714 (0.033)	0.699 (0.024)
9	0.722 (0.020)	0.629 (0.039)	0.693 (0.027)	0.712 (0.021)
10	0.720 (0.021)	0.565 (0.039)	0.720 (0.033)	0.703 (0.023)

5.4.2 DT Classifier

The DT classifier works with a criterion that measures the quality of the split. Since DT aims to partition the data into homogeneous groups where the split nodes are pure, purity in this context can be defined as minimizing the misclassification error. It can be measured with the Gini index or cross-entropy. The algorithm evaluates nearly all split points and selects the split point value that minimizes the purity criterion. However, many researchers suggest that the choice of splitting or purity criteria will not significantly affect the tree performance in most cases. The splitting process will be performed using the Gini index, and the strategy to split each node is random, selecting a set of features and splits that are less prone to overfitting.

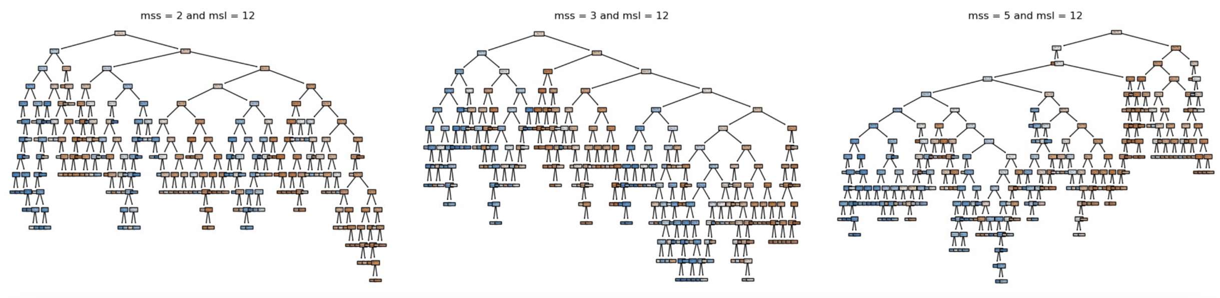

Figure 5.7: DT depth as mss increases.

Based on [109], the ideal *min_samples_split* (mss) values tend to be between 1 to 40 while *min_samples_leaf* (msl) values tend to be between 1 to 20 for the CART algorithm, which is the algorithm implemented in scikit-learn. Therefore, the model continues increasing the depth of the tree until the stopping criteria are met, which in this case are the mss and msl. Figure 5.7 and 5.8 illustrate the tree model's complexity and depth as the hyperparameters change. As seen in Figure 5.7, the tree classifier depth increases as the number of mss increases. On the other side, it can be seen in Figure 5.8 that the tree decreases in depth and complexity as the msl number increases.

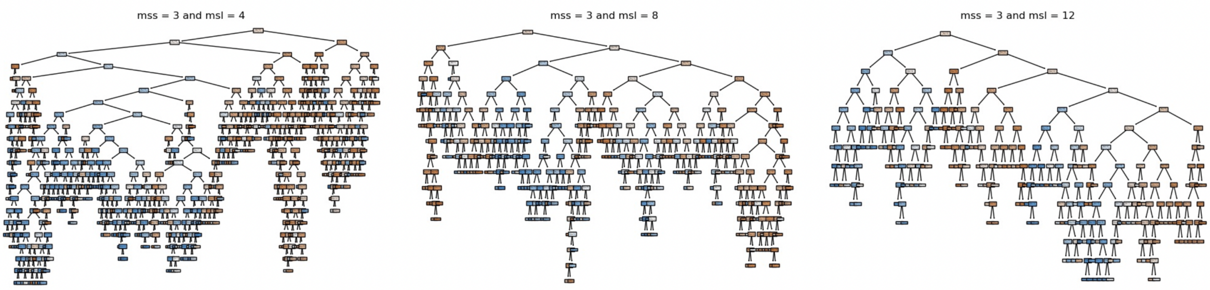

Figure 5.8: DT depth as msl increases.

After hyperparameters tunning, the best minimum number of samples to split an internal node value was 2, and the minimum number of samples required to be at a leaf node was 13 to obtain good performance for the DT. The data is split into train and test sets for this classification with a 70% and 30% ratio, respectively. Metrics are measured against the test or validation set to detect overfitting problems. Results of the different experiments are shown in Table 5.3, where *mss* refers to the minimum samples to split and *msl* to the minimum samples to be at a leaf node.

Table 5.3: DT stress detection performance

mss	msl	Accuracy	Precision	Recall	F-1 Score
Experiment 1 (PC results dataset)					
	4	0.646	0.665	0.758	0.709
	8	0.640	0.668	0.732	0.698
2	12	0.662	0.703	0.701	0.702
	13	0.655	0.712	0.661	0.686
	16	0.672	0.721	0.692	0.706
	4	0.632	0.664	0.715	0.689
	8	0.647	0.667	0.758	0.709
3	12	0.652	0.693	0.697	0.695
	13	0.692	0.706	0.789	0.745
	16	0.682	0.691	0.783	0.734
	4	0.683	0.688	0.795	0.738
	8	0.712	0.735	0.761	0.748
5	12	0.647	0.653	0.790	0.715
	13	0.680	0.687	0.790	0.735
	16	0.688	0.696	0.788	0.739
Experiment 2 (RFE results dataset)					
	4	0.703	0.732	0.794	0.762
	8	0.695	0.752	0.732	0.742
2	12	0.683	0.748	0.709	0.728
	13	0.730	0.751	0.800	0.775
	16	0.727	0.735	0.829	0.779
	4	0.686	0.708	0.781	0.743
	8	0.708	0.743	0.760	0.752
3	12	0.732	0.746	0.818	0.780
	13	0.712	0.758	0.742	0.750
	16	0.720	0.743	0.793	0.767
	4	0.712	0.746	0.765	0.755
	8	0.703	0.735	0.765	0.750
5	12	0.700	0.725	0.779	0.751
	13	0.715	0.740	0.786	0.762
	16	0.714	0.743	0.774	0.759

5.4.3 RF Classifier

As mentioned, RF is an assembly of random decision tree classifiers, where the prediction output is made of all the prediction outputs by the individual classifier. The two RF hyperparameters are the number of decision trees to be generated (Ntree) and the number of variables to be selected and tested for the best split when growing the trees (Mtry). RF classification model is characterized as being computationally efficient. It is a model that does not overfit because many uncorrelated models (classification trees) operate as a committee to outperform any individual trees. Therefore, Ntree can be as large as possible. For this stress classification task, Ntree works as the hyperparameter varying from 50 to 5000 to analyze the relationship between the number of trees and classification performance. Results are shown in Table 5.4.

Table 5.4: RF stress detection performance

Ntree	Accuracy	Precision	Recall	F-1 Score
50	0.753	0.781	0.801	0.791
100	0.747	0.778	0.794	0.786
200	0.744	0.781	0.783	0.782
300	0.753	0.789	0.787	0.788
400	0.753	0.786	0.792	0.789
500	0.758	0.792	0.794	0.793
600	0.751	0.786	0.789	0.788
700	0.754	0.792	0.785	0.788
800	0.754	0.786	0.796	0.791
900	0.781	0.799	0.839	0.814
1000	0.763	0.795	0.801	0.798
2000	0.755	0.789	0.794	0.791
3000	0.755	0.787	0.796	0.792
4000	0.758	0.790	0.799	0.794
5000	0.757	0.790	0.794	0.792

5.4.4 SVM Classifier

The SVM classifier is suited for both linear and non-linear data classifications. It aims to project separable samples onto another higher dimensional space by using different types of kernel functions. The use of varying kernel methods helps in dealing with computational complexity issues. Due to the popularity of SVM for classification tasks, there has been a focus on studying these kernel functions, given their significant role on SVMs. Choosing the correct kernel is crucial for classification; popular kernels are linear, polynomial, Gaussian, Radial Basis Function (RBF), Laplace RBF, Sigmoid, and Anove RBF Kernel, among others. Considering that the features used for this stress classification are linear HRV indices, a linear kernel is implemented for the SVM model. Another important tuning parameter is the regularization parameter (C), which tells the SVM model how much to avoid misclassifying each training example. C is inversely proportional to the strength of regularization and the number of support vectors.

