Master's Degree in Mechatronic Engineering

AI for health: Using AI to identify stress from wearable devices data

Supervisors

Prof. Eros Gian Alessandro Pasero

Prof. Maria A. Zuluaga

Candidate

Bianca D'Alpaos

2022/2023

Master's Degree in Mechatronic Engineering

Master's Degree Thesis

AI for health: Using artificial intelligence to identify stress from wearable devices data

Supervisors

Prof. Eros Gian Alessandro Pasero

Prof. Maria A. Zuluaga

Candidate

Bianca D'Alpaos

2022/2023

Abstract

Stress is an escalated psycho-physiological state of the human body emerging in response to a challenging event or a demanding condition and since it is one of the major issues in modern society, there is a growing interest in developing methods that make automatic detection possible. To this end, the adoption of wearable technology coupled with the implementation of machine learning (ML) techniques are emerging as an interesting approach to develop non-invasive stress detection systems.

The present work investigates the coupled use of ML and biosignals collected from wearables in a controlled environment to study the feasibility of non-invasive stress detection systems. This thesis uses a dataset that was obtained by acquiring three different biosignals from subjects during a stress test by using the *BiosignalsPlux* platform: Electrocardiogram (ECG), Electrodermal Activity (EDA) and Respiration Signal.

The collected data was then pre-processed and prepared by applying a features extraction and features selection algorithms in order to work with tabular data. The selected features are prepared for use in multiple AI algorithms, including Random Forest, XGBoost, Neural Network and Support Vector Machine.

The results of the thesis showed that machine learning algorithms were successful in detecting stress from biosignals with high accuracy level and it also shows that the use of multiple biosignals is more effective compared to single-signal-based system. XGBoost was the best performing algorithm achieving an accuracy of 84% for the binary classification (stress vs. no stress) and 70% for three class classification (no stress, medium stress and high stress).

The findings of this work may have important implications for the development of non-invasive wearable devices for early stress detection and management.

Table of Contents

List of Tables

List of Figures

Acronym

ECG
: Electrocardiogram

EEG
: Electroncephalogram

EDA
: Electrodermal Activity

SCL
: Skin Conductance Level

SCR
: Skin Conductance Response

BR
: Breathing Rate

RESP
: Respiration Signal

EMG
: Electromyography

PPG
: Photoplethysmography

CPT
: Cold Pressor Test

TSST
: Trier Social Stress Test

SCWT
: Stroop Color-Word Test

MIST
: Montreal Imaging Stress Test

MAST
: Maastricht Acute Stress Test

MAT
: Mental Arithmetic Task

LPF
: Low-Pass Filter

HPF
: High-Pass Filter

MA
: Motion Artifact

PLI
: Powerline Interference

FIR
: Finite Impulse Response

IIR
: Infinite Impulse Response

EMD
: Empirical Mode Decomposition

IMF
: Intrinsic Mode Function

DWT

Discrete Wavelet Transform

ICA

Independent Component Analysis

PCA

Principal Component Analysis

RL

Reinforcement Learning

SVM

Support Vector Machine

ANNs

Artificial Neural Networks

FPR

False Positive Rate

FNR

False Negative Rate

ROC

Receiver operating characteristic

AUC

Area Under the Curve

Chapter 1

Introduction

1.1 Review of the Presented Work

This thesis aims to assess the effectiveness of a stress detection algorithm by analyzing biosignals obtained from wearable sensors.

The initial phase of the study was dedicated to acquiring and creating a dataset to be utilized in the research. Data collection involved forty healthy subjects who were instructed to complete a stress test customized for the study. The stress test involved the acquisition of three distinct biosignals, namely Electrocardiogram (ECG), Electrodermal Activity (EDA), and Respiration Signal (RESP), which were all captured using the *Biosignalsplux* platform.

Following the acquisition of all signals and completion of the dataset, features were extracted from the signals and employed to train a machine learning model. The model was trained to classify the level of stress of a subject solely based on the extracted features. The classification could be either binary, distinguishing between "no stress" or "stress", or three-class, classifying stress as "no stress", "medium stress", or "high stress".

1.2 Stress and Anxiety

1.2.1 General definition of stress and anxiety

Stress can be defined as any type of change that causes physical, emotional or psychological strain. The framework of the stress system indicates that stress includes two types: eustress (good stress) and distress (bad stress), depending on the way stressful situations are handled. As the *American Psychology Association* says the "Distress is the negative stress response, often involving negative affect and physiological reactivity: a type of stress that results from being overwhelmed by demands, losses, or perceived threats. It has a detrimental effect by generating physical and psychological maladaptation and posing serious health risks for individuals" and "Eustress is the positive stress response, involving optimal levels of stimulation: a type of stress that results from challenging but attainable and enjoyable or worthwhile tasks (e.g., participating in an athletic event, giving a speech). It has a beneficial effect by generating a sense of fulfillment or achievement and facilitating growth, development, mastery, and high levels of performance" [1].

Type of stressor	Description and Examples
Physical	Strain on a body Ex: Injury, illness, pain, travel, infection, excess alcohol
Psychological	Anything interpreted as threatening or challenging for mind Ex: Money problem, exams, loss of employment, excessive worrying
Environmental	Associated with surrounding Ex: Noise, crowd, air quality, light, insects, temperature variation, war and disaster.
Psychosocial	Associated with social situation Ex: Divorce, unwanted change of residence, prolonged illness, highly competitive work situation.

Figure 1.1: Stressor Types and Examples

Stress that has a positive impact is marked by an increase in pulse rate but without any underlying feeling of threat or fear. Stress whit a negative impact is also known as the "fight-or-flight" response, which is the reaction of the body's symphathetic nervous system that reacts to a stressor by producing larger quantities of chemicals like cortisol, adrenaline, and nor adrenaline. This increases the heartbeat, causes breathlessness, and sharpens the senses. Stress can be either a triggering or aggravating factor for many diseases and pathological conditions, it can cause structural changes in the brain with long-term effects on the nervous system, it may increase the development of cardiac arrhythmias and can also alter the functional physiology of the intestine. Therefore, a greater appreciation to the significant role that stress may play in various diseases is needed. To prevent stress-related issues, it may be helpful to detect them in the burgeoning stages by using continuous monitoring through wearable devices and machine learning techniques.