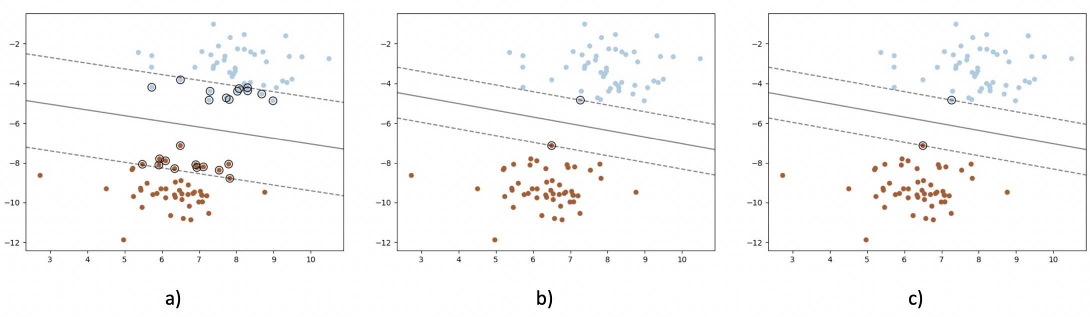

Figure 5.9: C influence on hyperplane.

As C is higher, the hyperplane will be smaller, and the training data miss classification rate will be lower. Figure 5.9 shows these changes as C takes different values tested in this study: a) 0.01, b) 1, and c) 100. However, increasing C does not always mean better accuracy since the model can overfit. In this case, tests have been performed with different C values, and the results are shown in Table 5.5.

Table 5.5: SVM stress detection performance

C	Accuracy	Precision	Recall	F-1 Score
0.001	0.634	0.655	0.790	0.716
0.01	0.676	0.739	0.719	0.729
0.1	0.674	0.747	0.697	0.721
1	0.684	0.750	0.692	0.720
10	0.683	0.751	0.687	0.718
100	0.690	0.760	0.687	0.722

The regularization parameter C set to 100 has achieved the highest accuracy. However, cross-validation has been performed to compare the accuracy to ensure the SVM model is not overfitting. It has shown that SVM models with C equal to or greater than 1 are overfitting. Therefore, it can be determined that the SVM classifier with C = 0.01 scores the highest accuracy with 67.6% and an F1 score of 0.729.

5.4.5 MLP Classifier

The MLP is a supervised learning algorithm that learns a function by training on a dataset. It can learn a non-linear function approximator for either classification or regression. It is different from logistic regression in that between the input and the output layer, there can be one or more non-linear layers, called hidden layers. MLP as a classifier comprises different hyperparameters to be tuned. Therefore, a cross-validation and grid search (CVGS) has been performed for this model to find the hyperparameters most suitable for this classification. For the CVGS technique, the parameters to vary are *hidden_layer_sizes* that specifies the number of hidden layers and hidden units, *activation* which is the function for the hidden layer, *solver* for weight optimization, *max_iter* which is the maximum number of iterations, and *alpha* which is an important parameter used for regularization that helps to avoid overfitting by penalizing weights with large magnitudes.

After CVGS, the technique determined the best hyperparameters as followed: *hidden_layer_sizes* = (100,), *activation* = logistic, *solver* = adam, *max_iter* = 300 and *alpha* = 0.1. Therefore, the dataset tested the MLP classifier using the mentioned hyperparameters. However, more experiments were performed, varying the regularization parameter, and the results are shown in Table 5.6.

Table 5.6: MLP stress detection performance

alpha	Accuracy	Precision	Recall	F-1 Score
0.001	0.732	0.752	0.791	0.771
0.01	0.735	0.751	0.800	0.775
0.1	0.724	0.746	0.781	0.763
1	0.670	0.683	0.784	0.730
10	0.571	0.570	0.995	0.725

Once the best hyperparameters of each ML model have been determined, we can perform other experiments changing the window size used for the signals. For the experiments performed, the window intervals have a size of 30 seconds. By splitting the signals into 20 and 40 seconds intervals and using the hyperparameters that acquired the highest accuracies on each model, we can see the performance scores in Table 5.7.

Table 5.7: ML performance with data split on different windows

Window size	Metric	KNN	DT	RF	SVM	MLP
20s	Accuracy	0.704	0.680	0.751	0.674	0.714
	Precision	0.726	0.693	0.751	0.711	0.722
	Recall	0.754	0.756	0.819	0.731	0.795
	F1 score	0.719	0.723	0.784	0.721	0.757
30s	Accuracy	0.722	0.732	0.781	0.676	0.735
	Precision	0.693	0.746	0.799	0.739	0.751
	Recall	0.629	0.818	0.839	0.719	0.800
	F1 score	0.712	0.780	0.814	0.729	0.775
40s	Accuracy	0.725	0.707	0.754	0.664	0.734
	Precision	0.767	0.739	0.778	0.712	0.753
	Recall	0.758	0.770	0.807	0.712	0.775
	F1 score	0.763	0.754	0.792	0.712	0.764

Finally, once obtained the best hyperparameters were found for each model, in order to increase the metrics scores, cross-validation (CV) was performed to evaluate the models. A k-fold-cross-validation generally results in a less biased estimate of the model skill than simple train-test split methods. A value of k=10 is widespread in applied ML; therefore, a 10-fold-cross-validation is performed in this case, and results are shown in Table 6.1 with their accuracy and F1 scores.

Table 5.8: ML models' results after CV

Model	Accuracy	F-1 Score
KNN	74.9%	78.8%
DT	73.2%	78.0%
RF	82.4%	87.8%
SVM	67.6%	72.9%
MLP	73.5%	77.5%

5.5 Experimental Setup

Domestic signals were successfully obtained, presenting noise that made it challenging to analyze the different peaks of a standard ECG signal, as seen in Figure 5.11.

After the filters were applied, the signals' results are shown in Figure 5.11, where the conditioned signal refers to the combination of the three filters mentioned. It can be seen how each filter changes the signal amplitude, and the noises are reduced with most of them. However, the signal resulting from the combination of the high and low pass filters looks

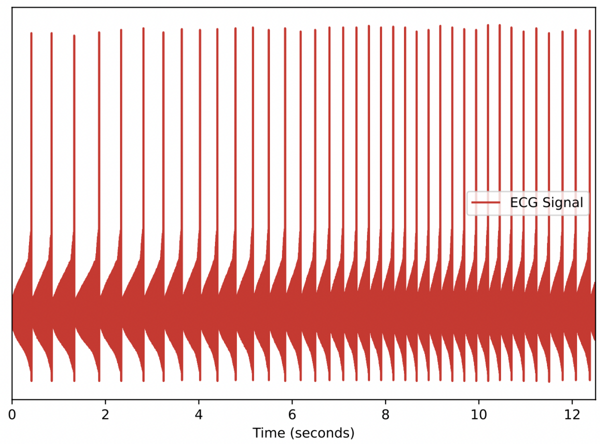

Figure 5.10: Raw ECG signal obtained with the AD8232 module.

cleaner and with fewer corruptions. Therefore, these results are used to obtain the HRV features.