1.2.2 Ground Truth of Stress

Some studies establish stress ground truth using the person's perceived stress as expressed in self-report ratings or scores from questionnaires. In other studies, stress ground truth is determined as a neutral or reference period and the stress state is determined by the presentation of stressors or the exposure to stressful situations. The baseline can be considered as the person's relaxed state achieved through relaxation videos or following relaxation instructions of a psychologist [2]. Ground truth of stress can be established from biosignals or biomarkers that are considered reliable for stress level identification. In some studies, the salivary cortisol levels [3] and the SDNN cardiovascular measure were employed to define stress groups or stressors. In general, stress ground truth determination is not a straightforward procedure mainly in real world conditions. The assessment through the use of self reports or ratings may have wide inconsistencies and involve subjective bias. In addition, stress self-assessment may not illustrate unconscious or subconscious psychological processes. Measurement of stress responses in different people requires the formulation of an objective measure framework [2]. Valence and arousal are two important parameters that are highly associated with an individual's emotional states. Arousal (or intensity) is the level of autonomic activation that an event creates, and ranges from calm (or low) to excited (or high) [4]. Arousal refers to a physiological activity dimension, ranging from quiet to active mood, is linked to the excitement level of individuals. Valence, on the other hand, is the level of pleasantness that an event generates and is defined along a continuum from negative to positive [4]. Valence refers to another physiological activity dimension orthogonal to arousal, ranging from misery to pleasant [3]. The psychological evidence suggests that these two dimensions are intercorrelated. Likewise, stress has both positive and negative interpretations based on its origin. High arousal and negative valence are characteristics of emotional stress, an affective state induced by threatening stimuli. High arousal and negative valence are also characteristic of acute affective states: the specific emotions of anger, disgust, and fear [5].

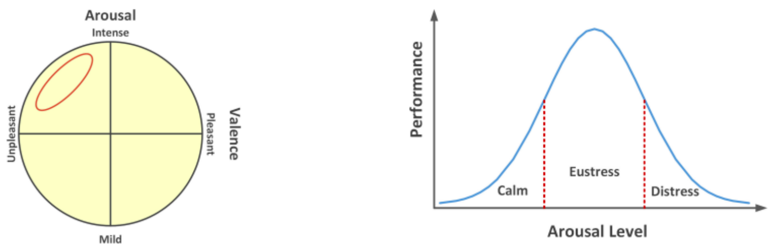


Figure 1.2: Correlation between valence, arousal and human stress [2].

Objective measures of stress include physiological and physical measures. Physiological measures of stress need sensors to be connected to the human body at some specified location e.g., EEG, ECG, and EDA, whereas, in the case of physical sensors, the measurement can be done at a distance from the subject without the need of any physical contact. Objective measures of stress are free from human intervention and hence cannot be biased like a subjective questionnaire, this being a major benefit of objective measures over the subjective assessment of stress.

1.3 Biosignals related to stress

When individuals experience stressful conditions, the autonomic nervous system is activated, thus causing an imbalance between the sympathetic and parasympathetic systems. Therefore, signs of stress often include physiological and physical reactions that stem from the nervous system. Physical biosignals are measures of body deformation as the result of muscle activity. These include:

- Pupil size,
- Eye movements or blinks,

- Respiration,
- Facial expressions
- Body and extremity semivoluntary position/movements.

Physiological signals are more directly related with body vital functions, such as [2]:

- Electrocardiogram (ECG),
- Electroencephalogram (EEG),
- Electrodermal Activity (EDA),
- Respiration and Breath Rate (BR),
- Photoplethysmography (PPG).

In the following subsections, biosignals are categorized according to their source on the body into those recorded from the head, the heart, and the remaining body parts.

1.3.1 Heart Rate (HR)

The electrocardiogram (ECG) is a a medical test that records the electrical activity of the heart. It is a non-invasive procedure that involves placing small electrodes on the chest, arms and legs of a patient, the electrodes detect the electrical signals that are generated by the heart's cells as they contract and relax, and transmit them to a machine that records the data. An ECG can be described also as the time-series signal of the electrical activity of the heart, where each heartbeat is displayed as a series of electrical waves. The changes in electrical potential difference (voltage) during depolarization and repolarization of the myocardial fibers are recorded by electrodes positioned on the surface of the chest, this voltage is the amplitude of the ECG signal. The ECG, as it is shown in Figure 1.3 , consists of three basic waves: P, QRS, and T. These waves correspond to the far field induced by specific electrical phenomena on the cardiac surface, namely, the atrial depolarization, P, the ventricular depolarization, QRS complex, and the ventricular repolarization, T.

Differences in the speed of wavefront propagation through the cardiac cycle are reflected by different frequencies content of ECG waves. The content of T wave lays mostly within a range from zero (DC) to 10 Hz. The content of P wave is characterized by 5-30 Hz frequencies. The content of QRS usually contains within 8-50 Hz frequencies while abnormal ventricular conduction is characterized by high frequencies (above 70Hz), forming notches on the QRS [7]. Stress leads to the activation of the Sympathetic Nervous System (SNS), resulting in the increase of heart rate and its force of contraction. As a result, the amount of blood circulates faster through the body to deliver more oxygen to the organs and skeletal muscles as an attempt to eliminate the stressor [2].

During times of stress, the body releases hormones like adrenaline and cortisol, which can increase heart rate, blood pressure, and cardiac output.

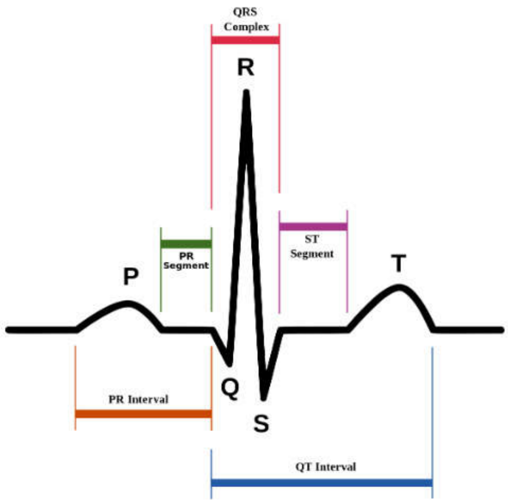


Figure 1.3: Typical ECG signal with characteristics peaks P,Q,R,S,T [6]

These changes in the body's physiology can be detected by an ECG, which will show an increase in the heart rate, changes in the shape of the waves on the ECG tracing, and other abnormalities. Heart rate (HR) is the most widely adopted and straightforward measure to estimate levels of stress, it is the number of beats per minute, heart rate increases significantly during states of stress. HR is one of the most widely used measures of human stress available in the literature as shown in 1.4. Heart rate is defined as the number of heartbeats in one minute (bpm). The RR interval of the ECG signal, which is defined as the interval between consecutive heartbeats has an inverse relationship with the heart rate of a person. Therefore, an ECG can be a useful tool in evaluating the impact of stress on the body and detecting potential cardiac problems related to stress.