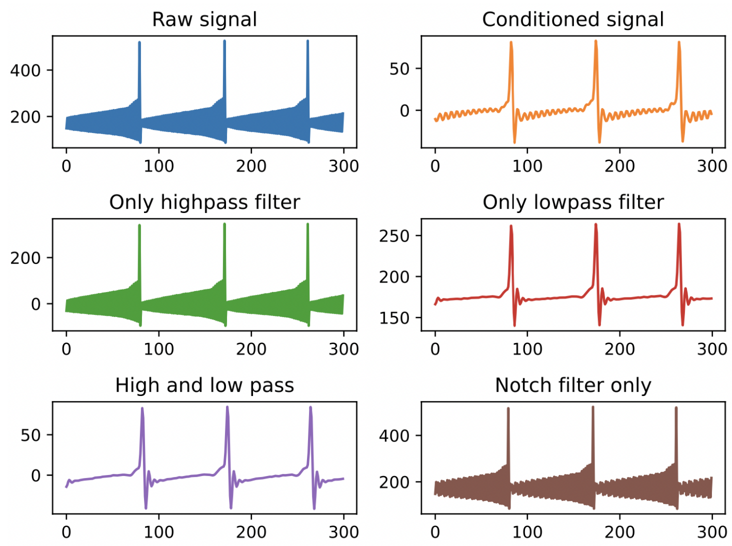

Figure 5.11: Raw ECG and ECG cleaned with different filters.

An interval with the 16 features was obtained the same way as those from the chosen databases, and classification was performed once the best classifier was determined to test this stress prediction protocol. The obtained interval was successfully predicted using the RF model, which was a nonstress phase, while the other models could not accurately predict it. However, domestically obtained signals are not long enough compared to the signals used to train and test the models, which can lead to errors in the frequency features that depend on the length of the signals. Consequently, more signal recordings are needed to test the developed protocol.

Chapter 6

Discussion

This work proposed two feature selection methods, as seen in Table 5.1; the PC method provided 9 HRV features while RFE eliminated more elements by selecting 16 for the classification of stress. Each set of features formed a dataset for stress classification experiments using two ML models: KNN and DT classifiers.

For the classification with KNN, both experiments were tested using the same hyperparameters to determine which dataset of features performs better for the model. Besides obtaining high results in the two experiments, the dataset consisting of features selected by the RFE method outperformed the dataset made by the PC feature selection. The accuracy and f1 scores in the second experiment were higher no matter the number of neighbors used in the model. In the same way, experiment 2 in the DT stress classification outperformed each test for the different number of samples at the node and the number of samples to split. It was enough information to determine the 16 features selected with the RFE technique, provided relevant information that most contributes to this stress detection task, and can be used to feed an ML model. Therefore, the other ML models tested in this study were fed with these 16 features as the dataset to analyze their performance after hyperparameter tunning and determine which model and hyperparameter can provide the highest accuracy for stress classification.

For the classification using the KNN classifier, as shown in Table 5.2, the No. of k neighbors = 9 offers the highest accuracy for each experiment. Analyzing the other metrics of the KNN model with 9 k neighbors, precision's performance is more outstanding than recall in both cases, which can be good since we want to detect as many subjects who suffer from stress as possible. Regarding the F1 score, both models with 9 k neighbors have the best balance between recall and precision. Therefore, we can suggest that the KNN classifier with nine neighbors is a good approach for stress detection using HRV features.

The DT classifier had two hyperparameters to be tuned. Experiments performed with three minimum samples to split an internal node showed higher performance scores than the other two experiments, Table 5.3. Additionally, 12 minimum samples at a leaf node are needed to outperform other models. Again, the performance was better for accuracy and the f1 score, which is essential for this medical task, by determining an excellent detection of subjects suffering from stress and a balance between precision and recall. DT performance results show a practical approach for stress detection since another study that tested this algorithm for detecting anxiety, depression, and stress [86] achieved an accuracy of 62.8% and an F1 score of 59.2% for stress classification.

When employing the RF classifier, all the tested models perform with a great accuracy greater than 70% besides the number of trees, showing no significant difference between their performance (Table 5.4. However, the classification using 900 and 1000 Ntress outperforms the other models, positioning the 900 Ntress model as the most accurate among the RF classifications with an accuracy of 78,1%. The RF has performed a reliable stress classification since other studies concluding that RF is the best model for stress detection tasks have achieved less accuracy than these results, such as in [110] where the maximum accuracy was 60%. This RF model with 900 trees also provides the highest F1 score of 0,814, showing an outstanding performance in balancing recall and precision.

The regularization parameter C was the hyperparameter tunned to analyze the misclassification rate for the SVM model. Since this hyperparameter is prone to overfitting, cross-validation showed that higher C values tended to overfit the data. Therefore, it was observed that SVM is a model that tends to overfit this task, and C is an essential hyperparameter to take into account for analyzing the performance. On the other side, C = 0.01 is a value that did not overfit and performed better than different values (Table 5.5. However, the accuracy of 67.6% as obtained with this model is not high enough compared to the previously discussed models. As seen in Table 3.2, many studies obtained a high performance when using an SVM model, and this algorithm outperformed other ML algorithms. In this case, the SVM classifier has been able to classify but has not outperformed models as the KNN and DT studied here.

In the MLP classifier, a previous step of CVGS was performed due to the number of hyperparameters. However, the alpha values, a critical hyperparameter, were also tuned after determining the best choices for the other hyperparameters. Besides the CVGS technique selected alpha = 0.001 as the best regularization value, additional experiments showed that with alpha = 0.01, the MLP achieved an accuracy, recall, and F1 score a little higher. On the other side, it can be observed in Table 5.6 that as the alpha value increases, the model performance decreases. Therefore, alpha remains at 0.01 for other experiments. The accuracy obtained is 73.5% which seems valid since [95] performed a stress detection with different levels where MLP achieved an accuracy of 67%. When performing a binary classification with non-filtered signals, the accuracy increased significantly; however, the authors focus on other metrics, such as the F1 score and the area under the curve.

By looking at the different machine learning models' performance, it is possible to suggest that the ECG preprocessing methods, features extraction, and selection has been successfully performed to obtain accurate stress classifications (Figure 6.1). Therefore, we can assume the RFE technique with cross-validation is a reliable methodology for selecting HRV features, as in [111], where RFE was employed to select the HRV top features that fed an ML algorithm.

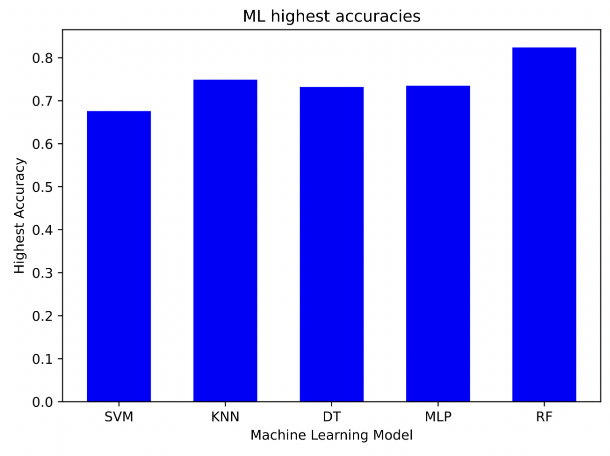

Figure 6.1: Bar chart of models' accuracies

Comparing the machine learning models used for classification, it is observed that besides every model performing an accurate classification by a correct hyperparameter tunning based on literature search and experimenting, the RF model obtained the highest one. Table 6.1 shows the highest accuracy of each model, its F1 score, and the tunned hyperparameters each model used for obtaining the mentioned performance scores. These results indicate that the stress classification methodology proposed using HRV features and an RF classifier with 900 the number of trees for classification can successfully detect stress in real-time.

Table 6.1: ML models with highest accuracy

ML model	Accuracy	F-1 Score	Hyperparameters
KNN	74.9%	78.8%	9 K-nearest neighbors
DT	73.2%	78.0%	3 mss and 12 msl
RF	82.4%	87.8%	900 numbers of trees
SVM	67.6%	72.9%	C = 0.01
MLP	73.5%	77.5%	alpha = 0.01

Since the objective of this study is to develop an accurate stress detection model with the potential to be applied to real-time applications, the evaluation of ML models was evaluated on different and small sliding window sizes of the ECG signal to determine which size could provide the best accuracy. Table 5.7 exposes the ML models' performance with their best hyperparameters in windows of 20, 30, and 40 seconds. Besides being the KNN model with a 40s window size, a model with better performance than the 30s window size, the 30s window size shows a generally more outstanding performance than the other window sizes in terms of accuracy. Regarding the F1 score, there are slight differences between some models and window sizes; however, generally speaking, the 30s window size has a higher F1 performance. As a result, it is correct to assume that the segmentation method for the ECG signals into a fixed window of 30s which coincides with the study of Rastgoo et al. [77] where the sliding window size was a parameter studied for the performance of a

DL model for stress classification, showing the 30s scored the best accuracy compared to 10 and 5 seconds.