Feature	Studies	↑	↓	=
HR	23 [109, 131, 132, 151, 154, 160, 165, 180, 182, 187-200]	18	0	5
STD HR	1 [198]	0	0	1
RR	8 [180, 198, 200-205]	0	6	2
SDNN	12 [180, 187, 193, 194, 197, 198, 200, 201, 203-206]	1	7	4
RMSSD	6 [187, 190, 197, 198, 203, 204]	0	5	1
NN50	2 [187, 200]	0	2	0
pNN50	6 [116, 194, 198, 200, 203, 207]	0	6	0
HRV triangular	2 [198, 200]	0	1	1
Total power	4 [133, 197, 204, 206]	0	4	0
VLF	3 [187, 204]	0	0	3
LF	12 [180, 187, 192-195, 197, 199, 203-205, 208]	5	3	4
HF	14 [180, 187, 192-194, 197, 199, 201, 203-205, 208-210]	1	6	7
LF/HF	17 [165, 180, 187, 188, 192-194, 198-200, 202-204, 207-210]	10	0	7
VLF relative	2 [187, 188]	2	0	0
LF relative	8 [187, 188, 200-202, 204, 208]	4	1	3
HF relative	7 [187, 200-202, 204, 208]	0	4	3
SD1	1 [211]	0	0	1
SD2	1 [211]	0	1	0
D2	2 [211]	0	2	0
BR	5 [165, 180, 193, 199, 204]	2	0	3
SBP	15 [129, 132, 151, 154, 160, 188-191, 195, 201, 206, 212-214]	15	0	0
DBP	15 [129, 132, 151, 154, 160, 188-191, 195, 201, 209, 212-214]	15	0	0
BP HF	1 [206]	1	0	0
ApEn	1 [211]	0	1	0
SampEn	1 [192]	0	0	1

↑: significant increase (p<0.05) during stress
↓: significant decrease (p<0.05) during stress

Figure 1.4: Main Features Related to stress detection for ECG signal [2]

Heart Rate is the most widely adopted and straightforward measure to estimate levels of stress. Alternative, the mean RR interval can be used, having an inverse relationship with heart rate. According to [2], the heart rate is the most important features used to detect stress being a reliable measure for the arousal part of stress. The HRV, instead, has a chaotic behaviour in states of anger, anxiety or sadness whose rhythmicity can be described by a measure known as cardiac coherence. SDNN and RMSSD are reduced during stress consitions.

In frequency domain, instead, the LF band is modulated by both sympathetic and parasympathetic activity, while the HF band corresponds only to parasympathetic activity. Thus, the ratio LF/HF is considered a distinctive approach for the sympathetic modulation and is the more prominent feature in frequency domain increasing during stress condition.

1.3.2 Electroencephalogram (EEG)

The electroencephalogram (EEG) is a recording of the electrical activity of the brain from the scalp. It is a non-invasive procedure that involves placing small electrodes on the scalp of a patient. The electrodes detect the electrical signals that are generated by the brain's neurons as they communicate with each other, and transmit them to a machine that records the data. The recorded data is represented graphically as a series of waves that correspond to different states of brain activiy. These waveforms reflect the cortical electrical activity, which is a widely used technique to estimate changes in neurophysiological activity associated with external stimuli and/or with the performance of specific tasks [2] and has a typical shape as shown in Figure 1.5.

An EEG has a frequency content ranging from 0.01 to around 100 Hz and varies from a few microvolts to approximately 100 µV, the slow components around 0.01 Hz correspond to slow cortical potentials that in clinical routine are usually not recorded (they are filtered out); however, they may be of interest for brain-computer interfaces. The most frequently used method to classify EEG waveforms is by the frequency, the most commonly studied waveforms include [8]:

- *Delta* with frequency range from 0.5 to 4Hz, it is physiologically seen in deep sleep and is prominent in the frontocentral head regions;

- *Theta* with frequency range from 4 to 7Hz, this is the rhythm which is brought on by drowsiness as well as early stages of sleep and it is most prominent in the fronto-central head regions;

- *Alpha* with frequency range from 8 to 12 Hz, the posterior dominant alpha rhythm is characteristically present during relaxation or conditions with minimal cognitive demands or emotional strain in the occipital head region. It is the defining feature of the normal background rhythm of the adult EEG recording;

- *Sigma* with frequency range from 12 to 16 Hz, this activity is also known as "sleep spindles", they may be slow (12 to 14Hz) or fast (14 to 16Hz) and are seen most prominently in the fronto-central head regions;

- *Beta* with frequency range from 13 to 30 Hz, which is the most frequently seen rhythm in normal adults and children during conditions with significant processing demands or high alertness levels, it is most prominent in the frontal and central head regions and attenuates as it goes posteriorly.

Stress conditions are considered to decrease the alpha activity and increase the beta activity waves. For instance, certain conditions such as sleep deprivation and performance of a cognitive task may only affect the lower alpha power (8-10.5 Hz) whereas the higher alpha power (11-12 Hz) may actually be reduced in both conditions. Power of higher frequency rhythms (between 30 and 70 Hz) may provide a sensitive index of stress response magnitude [2]. Alpha band activity is dominant in the relaxation phase when the cognitive demands are minimal whereas, on the contrary, it has been found that situations involving high strain or alertness, beta-band activity is found to be significant. Stress is found to be correlated to the beta wave in the temporal part of the brain. When a subject is having a negative mood or depression the alpha and beta band activity is dominant.

Prolonged exposure to stress can also have negative effects on brain function and lead to the development of various neurological conditions, such as anxiety, depression, and post-traumatic stress disorder (PTSD). These conditions can also be detected and diagnosed through changes in the electrical patterns recorded by an EEG. The EEG can provide rich and reliable information on the factors that cause stress, since continuous recording of EEG signals on a real-time basis can greatly contribute to an understanding of field subjects' stress patterns because these signals are rapidly responding to many different stressors. Therefore, EEG can be very useful in studying field stressors or developing online physiological monitoring systems.

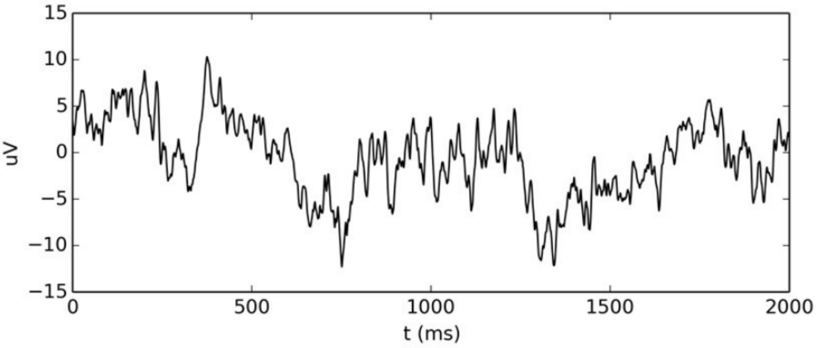


Figure 1.5: Typical raw EEG acquired with *biosignalplux* [9]

1.3.3 Electrodermal Activity (EDA)

Sweat glands produce moisture through pores towards the surface of the skin, whenever they are triggered. When the balance of positive and negative ions in this secreted fluid changes, the eletrical current flows more readily. This results in decreased skin resistance, or in other words, increased skin conductance. The electrodermal activity (EDA), also known as galvanic skin response (GSR), is a physiological measurement of electricity flow through the skin. An EDA sensor measures the change in skin conductance, these changes are caused by alterations in sweat secretion and sweat gland activity as a result of changing sympathetic nervous system activity, even moderate amounts of sweating that are not observable at the skin surface can alter skin electrical conductivity [2].