Consequently, analyzing the ML models' performance with the dataset obtained by segmenting ECG signals into 30s window size and the 16 features determined after an RFE technique is correct. That being the case, RF is the most accurate model (Table 5.7).

RF algorithms have already been used for stress classification problems, as shown in Table 3.2. However, in most cases, other models have outperformed the RF technique. For example, a study classifying stress compared different ML models, including an RF classifier; however, the highest score was achieved with a Boosting model that performed 75,13% [112]. In the same way, Priya et al. [86] studied the prediction of anxiety, depression, and stress using different ML algorithms, where the results showed 74.2% as the highest accuracy for stress classification using an NB algorithm; however, the F1 score performed with the model was 55.8%, which as mentioned before, suggests poor precision and poor recall meaning the number of FP and FN are not as low as expected. On the other side, a study in 2017 that classified stress with some models, including RF and NB, concluded that RF got the highest prediction accuracy, matching these results. Other studies described in Table 3.2 did not get the highest accuracy with RF as the classifier. It can be due to the different techniques used for classification, such as filtering or feature selection. This study explains that many parameters, from the data to the model selection, can vary the final performance.

One aim of this study is to provide a methodology for stress detection that can be replied in a real-life scenario and outside laboratories in an environment not necessarily strictly controlled, such as a patient home, so that with simple ECG sensors, a stressed person can be detected based on this ML classification methodology. Therefore, the results in this study perform remarkably considering the number of features used, especially the number of signals, which is 1: ECG. Additionally, experimentation using an AD8232 module and an Arduino was performed to obtain nonstressed ECG signals, proving the whole protocol can classify and detect stress or the lack of it. Most studies use two or more signals for stress classification tasks. For example, [87] used four physiological signals such as ECG, EMG, GSR, and RESP, achieving an accuracy higher than 90%. Many previous works have used this multi signals methodology (Table 3.2). Nevertheless, although obtaining a high accuracy, multiparameter techniques include more sensors, more computational cost, and significantly more discomfort for the subject to whom the experiment can be replied, which can induce stress and lead to the accurate detection of temporal stress that does not need to be treated.

Limitations

This thesis is trying to develop a system to detect stress using physiological signals, focusing on using a supervised ML model to classify ECGs since these signals can be easily monitored in an ambulant environment. The main limitation of this classification is the lack of data from the database. The “Stress Recognition in Automobile Drivers” database lacks signals for some participants, or the signal data are incomplete, which reduces the amount of data. It reduces the dimensions of the dataset, which connotes less data for training and testing the model. As mentioned in 4, the complete data from the original experimental acquisition study is not publicly available, consisting of hours or signal recordings.

Additionally, stress markers are not present in all the experiments, so the GSR signals have been employed to determine the presence of stress and label the data. Consequently, the classification models have been unable to improve their performance by changing hyperparameters or using cross-validation techniques. However, the algorithms used have obtained acceptable results considering the data dimensions and the number of signals used.

The room for improvement is significant. More ECG recordings from stress experiments can yield a considerable amount of data and add up an improvement in the dataset. In this case, the classification models have more information for training and testing, suggesting an improvement in the metrics.

Chapter 7

Conclusions

This project discussed three main topics: stress definition, the physiological stress response, and machine learning techniques for stress classification. Stress has been defined as focused on the homeostasis changes induced in the body and how these alterations lead to severe health problems. Therefore, the stress definition from Hans Selye has been chosen for this study, which addresses stress as psychological or physiological agents (stressors) that elicit a behavioral and physiological stress response.

The most common physiological signals and machine learning techniques for stress detection in the past five years were discussed. However, important information is lacking in the literature. For example, the features extracted and selected from physiological signals and their combinations for use in ML classification are not explicitly stated. Although authors have addressed the obtention of the data in the methodology, it is inconclusive due to differences in protocols, stressors, and analysis techniques, so it is not clear the processed dataset or set of features that provide the accuracies the studies obtain as results. On the other side, most studies or the signals for their datasets have been executed in laboratories or under controlled environment conditions. Thus, whether insights based on these studies can be extended to the ambulant environment in real-life events should be investigated.

Furthermore, the literature review shows which ML techniques are most suitable for stress detection using physiological signals is unclear. Each study used specific biosignals or sets to obtain features that are not always detailed. The differences in data, methodologies, and results presume that a one-model-fits-all solution is not feasible for this task. Each model should be personalized based on the dataset input, methods, and purpose.

The optimal physiological signals for stress detection were identified based on the fundamentals of stress, its effects on the body, and the different systems used to measure it. ECG was selected, which provides HR data and HRV features in time and frequency domains. This signal was chosen considering the environment of the experiments for their obtention, which is a real-life task, and considering that ECGs can be obtained both in controlled and not controlled settings, which is crucial for the objective of this project and future research. Additionally, essential factors for this signal choice are the computational cost that signals carry and their feasibility in the acquisition, which implies it does not produces discomfort to the subjects or induces psychological or physical stress that leads to misclassification.

Experiments without the cleaning or prepossessing step led to an error in the peaks detection, which did not permit the feature extraction of the signals. Therefore, ECG signals were cleaned using two filters to eliminate the most common noises. Then a window sliding technique was applied to segment the signals into intervals of 30 seconds before HRV analysis. Afterward, eighteen features from the ECG segments were extracted, and the most relevant ones were identified as essential data for the classification task using two feature selection techniques. In the same way, GSR signals were segmented and used for labeling the data to obtain an appropriate dataset for supervised classification. The features selected with each technique, PC and RFE, formed a dataset that was tested for stress classification using two models: KNN and DT. After experiments were performed, based on the results, it was concluded that the RFE technique with a CV provides the set of features that primarily contribute to this stress classification task.

Five supervised machine learning algorithms were employed for stress classification once the features were selected and the data was normalized. KNN, DT, RF, SVM, and ANN were used for different experiments. Besides considering the hyperparameters set on previous works for each classification model, it was necessary to perform hyperparameter tuning to determine the values that can obtain high performance of the classifier. Therefore, experiments were carried out with varying hyperparameter values, and the success of each model was measured using four different metrics: accuracy, precision, recall, and F1 score. Accuracy and F1 scores were mainly considered since they provide information about the correct and incorrect predictions of the model and yield a general overview of the balance between precision and recall, which are essential information for health-related classifications.

The segmentation of the signals performed in the preprocessing step was another parameter analyzed and tested using the models with their best hyperparameters. It was concluded that segmenting the signals in window sizes of 30s preserves enough data for the peaks detection and features extraction and their posterior use as data input for classification. The algorithms' performance showed that all five could classify stress with more than 65% accuracy. However, it is concluded that the RF classifier outperformed the other models with an accuracy of 82.4% and an F1 score of 87.8% when setting the number of trees to 900 and performing a binary stress classification based on 16 HRV features obtained from the time and frequency domain.

On the other side, the classifier that performed the worst is the SVM with an accuracy of 67.6%, an F1 score of 72.9%, and setting its regularization parameter to 0.01; metrics obtained after a cross-validation technique to avoid overfitting of the model when increasing its regularization parameter. The analysis showed that some of these results agree with previous studies where the RF classifier has outperformed other models when detecting stress. However, the SVM is the model mainly used for this task based on the review, and besides performing less than other complex models in some studies, it has been characterized as the most accurate model in others. After examining the difference in data and signals used, the features extracted, and the methodology proposed, it is proper to suggest these factors create a significant difference in each model's performance resulting in differences in some models' efficiency compared to previous works.

The current work aims to provide a methodology to solve the problem of stress detection out of laboratories and controlled environments, which is fundamental to the goal of continuous, physiological stress detection in daily life. One physiological signal that can be easily acquired and monitored in ambulant environments has been chosen to keep the complexity of this system and computational cost low, so the proposed protocol can be implemented in daily-live activities to detect stress in modern life. Therefore, considering the amount of data, the methodology, and the classifiers used, it can be concluded that these results provide a baseline for ambulatory population monitoring to detect physiological responses to stress and provide early personalized treatment to avoid future diseases. Other devices can also be used to determine stress by applying the same protocol described, such as smartwatches, which can acquire ECG signals, making it easier for some people to monitor their stress levels in daily life continuously and track their daily levels of stress, whether it is for personal purposes such as being aware of the activities that can lead to stress or for medical purposes when a person has been detected with problems where stress and cortisol levels must stay low. However, further research is needed to investigate if these conclusions generalize to multiple clinical diagnoses. These results highlight the potential of using electrocardiogram signals to detect stress by measuring the body's responses. Further longitudinal research using wearable technology to investigate stress development could better understand the development of stress-related disorders.

Future Work

The application of this thesis can result in a real-time personalized stress detection model. Controlled experiments for data collection can be performed in daily-life activities that can induce physiological or psychological stress, using wearable devices that collect this data. The signals obtained can feed and train the proposed classification models, improving the performance. This stress classification protocol can be used in hospital and clinical environments; however, it would have a better impact outside these settings, like in-house environments for regular analysis of stress in people so they can analyze their stress records at the end of each day to recognize their emotions and be alerted in the early stage of chronic stress.

Bibliography

[1] A. Psychological Association, "STRESS IN AMERICA ™ 2021 Stress and Decision-Making During the Pandemic."

[2] S. Bennett Johnson, C. D. Belar, S. J. Breckler, K. C. Nordal, D. Ballard, L. F. Bufka, L. Bossolo, M. Salomon, S. Research Director, C. Jim Blumenthal, P. in Psychiatry, B. Health Sciences, T. L. Bourdeau, C. Assistant Professor of Behavioral, R. Millard, M. Vanguard Communications Brenda Foster, B. Horton, S. Account Supervisor Stephanie Dukes, and S. Account Executive, "Our Health at Risk."

[3] V. SS, A. A, A. HJ, B. EJ, B. MS, C. CW, C. AP, C. AM, C. S, D. FN, E. MSV, E. KR, F. JF, G. DK, K. SS, K. BM, K. KL, L. CD, L. TT, L. J, L. MS, L. PL, M. J, M. J, M. SS, M. DB, M. ME, N. SD, P. AM, R. GA, S. Z, S. GM, S. EB, S. SH, S. CM, S. A, V. LB, W. NY, and T. CW, "Heart Disease and Stroke Statistics-2021 Update: A Report From the American Heart Association," *Circulation*, vol. 143, no. 8, pp. E254–E743, 2021. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/33501848/

[4] S. Sarkar, R. Gupta, and V. Menon, "A systematic review of depression, anxiety, and stress among medical students in India," *Journal of Mental Health and Human Behaviour*, vol. 22, no. 2, p. 88, 2017. [Online]. Available: https://www.jmhhb.org/article.asp?issn=0971-8990;year=2017;volume=22;issue=2;spage=88;epage=96;aulast=Sarkarhttps://www.jmhhb.org/article.asp?issn=0971-8990;year=2017;volume=22;issue=2;spage=88;epage=96;aulast=Sarkar;type=0

[5] "Middle income — Data." [Online]. Available: https://data.worldbank.org/income-level/middle-income

[6] H. Selye, "A Syndrome produced by Diverse Nocuous Agents," *Nature 1936 138:3479*, vol. 138, no. 3479, pp. 32–32, 1936. [Online]. Available: https://www.nature.com/articles/138032a0

[7] H. SELYE, "Stress and psychiatry." *The American journal of psychiatry*, vol. 113, no. 5, pp. 423–427, 1956.

[8] A.-B. Hans Selye, "The Physiology and Pathology of Exposure to Stress. A Treatise Based on the Concept of the General Adaptation Syndrome and the Diseases of Adaptation," *The Indian Medical Gazette*, vol. 87, no. 9, p. 431, 1952. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5191521/

[9] B. S. McEwen, "Stress, Definitions and Concepts of," *Encyclopedia of Stress*, p. 653, 1 2007.

[10] A. Steptoe, "Control and Stress," *Encyclopedia of Stress*, pp. 568–573, 1 2007.

[11] N. Kessel, "Society, Stress and Disease, Volume I. The Psychosocial Environment and Psychosomatic Diseases," *British Journal of Industrial Medicine*, vol. 29, no. 3, p. 346, 1972. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1009441/

[12] B. S. McEwen and J. C. Wingfield, "The concept of allostasis in biology and biomedicine," *Hormones and behavior*, vol. 43, no. 1, pp. 2–15, 2003. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/12614627/

[13] P. X. Yuan, R. Zhou, N. Farzad, T. D. Gould, N. A. Gray, J. Du, and H. K. Manji, "Enhancing resilience to stress: the role of signaling cascades," *Techniques in the Behavioral and Neural Sciences*, vol. 15, no. PART 1, pp. 751–771, 1 2005.

[14] A. Rosenblueth and W. B. Cannon, "STUDIES ON CONDITIONS OF ACTIVITY IN ENDOCRINE ORGANS," *https://doi.org/10.1152/ajplegacy.1932.99.2.398*, vol. 99, no. 2, pp. 398–407, 1 1932. [Online]. Available: https://journals.physiology.org/doi/abs/10.1152/ajplegacy.1932.99.2.398

[15] C. Landis, "Walter B. Cannon. Bodily Changes in Pain, Hunger, Fear and Rage. (2nd ed., revised and enlarged.) New York: Appleton, 1929. Pp. xvi+404." *http://dx.doi.org/10.1080/08856559.1930.10532290*, vol. 38, no. 1-4, pp. 527–531, 2012. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/08856559.1930.10532290

[16] A. Steptoe, "Stress Effects, Overview," *Encyclopedia of Stress*, pp. 599–600, 1 2007.

[17] L. J. Luecken and K. S. Lemery, "Early caregiving and physiological stress responses," *Clinical Psychology Review*, vol. 24, no. 2, pp. 171–191, 5 2004.

[18] R. M. Nesse, S. Bhatnagar, and E. A. Young, "Evolutionary Origins and Functions of the Stress Response," *Encyclopedia of Stress*, pp. 965–970, 1 2007.

[19] H. Hofer and M. L. East, "Biological Conservation and Stress," *Advances in the Study of Behavior*, vol. 27, no. C, pp. 405–525, 1 1998.

[20] B. M. Wen, "Stress, adaptation, and disease. Allostasis and allostatic load," *Annals of the New York Academy of Sciences*, vol. 840, pp. 33–44, 5 1998. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/9629234/

[21] J. H. Markovitz and K. A. Matthews, "Platelets and coronary heart disease: Potential psychophysiologic mechanisms," *Psychosomatic Medicine*, vol. 53, no. 6, pp. 643–668, 1991. [Online]. Available: /record/1992-16955-001

[22] L. Saturday and H. Selye, "BRITISH MEDICAL JOURNAL STRESS AND THE GENERAL ADAPTATION SYNDROME*," 1950.

[23] B. M. Wen, “Stress, adaptation, and disease. Allostasis and allostatic load,” *Annals of the New York Academy of Sciences*, vol. 840, pp. 33–44, 5 1998. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/9629234/

[24] B. S. McEwen, “Protective and damaging effects of stress mediators,” *The New England journal of medicine*, vol. 338, no. 3, pp. 171–179, 1 1998. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/9428819/

[25] S. General, “Healthy people : the Surgeon General’s report on health promotion and disease prevention,” *undefined*, 1979.