EDA can be measured using electrodes placed on the skin, typically on the fingers or palms of the hands, and is commonly used in research and clinical settings as a physiological marker of emotional arousal and stress and has a shape as shown in Figure 1.8. EDA has been used to study a range of phenomena, including emotional responses, cognitive processing and physiological reactivity to stress. The measurements of EDA signals are composed of the convolution of two signals, as shown in Figure 1.6:

- Skin conductance level (SCL), tonic part: this signal relates to the slower acting components and background characteristics of the signal;
- Skin Conductance Response (SCR), phasic part: this signal refers to the faster changing elements of a signal.

The SCL signal establishes the base level of the signal. Changes in the SCL are thought to reflect general changes in autonomic arousal, which is a physiological response that occurs when the autonomic nervous system (ANS) is activated, resulting in changes in bodily functions that are not under voluntary control. The tonic level of EDA signal, can differ significantly across different individuals. Due to this, the actual tonic level on its own is not completely informative.

The SCR is closely related to the activity of the sweat motor system which, at the same time, is closely associated with the parasympathetic nervous system

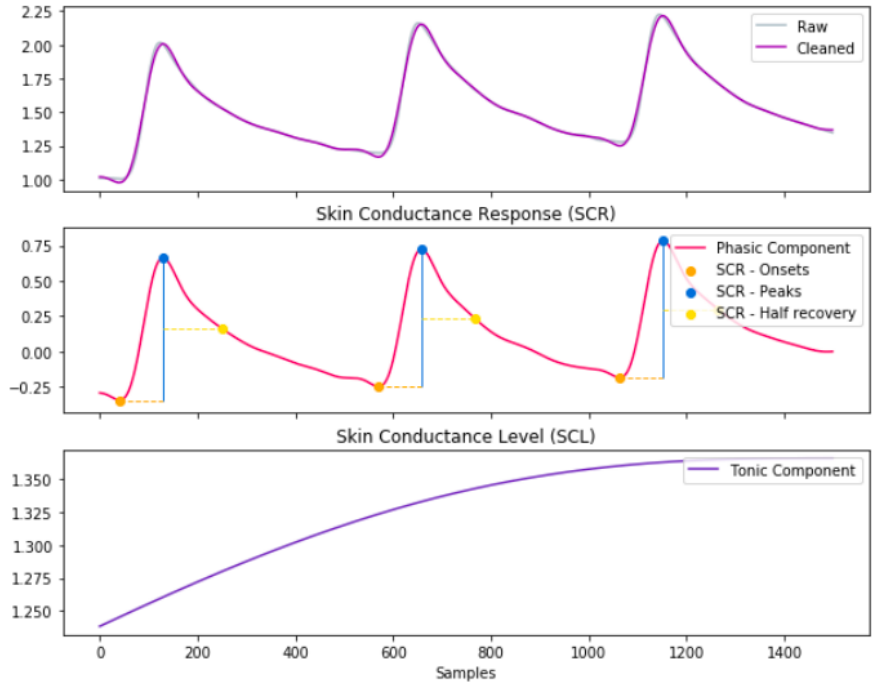


Figure 1.6: Typical components of EDA signal.

[10]. The phasic component is sensitive to specific emotionally arousing stimulus events; these bursts can occur between 1-5 seconds after the onset of emotional stimuli. The pattern of the skin response data is distinct according to the state of the person and is considered as one of the reliable stress measurement methods. SCR part of EDA increases when encountered with an emotionally arousing situation.

There is a well-established correlation between electrodermal activity and stress. Whenever the person is under stress, their sympathetic nervous system becomes activated, which lead to an increase of the moisture in the human skin resulting in an increase in the SCL and SCR part of the electrodermal acitivy.

Feature	Studies	↑	↓	=
SCR	9 [7], [64], [127], [180], [183], [184], [185], [186], [187]	7	0	2
SCL	5 [116], [180], [181], [182], [183]	5	0	0
Ns-SCR	1 [180]	1	0	0
SCR frequency	0	0	0	0
SCR amplitude	2 [7], [116], [187]	1	0	1
SCR latency	0	0	0	0
SCR rise time	1 [187]	1	0	0
SCR 50% recovery time	2 [187]	1	0	1

↑: significant increase ($p < 0.05$) during stress.
↓: significant decrease ($p < 0.05$) during stress.
=: no significant difference.

Figure 1.7: Main Features Related to stress detection for EDA signal [2]

When a person is under stress, the two most relevant features of the EDA are both tonic part SCL and phasic part SCR which increases due to skin moisture increase. According to [2], SCL was considered the most effective stress correlate among features from HRV, RSP and EMG. Other common features that are used in stress studies are the SCR frequency, SCR amplitude, SCR latency, SCR rise time, SCR half recovery and SCR response onset.

For example, researchers have found that people who experience high levels of chronic stress, such as those with post-traumatic stress disorder (PTSD), tend to have higher levels of EDA than those who do not. In addition, EDA has been used in biofeedback interventions to help individuals manage stress and anxiety. By monitoring their EDA, individuals can learn to identify and regulate their physiological responses to stress, which can improve their overall well-being.

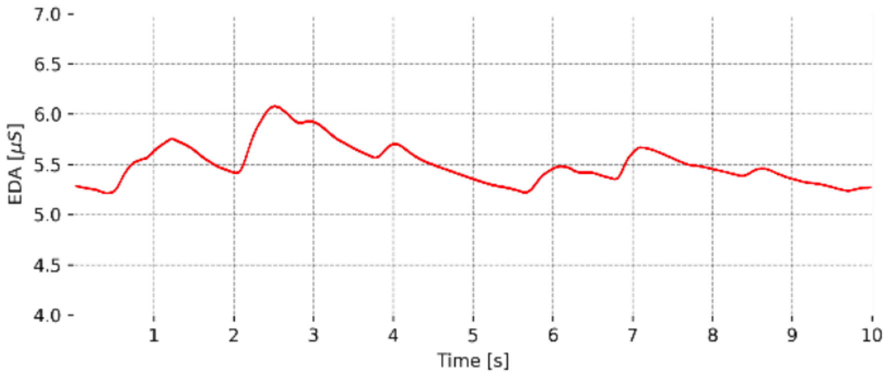


Figure 1.8: Typical raw EDA acquired with *biosignalplux* [11]

Although one of the main purposes of sweating is thermoregulation, sweating is also triggered whenever a person is exposed to a stimulus, such as emotionally loaded images. This type of sweating is called emotional sweating. Sweat secretion, which reflects the changes in arousal, is driven unconsciously by the automatic nervous system (ANS) in order to meet behavioral requests. When a person is under stress, both tonic part SCL and phasic part SCR increases due to skin moisture increase. Even the expectation of a painful or stressful event can elicit increases in the EDA. The peaks of SCR usually appear between 1.5 and 6.5 seconds after the onset of stressor stimuli. [2] The response of the human skin is not under human conscious control and is dependent on the changes in the sweating pattern of a subject and thus reflects the behavior of the sympathetic nervous system. GSR measurement locations are part of the body with a large number of sweat glands. There exist a variety of possible locations on the human body for the measurement of GSR.