[26] D. L. Delahanty and J. K. Cremeans-smith, “Behavioral Neuroimmunology,” *Encyclopedia of the Human Brain*, pp. 393–404, 1 2002.

[27] A. E. Moyer, J. Rodin, C. M. Grilo, N. Cummings, L. M. Larson, and M. Rebuffé-Scrive, “Stress-induced cortisol response and fat distribution in women,” *Obesity research*, vol. 2, no. 3, pp. 255–262, 1994. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/16353426/

[28] M. Maciejczyk, E. Żebrowska, and A. Chabowski, “Insulin Resistance and Oxidative Stress in the Brain: What’s New?” *International Journal of Molecular Sciences 2019, Vol. 20, Page 874*, vol. 20, no. 4, p. 874, 2 2019. [Online]. Available: https://www.mdpi.com/1422-0067/20/4/874/htmhttps://www.mdpi.com/1422-0067/20/4/874

[29] J. C. Da Silva, A. Garcez, G. H. Cibeira, H. Theodoro, and M. T. A. Olinto, “Relationship of work-related stress with obesity among Brazilian female shift workers,” *Public Health Nutrition*, vol. 24, no. 2, pp. 260–266, 2 2021. [Online]. Available: https://www.cambridge.org/core/journals/public-health-nutrition/article/relationship-of-workrelated-stress-with-obesity-among-brazilian-female-shift-workers/B6C2B53B452905DE3318C044E3E34CB2

[30] G. Siasos, S. A. Paschou, and D. Tousoulis, “Mitochondria and diabetes,” *Annals of Translational Medicine*, vol. 8, no. 6, pp. 262–262, 3 2020. [Online]. Available: /pmc/articles/PMC7186614/https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7186614/

[31] M. Kivimäki and A. Steptoe, “Effects of stress on the development and progression of cardiovascular disease,” *Nature Reviews Cardiology 2017 15:4*, vol. 15, no. 4, pp. 215–229, 12 2017. [Online]. Available: https://www.nature.com/articles/nrcardio.2017.189

[32] I. Kyrou, H. S. Randeva, and C. Tsigos, “Stress, Insulin Resistance, and Type 2 Diabetes,” *Stress: Neuroendocrinology and Neurobiology*, vol. 2, pp. 351–358, 1 2017.

[33] T. Vargas, R. E. Conley, and V. A. Mittal, “Chronic stress, structural exposures and neurobiological mechanisms: A stimulation, discrepancy and deprivation model of psychosis,” *International Review of Neurobiology*, vol. 152, pp. 41–69, 1 2020.

[34] Y. Li, Y. He, H. Fan, Z. Wang, J. Huang, G. Wen, X. Wang, Q. Xie, and P. Qiu, “Brain-derived neurotrophic factor upregulates synaptic GluA1 in the amygdala to promote depression in response to psychological stress,” *Biochemical Pharmacology*, vol. 192, p. 114740, 10 2021.

[35] S. Sydeman, "State-Trait Anxiety Inventory," *Encyclopedia of Personality and Individual Differences*, pp. 1–3, 2018. [Online]. Available: https://link.springer.com/referenceworkentry/10.1007/978-3-319-28099-8_950-1

[36] T. P. Oei, S. Sawang, Y. W. Goh, and F. Mukhtar, "Using the Depression Anxiety Stress Scale 21 (DASS-21) across cultures," *International Journal of Psychology*, vol. 48, no. 6, pp. 1018–1029, 12 2013. [Online]. Available: https://onlinelibrary.wiley.com/doi/full/10.1080/00207594.2012.755535https://onlinelibrary.wiley.com/doi/abs/10.1080/00207594.2012.755535https://onlinelibrary.wiley.com/doi/10.1080/00207594.2012.755535

[37] S. Cohen, T. Kamarck, and R. Mermelstein, "A global measure of perceived stress." *Journal of health and social behavior*, vol. 24, no. 4, pp. 385–396, 1983.

[38] R. Hicks, M. Bahr, D. F. P. differences, individual, and u. 2010, "The Occupational Stress Inventory-Revised: Confirmatory factor analysis of the original inter-correlation data set and model," *Elsevier*, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0191886909004358

[39] C. Kirschbaum, K. M. Pirke, and D. H. Hellhammer, "The 'Trier Social Stress Test'–a tool for investigating psychobiological stress responses in a laboratory setting," *Neuropsychobiology*, vol. 28, no. 1-2, pp. 76–81, 1993. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/8255414/

[40] D. H. Hellhammer, A. A. Stone, J. Hellhammer, and J. Broderick, "Measuring Stress," *Encyclopedia of Behavioral Neuroscience*, pp. 186–191, 1 2010.

[41] G. D. Sherman and P. H. Mehta, "Stress, cortisol, and social hierarchy," *Current Opinion in Psychology*, vol. 33, pp. 227–232, 6 2020.

[42] C. Schiweck, D. Piette, D. Berckmans, S. Claes, and E. Vrieze, "Heart rate and high frequency heart rate variability during stress as biomarker for clinical depression. A systematic review," *Psychological Medicine*, vol. 49, no. 2, pp. 200–211, 1 2019. [Online]. Available: https://www.cambridge.org/core/journals/psychological-medicine/article/abs/heart-rate-and-high-frequency-heart-rate-variability-during-stress-as-biomarker-for-clinical-dep
FA90B43DA846B6F88AC14EFBFD7EB01A

[43] S. P. Rowland, J. E. Fitzgerald, T. Holme, J. Powell, and A. McGregor, "What is the clinical value of mHealth for patients?" *npj Digital Medicine 2020 3:1*, vol. 3, no. 1, pp. 1–6, 1 2020. [Online]. Available: https://www.nature.com/articles/s41746-019-0206-x

[44] A. L. Goldberger, Z. D. Goldberger, and A. Shvilkin, "Essential Concepts: What Is an ECG?" *Goldberger's Clinical Electrocardiography*, pp. 2–5, 1 2018.

[45] J. Feher, "The Electrocardiogram," *Quantitative Human Physiology*, pp. 467–476, 1 2012. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/B9780123821638000505

[46] G. Lee, IEEE Computer Society., and Institute of Electrical and Electronics Engineers., "Detection of Acute Stress by Heart Rate Variability Using a Prototype Mobile ECG Sensor," *2006 International Conference on Hybrid Information Technology*, pp. 453–459, 11 2006.

[47] M. Malik, A. John Camm, J. Thomas Bigger, G. Breithardt, S. Cerutti, R. J. Cohen, P. Coumel, E. L. Fallen, H. L. Kennedy, R. E. Kleiger, F. Lombardi, A. Malliani, A. J. Moss, J. N. Rottman, G. Schmidt, P. J. Schwartz, and D. H. Singer, "Heart Rate Variability," *Circulation*, vol. 93, no. 5, pp. 1043–1065, 3 1996. [Online]. Available: https://www.ahajournals.org/doi/abs/10.1161/01.CIR.93.5.1043

[48] L. C. Johnson and A. Lubin, "SPONTANEOUS ELECTRODERMAL ACTIVITY DURING WAKING AND SLEEPING," *Psychophysiology*, vol. 3, no. 1, pp. 8–17, 1966. [Online]. Available: /record/1967-00144-001

[49] T. V. Damodaran, "Peripheral nervous system toxicity biomarkers," *Biomarkers in Toxicology*, pp. 169–198, 1 2014.

[50] W. Boucsein, "Principles of Electrodermal Phenomena," *Electrodermal Activity*, pp. 1–86, 2012.

[51] G. Turpin and T. Grandfield, "Electrodermal Activity," *Encyclopedia of Stress*, pp. 899–902, 1 2007.

[52] D. Leiner, A. Fahr, and H. Früh, "EDA Positive Change: A Simple Algorithm for Electrodermal Activity to Measure General Audience Arousal During Media Exposure," *http://dx.doi.org/10.1080/19312458.2012.732627*, vol. 6, no. 4, pp. 237–250, 10 2012. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/19312458.2012.732627

[53] H. Sequeira, P. Hot, L. Silvert, and S. Delplanque, "Electrical autonomic correlates of emotion," *International Journal of Psychophysiology*, vol. 71, no. 1, pp. 50–56, 1 2009.

[54] J. Scheirer, R. Fernandez, J. Klein, and R. W. Picard, "Frustrating the user on purpose: a step toward building an affective computer," *Interacting with Computers*, vol. 14, no. 2, pp. 93–118, 2 2002.

[55] G. Giannakakis, D. Grigoriadis, K. Giannakaki, O. Simantiraki, A. Roniotis, and M. Tsiknakis, "Review on psychological stress detection using biosignals," *IEEE Transactions on Affective Computing*, 2019.

[56] S. Alehagen, B. Wijma, U. Lundberg, and K. Wijma, "Fear, pain and stress hormones during childbirth," *http://dx.doi.org/10.1080/01443610400023072*, vol. 26, no. 3, pp. 153–165, 9 2009. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/01443610400023072

[57] S. Ray, "A Quick Review of Machine Learning Algorithms," *Proceedings of the International Conference on Machine Learning, Big Data, Cloud and Parallel Computing: Trends, Prespectives and Prospects, COMITCon 2019*, pp. 35–39, 2 2019.

[58] T. Mitchell, "Machine Learning (McGraw-Hill International Editions Computer Science Series): Tom M. Mitchell: 9780071154673," p. 414, 1997.

[59] R. Saravanan and P. Sujatha, "A State of Art Techniques on Machine Learning Algorithms: A Perspective of Supervised Learning Approaches in Data Classification," *Proceedings of the 2nd International Conference on Intelligent Computing and Control Systems, ICICCS 2018*, pp. 945–949, 3 2019.

[60] J. D. Kelleher, "Deep Learning (MIT Press Essential Knowledge Series)," 2019.

[61] W. D. Scherz, J. Baun, R. Seepold, N. M. Madrid, and J. A. Ortega, "A portable ECG for recording and flexible development of algorithms and stress detection," *Procedia Computer Science*, vol. 176, pp. 2886–2893, 1 2020.

[62] M. Zubair and C. Yoon, "Multilevel mental stress detection using ultra-short pulse rate variability series," *Biomedical Signal Processing and Control*, vol. 57, p. 101736, 3 2020.

[63] A. Cheema and M. Singh, "An application of phonocardiography signals for psychological stress detection using non-linear entropy based features in empirical mode decomposition domain," *Applied Soft Computing*, vol. 77, pp. 24–33, 4 2019.

[64] B. García-Martínez, A. Fernández-Caballero, R. Alcaraz, and A. Martínez-Rodrigo, "Assessment of dispersion patterns for negative stress detection from electroencephalographic signals," *Pattern Recognition*, vol. 119, p. 108094, 11 2021.

[65] A. Affanni, R. Bernardini, A. Piras, R. Rinaldo, and P. Zontone, "Driver's stress detection using Skin Potential Response signals," *Measurement*, vol. 122, pp. 264–274, 7 2018.

[66] I. K. Pribék, K. F. Szűcs, M. Süle, G. Grosz, E. Ducza, D. Vigh, E. Tóth, Z. Janka, J. Kálmán, Z. L. Datki, R. Gáspár, and B. Andó, "Detection of acute stress by smooth muscle electromyography: A translational study on rat and human," *Life Sciences*, vol. 277, p. 119492, 7 2021.

[67] N. Jin, X. Zhang, Z. Hou, I. Sanz-Prieto, and B. S. Mohammed, "IoT based psychological and physical stress evaluation in sportsmen using heart rate variability," *Aggression and Violent Behavior*, p. 101587, 3 2021.

[68] L. l. Chen, Y. Zhao, P. f. Ye, J. Zhang, and J. z. Zou, "Detecting driving stress in physiological signals based on multimodal feature analysis and kernel classifiers," *Expert Systems with Applications*, vol. 85, pp. 279–291, 11 2017.

[69] J. Rodríguez-Arce, L. Lara-Flores, O. Portillo-Rodríguez, and R. Martínez-Méndez, "Towards an anxiety and stress recognition system for academic environments based on physiological features," *Computer Methods and Programs in Biomedicine*, vol. 190, p. 105408, 7 2020.

[70] J. H. Kim, K. H. Lee, J. W. Lee, and K. S. Kim, "Semi-real-time removal of baseline fluctuations in electrocardiogram (ECG) signals by an infinite impulse response low-pass filter (IIR-LPF)," *The Journal of Supercomputing*

2018 74:12, vol. 74, no. 12, pp. 6785–6793, 9 2018. [Online]. Available: https://link.springer.com/article/10.1007/s11227-018-2608-y

[71] C. Lastre-Dominguez, Y. S. Shmaliy, O. Ibarra-Manzano, and M. Vazquez-Olguin, "Denoising and features extraction of ecg signals in state space using unbiased fir smoothing," *IEEE Access*, vol. 7, pp. 152 166–152 178, 2019.

[72] R. R. Sharma and R. B. Pachori, "Baseline wander and power line interference removal from ECG signals using eigenvalue decomposition," *Biomedical Signal Processing and Control*, vol. 45, pp. 33–49, 8 2018.

[73] A. Singhal, P. Singh, B. Fatimah, and R. B. Pachori, "An efficient removal of power-line interference and baseline wander from ECG signals by employing Fourier decomposition technique," *Biomedical Signal Processing and Control*, vol. 57, p. 101741, 3 2020.

[74] A. Uma and P. Kalpana, "ECG Noise Removal Using Modified Distributed Arithmetic Based Finite Impulse Response Filter," *Journal of Medical Imaging and Health Informatics*, vol. 11, no. 5, pp. 1444–1452, 4 2021.

[75] B. R. Manju and M. R. Sneha, "ECG Denoising Using Wiener Filter and Kalman Filter," *Procedia Computer Science*, vol. 171, pp. 273–281, 1 2020.

[76] N. T. Bui and G. S. Byun, "The Comparison Features of ECG Signal with Different Sampling Frequencies and Filter Methods for Real-Time Measurement," *Symmetry 2021, Vol. 13, Page 1461*, vol. 13, no. 8, p. 1461, 8 2021. [Online]. Available: https://www.mdpi.com/2073-8994/13/8/1461/htmhttps://www.mdpi.com/2073-8994/13/8/1461

[77] M. N. Rastgoo, B. Nakisa, F. Maire, A. Rakotonirainy, and V. Chandran, "Automatic driver stress level classification using multimodal deep learning," *Expert Systems with Applications*, vol. 138, 12 2019.

[78] J. Pan and W. J. Tompkins, "A Real-Time QRS Detection Algorithm," *IEEE Transactions on Biomedical Engineering*, vol. BME-32, no. 3, pp. 230–236, 1985.

[79] R. Costa, T. Winkert, A. Manhães, and J. P. Teixeira, "QRS Peaks, P and T Waves Identification in ECG," *Procedia Computer Science*, vol. 181, pp. 957–964, 1 2021.

[80] S. Katsigiannis and N. Ramzan, "DREAMER: A Database for Emotion Recognition Through EEG and ECG Signals from Wireless Low-cost Off-the-Shelf Devices," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 1, pp. 98–107, 1 2018.

[81] C. M. Durán Acevedo, J. K. Carrillo Gómez, and C. A. Albarracín Rojas, "Academic stress detection on university students during COVID-19 outbreak by using an electronic nose and the galvanic skin response," *Biomedical Signal Processing and Control*, vol. 68, p. 102756, 7 2021.

[82] M. Safa and A. Pandian, "Applying machine learning algorithm to sensor coupled IoT devices in prediction of cardiac stress – An integrated approach," *Materials Today: Proceedings*, 3 2021.