1.3.4 Respiration and Breath Rate (BR)

The respiration signal refers to the physiological process of breathing, which involves inhaling oxygen-rich air and exhaling carbon dioxide-rich air. Respiration is controlled by the respiratory center in the brainstem, which receives input from various sensors in the body that detect changes in oxygen and carbon dioxide levels in the blood, as well as other factors such as pH and temperature. In general, this signal may have the shape as shown in Figure 1.9.

The breath rate is the rate at which breathing occurs, which is usually measured by manually counting chest wall movements per minute. The respiratory center is located in the medulla oblongata (lower brainstem) and is involved in the minute-to-minute control of breathing [12]. This rate can vary depending on various factors, such as age, physical activity, and overall health. A normal respiratory rate for adults at rest is typically between 12 and 20 breaths per minute, while in children it can range from 20 to 30 breaths per minute.

Under stress conditions, breath rate generally increases with emotional arousal and decreases with relaxation, while tense situations may cause momentary interruptions in breath. Negative emotions such as stress are linked to irregularities in the respiration pattern, increase of the minute volume, the shift from abdominal to thoracic breathing and faster and shallower breathing [2]. Oxygen consumption rate can be extracted from the respiratory rate of the person and is considered a reliable measure of human stress too because the oxygen demand is increased under stress [13].

For the Respiration Signal (RESP), breath rate and breath depth (amplitude) are the most common measures of respiration. Under stress conditions, breath rate generally increases with emotional arousal and decreases with relaxation, while tense situations may bause momentary interruption in breath. The breath rate (BR) significant changes during stress as shown in 1.4. Besides, negative emotions such as stress are linked to irregularities in the respiration pattern, increase of the minute volume, the shift from abdominal to thoracic breathing and faster and shallower breathing [2].

During times of stress, the body's "fight or flight" response is activated and the respiration rate may increase. This increased respiration rate can be detected through changes in the respiration signal, such as an increase in the frequency and depth of breaths. These changes are designed to provide the body with the extra oxygen needed to respond to the stressful situation. However, prolonged exposure to stress can also have negative effects on respiratory function, and lead to the development of various respiratory conditions, such as asthma, COPD, and hyperventilation syndrome. These conditions can also be detected and diagnosed through changes in the respiration signal.

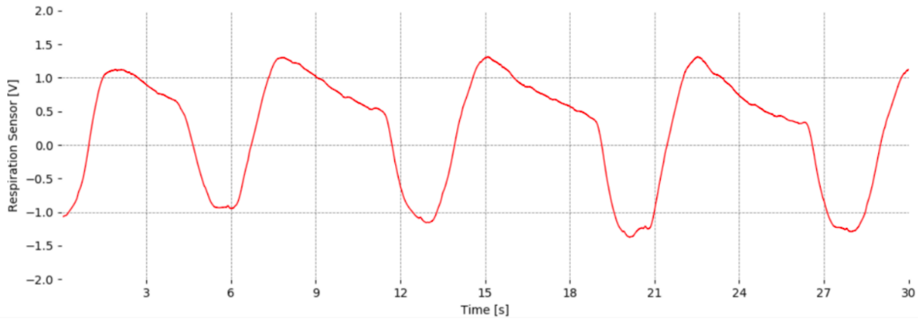


Figure 1.9: Raw Respiration signal acquired with *biosignalplux* [14]

Breathing rate (BR) can be estimated by extracting respiratory signals from the electrocardiogram (ECG) or photoplethysmogram (PPG) using either feature- or filter-based techniques. There are three idealised types of respiratory modulation of the ECG and PPG: baseline wander (BW), amplitude modulation (AM), and frequency modulation (FM). If the amplitude of the respiratory signal is too small compared to the underlying noise, then the signal may not be distinguishable from the noise, preventing the precise estimation of BR. The morphology of the ECG is modulated by respiration because the filling and emptying of the lungs cause change in the electrical impedance of the chest. Monitoring of breath regulation has also been proposed as a way to minimize the oscillations in HRV due to respiration, respiratory rate can be also used as a feature, combine it with additional biosignals for calculating the stress level of the subjects.

1.3.5 Photoplethysmography (PPG)

Photoplethysmography (PPG) is an optical non-invasive method measuring variations of skin hue associated with concurrent changes in blood volume in subcutaneous blood vessels during the cardiac cycle. This measurement provides valuable information about the cardiovascular system [2].

A typical PPG device contains a light source that generates pulses and receive the reflected light a photodetector. The light source emits light to a tissue and the photodetector measures the reflected light from the tissue. The reflected light is proportional to blood volume variations [15]. From the PPG signals the pulse rate (PR), pulse rate variability (PRV) and blood pressure (BP) can be extracted; it is also used for HRV parameters estimation as it present high temporal peak agreement in relation to ECG.

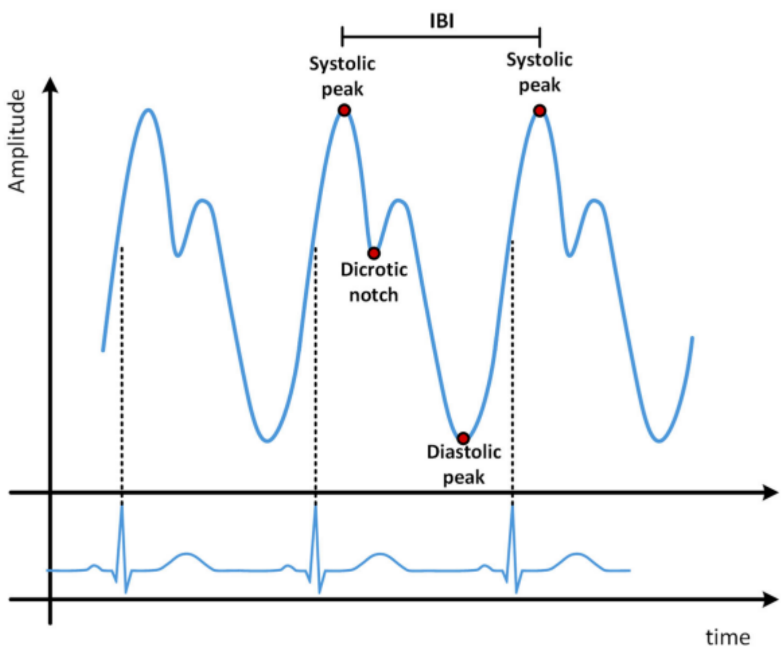


Figure 1.10: Raw PPG signal and its corresponding ECG.