[83] R. Ahuja and A. Banga, "Mental Stress Detection in University Students using Machine Learning Algorithms," *Procedia Computer Science*, vol. 152, pp. 349–353, 1 2019.

[84] F. Patlar Akbulut, B. Ikitimur, and A. Akan, "Wearable sensor-based evaluation of psychosocial stress in patients with metabolic syndrome," *Artificial Intelligence in Medicine*, vol. 104, p. 101824, 4 2020.

[85] M. Sevil, M. Rashid, I. Hajizadeh, M. R. Askari, N. Hobbs, R. Brandt, M. Park, L. Quinn, and A. Cinar, "Discrimination of simultaneous psychological and physical stressors using wristband biosignals," *Computer Methods and Programs in Biomedicine*, vol. 199, p. 105898, 2 2021.

[86] A. Priya, S. Garg, and N. P. Tigga, "Predicting Anxiety, Depression and Stress in Modern Life using Machine Learning Algorithms," *Procedia Computer Science*, vol. 167, pp. 1258–1267, 1 2020.

[87] W. Hadi, N. El-Khalili, M. AlNashashibi, G. Issa, and A. A. AlBanna, "Application of data mining algorithms for improving stress prediction of automobile drivers: A case study in Jordan," *Computers in Biology and Medicine*, vol. 114, p. 103474, 11 2019.

[88] S. Carreiro, K. K. Chintha, S. Shrestha, B. Chapman, D. Smelson, and P. Indic, "Wearable sensor-based detection of stress and craving in patients during treatment for substance use disorder: A mixed methods pilot study," *Drug and Alcohol Dependence*, vol. 209, p. 107929, 4 2020.

[89] E. Resendiz-Ochoa, I. A. Cruz-Albarran, M. A. Garduño-Ramon, D. A. Rodriguez-Medina, R. A. Osornio-Rios, and L. A. Morales-Hernández, "Novel expert system to study human stress based on thermographic images," *Expert Systems with Applications*, vol. 178, p. 115024, 9 2021.

[90] H. Jebelli, S. Hwang, and S. H. Lee, "EEG-based workers' stress recognition at construction sites," *Automation in Construction*, vol. 93, pp. 315–324, 9 2018.

[91] L. Xia, A. S. Malik, and A. R. Subhani, "A physiological signal-based method for early mental-stress detection," *Biomedical Signal Processing and Control*, vol. 46, pp. 18–32, 9 2018.

[92] A. Arsalan and M. Majid, "Human stress classification during public speaking using physiological signals," *Computers in Biology and Medicine*, vol. 133, p. 104377, 6 2021.

[93] R. Zhou, C. Wang, P. Zhang, X. Chen, L. Du, P. Wang, Z. Zhao, M. Du, and Z. Fang, "ECG-based biometric under different psychological stress states," *Computer Methods and Programs in Biomedicine*, vol. 202, p. 106005, 4 2021.

[94] A. Asif, M. Majid, and S. M. Anwar, "Human stress classification using EEG signals in response to music tracks," *Computers in Biology and Medicine*, vol. 107, pp. 182–196, 4 2019.

[95] E. Perez-Valero, M. A. Lopez-Gordo, and M. A. Vaquero-Blasco, "EEG-based multilevel stress classification with and without smoothing filter," *Biomedical Signal Processing and Control*, vol. 69, p. 102881, 8 2021.

[96] N. Hakimi, A. Jodeiri, M. Mirbagheri, and S. K. Setarehdan, "Proposing a convolutional neural network for stress assessment by means of derived heart rate from functional near infrared spectroscopy," *Computers in Biology and Medicine*, vol. 121, p. 103810, 6 2020.

[97] W. Sanchez, A. Martinez, Y. Hernandez, H. Estrada, and M. Gonzalez-Mendoza, "A predictive model for stress recognition in desk jobs," *Journal of Ambient Intelligence and Humanized Computing*, 2018.

[98] B. Hwang, J. You, T. Vaessen, I. Myin-Germeys, C. Park, and B. T. Zhang, "Deep ECGNet: An Optimal Deep Learning Framework for Monitoring Mental Stress Using Ultra Short-Term ECG Signals," *https://home.liebertpub.com/tmj*, vol. 24, no. 10, pp. 753–772, 10 2018. [Online]. Available: https://www.liebertpub.com/doi/abs/10.1089/tmj.2017.0250

[99] P. Arpaia, N. Moccaldi, R. Prevete, I. Sannino, and A. Tedesco, "A Wearable EEG Instrument for Real-Time Frontal Asymmetry Monitoring in Worker Stress Analysis," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 10, pp. 8335–8343, 10 2020.

[100] H. Jebelli, M. M. Khalili, and S. Lee, "Mobile EEG-Based Workers' Stress Recognition by Applying Deep Neural Network," *Advances in Informatics and Computing in Civil and Construction Engineering*, pp. 173–180, 2019.

[101] A. Golgouneh and B. Tarvirdizadeh, "Fabrication of a portable device for stress monitoring using wearable sensors and soft computing algorithms," *Neural Computing and Applications*, vol. 32, no. 11, pp. 7515–7537, 6 2020. [Online]. Available: https://link.springer.com/article/10.1007/s00521-019-04278-7

[102] "Rapid GUI Programming with Python and Qt," 2007.

[103] M. J. Nelson and A. K. Hoover, "Notes on Using Google Colaboratory in AI Education," *Annual Conference on Innovation and Technology in Computer Science Education, ITiCSE*, pp. 533–534, 6 2020.

[104] J. A. Healey and R. W. Picard, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Transactions on Intelligent Transportation Systems*, vol. 6, no. 2, pp. 156–166, 6 2005.

[105] P. Schmidt, A. Reiss, R. Duerichen, and K. Van Laerhoven, "Introducing WeSAD, a multimodal dataset for wearable stress and affect detection," *ICMI 2018 - Proceedings of the 2018 International Conference on Multimodal Interaction*, pp. 400–408, 10 2018.

[106] "PhysioNet." [Online]. Available: https://physionet.org/

[107] D. Singh and B. Singh, "Investigating the impact of data normalization on classification performance," *Applied Soft Computing*, vol. 97, p. 105524, 12 2020.

[108] N. El Haouij, J. M. Poggi, R. Ghozi, S. Sevestre-Ghalila, and M. Jaïdane, "Random forest-based approach for physiological functional variable selection for driver's stress level classification," *Statistical Methods & Applications 2018 28:1*, vol. 28, no. 1, pp. 157–185, 2 2018. [Online]. Available: https://link.springer.com/article/10.1007/s10260-018-0423-5

[109] R. G. Mantovani, T. Horváth, R. Cerri, S. B. Junior, J. Vanschoren, and A. C. P. d. L. F. de Carvalho, "An empirical study on hyperparameter tuning of decision trees," 12 2018. [Online]. Available: https://arxiv.org/abs/1812.02207v2

[110] S. Yadav and A. Hashmi, "An Investigation of Occupational stress Classification by using Machine Learning Techniques," *International Journal of Computer Sciences and Engineering*, vol. 6, no. 6, pp. 842–850, 6 2018.

[111] M. Vigier, B. Vigier, E. Andritsch, and A. R. Schwerdtfeger, "Cancer classification using machine learning and HRV analysis: preliminary evidence from a pilot study," *Scientific Reports 2021 11:1*, vol. 11, no. 1, pp. 1–12, 11 2021. [Online]. Available: https://www.nature.com/articles/s41598-021-01779-1

[112] U. S. Reddy, A. V. Thota, and A. Dharun, "Machine Learning Techniques for Stress Prediction in Working Employees," *2018 IEEE International Conference on Computational Intelligence and Computing Research, ICCIC 2018*, 12 2018.

,bib/references

Appendices

Appendix

GitHub Repository

All the codes and work carried out for the present research project can be found at the following GitHub repository:

`https://github.com/paolavasquez98/Stress_Detection`