PPG signal basically consists of four points which are diastolic points, systolic points, dicrotic notch and dicrotic wave as shown in Figure 1.10. The diastolic and systolic points are the important mechanism in PPG signal as it can provide useful information regarding the cardiovascular system. The systolic phase starts with a valley and ends with the pulse wave systolic peaks, the pulse wave end is marked by another valley at the end of the diastolic phase. While the time duration between two consecutive systolic points in the signal determine the instantaneous heart rate of an individual [16]. The stress-induced vascular response index (sVRI), which is a PPG-based measure is proposed to assess stress levels. Stress can also be reflected in peripheral vasoconstriction, being related to decreased pulse wave amplitude (PWA) in PPG signals [16].

1.4 Stress Inducing Tasks

Stressors may be either physical (environmental and physiological) or psychological/mental (cognitive and emotional) or mixed. There may be several types of stressors used in psychophysiology research:

- *Physical*, as strenuous physical activity, sleep deprivation, tiredness, painful stimuli, acute injury or medical emergency;
- *Environmental*, such as extreme temperature conditions, high levels of humidity, high levels of noise;
- *Mental task*, for example task demands and conditions taxing the person's cognitive capacities, inconsistent reward/reinforcement schedule;
- *Social*, as disturbances in social interactions, undesirable social roles, criticism, self-criticism;
- *Psychological/emotional*, as disturbances in personal life, intense emotional states, mental disorder affecting daily function;
- *Chronic*, as severe financial difficulties, poor living conditions, job insecurity;
- *Traumatic*, as memory of past traumatic experience that intrudes into consciousness and still affects the psycho-emotional state of a person [2].

The physical stressor has a direct effect on the body and induces direct metabolic or physiological changes. Physical stress may be an external environmental condition or the internal physical/physiologic demands of the human body. Psychological or mental stressors primarily activate the brain centers to disturb homeostasis without having any direct effect on the body. Numerous laboratory methods have been established to induce stress in humans, including the cold pressor test (CPT), Trier Social Stress Test (TSST), Maastricht Acute Stress Test (MAST), Montreal Imaging Stress Task (MIST), Stroop Color-Word Test (SCWT) and Mental Arithmetic (MA) [17].

1.4.1 Cold Pressor Test (CPT)

The cold pressor test was developed as a tool to study blood pressure variability by Edgar A. Hines. The test consisted of a 30-min baseline period during which blood pressure measurements were obtained, followed by a 60-s immersion of either a hand or a foot in ice water at 4-5 °C during which blood pressure was taken after 30 and 60 *s* with the cuff placed on the opposite arm. This test has a major limitation: it involves only physical pain sensation for stress induction without involving psychosocial evaluative threat. Being a physical stress, it only elicits the activation of the brain stem region, which in turn activates the sympathetic adreno-medullary system. It also lacks the uncontrollability and unpredictability of the procedure, which is an essential feature for robust activation of stress response [17].

1.4.2 Trier Social Stress Test (TSST)

The Trier Social Stress Test (TSST) emerged as a psychosocial stress protocol that reliably and consistently produced hypothalamic-pituitary-adrenal (HPA) axis stimulation. The TSST imposes a 15-minute period of psychosocial stress that includes 5 minutes of anticipatory stress, 5 minutes of public speaking and 5 minutes of mental arithmetic performed before a panel of evaluators. The consistent and large endocrine response to the TSST is believed to result from participants' concern about poor performance, which engenders a combination of ego involvement and awareness of social judgment with the potential for negative consequences. Physiologic markers of the stress response are obtained from salivary samples gathered before, during, and 1 hour after the induced stress period [18].

1.4.3 Stroop Color-Word Test (SCWT)

The Stroop Color-Word Test (SCWT) is a neuropsychological test extensively used for both experimental and clinical purposes. It assesses the ability to inhibit cognitive interference, which occurs when the processing of a stimulus feature affects the simultaneous processing of another attribute of the same stimulus. In the most common version of the SCWT, subjects are required to read three different tables as fast as possible. Two of them represent the "congruous condition" in which participants are required to read names of colors (henceforth referred to as color-words) printed in black ink (W) and name different color patches (C). Conversely, in the third table, named color-word (CW) condition, color-words are printed in an inconsistent color ink (for instance the word "red" is printed in green ink). Thus, in this incongruent condition, participants are required to name the color of the ink instead of reading the word [19].

1.4.4 Montreal Imaging Stress Task (MIST)

The Montreal Imaging Stress Task (MIST) consists of a series of computerized mental arithmetic challenges along with social evaluative threat components. To allow the effects of stress and mental arithmetic to be investigated separately, the MIST has 3 test conditions (rest, control and experimental). In the rest condition, subjects look at a static computer screen on which no tasks are shown. In the control condition, a series of mental arithmetic tasks are displayed on the computer screen, and subjects submit their answers by means of a response interface. In the experimental condition, the difficulty and time limit of the tasks are manipulated to be just beyond the individual's mental capacity. Upon completion of each task, the program presents a performance evaluation to further increase the social evaluative threat of the situation [20].

1.4.5 Social Evaluative Tasks

Psycho-social stress is a type of human stress which occurs when an individual has to face people or a group of people as in a public speaking task. When a socially threatening situation occurs, two mechanisms of the human body are affected, which include the autonomic nervous system and the neuroendocrine system. Instead of real-life events exposure, virtual reality has also been used as a stressor. Virtual reality exposure therapy (VRET) is an intermediate phase between thoughts and real-life events. Virtual reality is useful for a person who has difficulty imagining fearful tasks, it has also the advantage that if the stimuli become too threatening for the patient, the therapist has the control to stop the stimuli. The public speaking task as a social stressor has been a focus on very few studies. Existing literature either focuses on the real audience or a virtual audience. It has been shown that when men and women are both subjected to real-life stressors, no significant difference based on gender was found [13].

1.4.6 Maastricht Acute Stress Test (MAST)

The Maastricht Acute Stress Test (MAST) is designed to be a simple, quick, and non-invasive procedure aimed at activating the human stress system. The MAST procedure combines the most stressful features from two of the most common experimental paradigms, the TSST (involving novelty, unpredictability, ego involvement) and the CPT (involving physical pain). In direct comparison to a range of other validated stress protocols, the MAST induces similar, if not greater, changes in Blood Pressure (BP) immediately and 5 min following the conclusion of the stress test, and significant increases following the procedure in subjective experiences of stress, pain, and unpleasantness, as measured on Visual Analog Scales (VASs). In addition, the procedure has incorporated lack of control by not allowing participants to know how long their hand will be submerged in water in each trial [21].

1.4.7 Mental Arithmetic Task (MAT)

MAT is one of the most commonly used stimuli for inducing stress. Mental arithmetic task is a mechanism to increase the mental workload by performing a series of arithmetic operations with a varying range of difficulty. This stimulus is easy to implement and does not requires any special instrument [13].

Chapter 2

Noise Analysis

2.1 Noise Analysis in Biosignals

Biosignals have quite low signal-to-noise ratio and are often corrupted by different types of artifacts and noises originated from both external and internal sources. The presence of such artifacts and noise poses a great challenge for biosignals analysis. Internal sources of artifacts are due to different body activities.

External artifacts arise from coupling due to unwanted external interferences. For these reasons, biosignals often require to be pre-processed properly by removing such artifacts and interferences before any further analysis and feature extraction. There are several ways to pre-process a signal, the first way is to use digital filters. However, digital filtering is not suitable enough to effectively remove such artifacts, so advanced signal processing techniques have been proposed in the literature for this purpose.

2.1.1 Noise in ECG

A typical ECG signal of a normal subject is shown in Figure 1.3. Artifacts (noise) are the unwanted signals that are merged with ECG signal and sometimes create obstacles for the physicians from making a true diagnosis. There are mainly four types of artifacts encountered in ECG signals: baseline wander, powerline interference, EMG (Electromyography) noise and electrode motion artifacts. Electromyography is a technique for evaluating and recording the electrical activity produced by skeletal muscles. EMG is performed using an instrument called an electromyograph to produce a record called an electromyogram. An electromyograph detects the electric potential generated by muscle cells when these cells are electrically or neurologically activated [22].

Baseline wander, is the effect where the base axis (x-axis) of a signal appears to move up and down rather that be straight. This causes the entire signal to shift from its normal base as shown in Figure 2.1 and distorts ST-segment and other low-frequency components of the ECG signal. In the ECG, this effect is caused due to improper electrodes, patient's movement and breathing. The amplitude and duration of the wander depend on electrode properties,

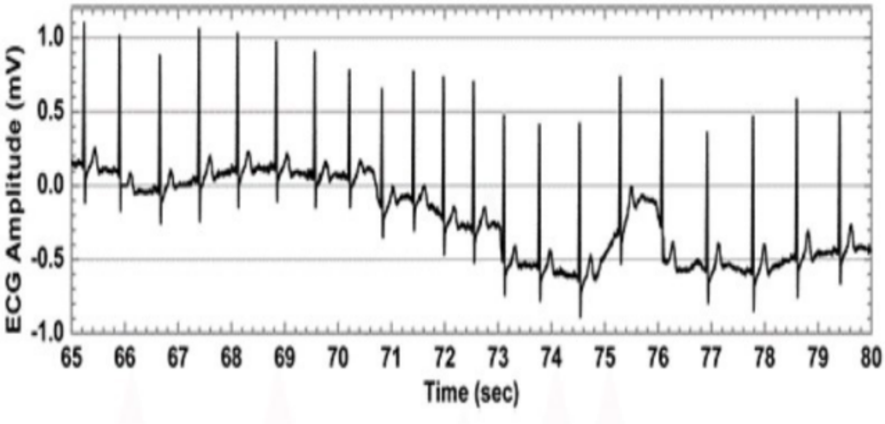


Figure 2.1: Baseline wander in ECG [23]

electrolyte properties, skin impedance, and body movements. The frequency content of the baseline wander is in the range between 0.05 and 1 Hz, it means that is a low-frequency artefact in ECG [23].

The powerline interference is introduced because of the electromagnetic interference of the alternating supply. This represents a common noise source in the ECG. The interference may be due to stray effect of the alternating current fields due to loops in the patient's cables. Other causes are loose contacts on the patient's cable, as well as dirty electrodes. When the machine or the patient are not properly grounded, power line interference may even completely obscure the ECG waveform. It is necessary to remove powerline interference from ECG signals as it completely superimposes the low frequency ECG waves like the P wave and T wave [23]. Such noise is characterized by 50 or 60 Hz sinusoidal interference, possibly accompanied by a number of harmonics. Such narrowband noise renders the analysis and interpretation of the ECG more difficult, since the delineation of low-amplitude waveforms becomes unreliable and spurious waveforms may be introduced.

EMG noise is caused by electrical activities in muscles, which arise from eye and muscle movements and heartbeat. Typical sources of MA are muscle movements near the head region, like neck movements, swallowing, and so on. The presence of muscle noise represents a major problem in many ECG applications, especially in recordings acquired during exercise, since low amplitude waveforms may become completely obscured, and in general can alter the shapes of local waves of the ECG signal. Muscle noise presents a much more difficult filtering problem since the spectral content of muscle activity considerably overlaps that of the PQRST complex [23]. EMG leads to distortion of local waves of the ECG signals due to a frequency match in the range of 0.01-100 Hz, this makes it challenging to denoise the signals for proper recognition of various ECG arrhythmias.

Electrode motion artifact is the noise introduced to the ECG that results from motion of the ECG electrode.

In order to record an ECG signal, electrodes (transducers) are placed at

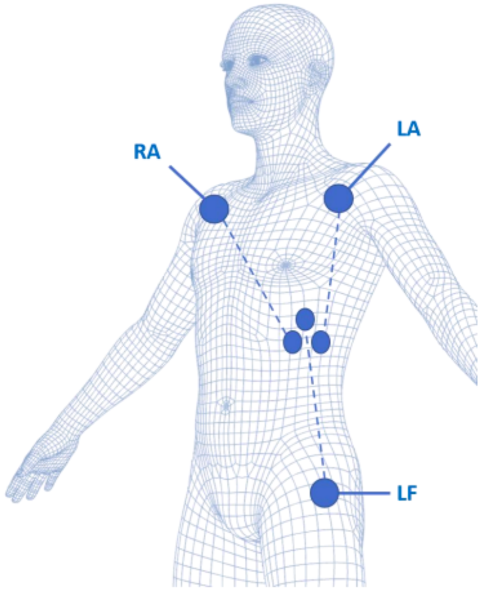


Figure 2.2: Typical Electrode placements

specific positions on the human body. In particular, the *biosignalplux* ECG is primarily designed for a single-lead ECG acquisitions in the Einthoven configuration. in Figure 2.2 the electrode placements is in the configuration by using the standard ECG sensor with electrode cable lenghts of 1.5 cm and 3 cm (reference).

Specifically, electrode movement causes deformations of the skin around the electrode site, which in turn cause changes in the electrical characteristics of the skin around the electrode and alters the impedance of the skin around the electrode. These electrical changes appear in the ECG as motion artifacts. Motion artifacts can produce large amplitude signals in the ECG and can resemble the P, QRS, and T waveforms of the ECG of baseline wander, but are more problematic to combat since their spectral content considerably overlaps that of the PQRST complex. These artifacts occur mainly in the range from 1 to 10 Hz, and in the ECG are manifested as large-amplitude waveforms which are sometimes mistaken for QRS complexes [24].

2.1.2 Noise in EEG

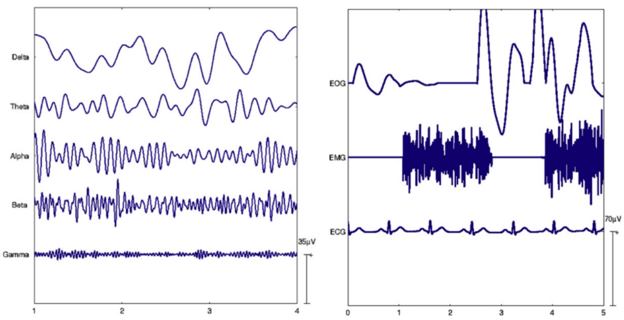
The various waveforms of the EEG convey clinically valuable information. The presence of artifacts in EEG signals can increase the difficulty in analyzing the EEG and to obtaining clinical information. Removing these artifacts is essential as they can affect the detection and extraction of features from the EEG signal. Among other possible categorizations, artifacts can be coarsely separated into those of physiological and non-physiological of technical origin. The most prevalent kind of artifacts are the following [25]:

- *Ocular artifacts.* The ocular artifact is a biological, non-neural disturbance generated by eye blinks and eye movements. The amplitude of these artifacts can be much larger than EEG signals, hence posing a serious problem for further analysis.

- *Muscle artifacts.* The electromyogram (EMG) measures the electrical activity on the body surface caused by contracting muscles. This artifact is typical of patients who are awake and occurs when the patient swallows, talks or walks. These signals have a wide frequency range and can be distributed across different sets of electrodes depending on the location of the source muscles. EMG presents a wide spectral distribution, thus perturbing all classic EEG bands: in particular, it considerably overlaps with beta activity in the 15-30 Hz range but may be as low as 2 Hz, making the widely used alpha band also vulnerable to muscle artifacts.

- *Cardiac artifact.* The amplitude of the cardiac activity on the scalp is usually of low amplitude; however this greatly depends on the electrode positions and differs for certain body types. When an electrode is placed on or near a blood vessel, it causes pulse, or heart beat artifacts and the expansion and contraction of the vessel introduce voltage changes into the recordings. The artifact signal has a frequency around 1.2 Hz, but can vary with the state of the object. This artifact can appear within EEG as a sharp spike or smooth wave. ECG artifact is recognized easily by its rhythmicity, if EEG and ECG are simultaneously taken. The identification of ECG artifacts is simple; they can be easily identified by coinciding with the ECG tracing.

- *Electrode artifact.* Morphologically, this appears as single or multiple sharp waveforms due to abrupt impedance change. It is identified easily by its characteristic appearance and its usual distribution, which is limited to a single electrode.



(a) Brain Rhythms (b) Artifacts

Figure 2.3: Five normal brain rhythms and three different kinds of artifacts [25]

2.1.3 Noise in EDA

Wearable devices are more prone to noise and artifacts and EDA is not immune from this issue. The main characteristic of EDA are peaks, known as skin conductance responses (SCRs), which occur as a reaction to a stimuli. Measurement of EDA in ambulatory settings creates uncertainty on what causes the peaks in the signal and hence, makes the signal vulnerable to the presence of artifacts. Among the factors that influence the presence of artifacts in ambulatory EDA signals are the recording procedure, such as the stability of electrodes, the influence of environment temperature and the user's physical activity. These factors cause artifacts that could resemble or not SCRs.

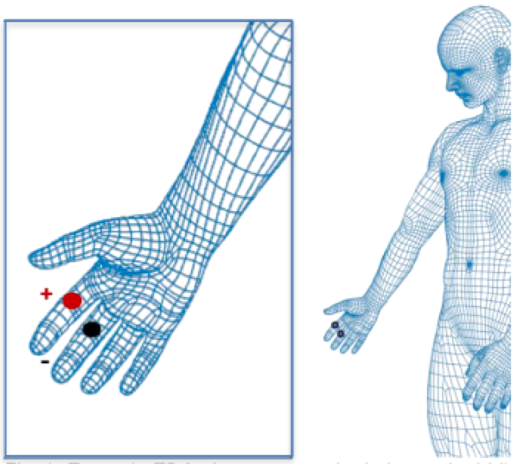


Figure 2.4: Typical Electrode placements for EDA

For this reason the artifacts can be divided into two sub-groups described as follows:

- *Shape artifacts* refer to artifacts that do not resemble physiological responses. Improper placement of electrodes or their movement for instance causes abrupt changes in the signal that cannot be generated by the electrodermal system itself and do not conform to the specifications of physiological responses. In Figure 2.4 is shown an example of the placement of the electrodes in the anterior side of the hand,

- *Thermoregulation responses* refer to physiological responses that are similar to EDA responses, but are not caused by electrodermal system. High physical activity or even increase in environmental temperature rises user's sweating, hence leading to physiological responses in EDA signal caused by thermoregulation rather than the electrodermal system. Such artifacts might be misinterpreted as physiological responses elicited by for instance an emotion, reducing the reliability of an emotion recognition system [26].

Noises in EDA can be also categorized into two another types: *extrinsic* noises and *intrinsic* noises. Extrinsic noises are defined as noises that are generated from environments outside of human body, for example electromagnetic field related noises or variability in temperature and humidity on skin. Besides the extrinsic noises, several human physiological activities other than ones related to stress, such as activation of muscles and thermoregulatory sweating, also cause undesired modulation in EDA, which is called intrinsic noises, like for example respiration noise.

Recently, a couple of studies have been conducted to better alleviate motion artifact. Since part of the motion artifact has similar frequency range with the desired EDA signals, the denoising methods have primarily depended on prior knowledge about the morphological characteristics of noise-free EDA signals. These methods have been validated to effectively suppress the motion artifact. Overall, the extrinsic noises are appropriately alleviated by previous denoising methods based on differences in signal characteristics. However, theses denoising methods based on differences in signal characteristics, are limited to attenuate intrinsic respiration noise because the respiration noise has similar signal characteristics with EDR, the EDA's reactivity to stress [27].

It should be noted that unlike other biosignals, such as ECG and PPG, EDA does not exhibit periodicity. Hence, manual adjudication of clean versus noisy EDA can be rather tricky.

2.1.4 Noise in Respiration Signal (RESP)

A good quality respiration signal as shown in Figure 1.9 should have little to no baseline wandering, which means there is no shifts parallel to the x-axis, it should also have clean and periodic peaks and valleys and approximate a sinusoidal form. The typical kind of artifact in the respiration signal is the motion artifact. In the case of jumping, the signal is corrupted by having in the peak that corresponds to the jump a vertical shift, so the valley is much slower and the peak is much lower as well compared to the expected signal. Another motion artifact can be caused by a torso rotation that leads to significant changes in the signal, where the peak and the valley are much earlier and much shorter than in the expected signal. Slow breathing dynamics and low respiration frequencies break out of the characteristic sinusoidal signal form of the respiration signal.

More than motion artifacts and baseline drift, other kinds of noise can be:

- Muscle noise is an high-frequency noise that can be caused by the contraction of the chest muscles during respiration,
- Thermal noise is random noise that arises due to the random motion of electrons within the recording circuity,
- Environmental noise includes noise from sources such as air conditioning, fans, and other electrical appliances that can be picked up by the recording device.

2.1.5 Noise in PPG

PPG measures the blood pulse wave from which the heart rate, its variations and even the respiratory rate can be extracted. The largest problem with the proper extraction of these health parameters is that the PPG signals are often measured during various kinds of movement and therefore are corrupted with motion noise. This noise can appear in the form of unruly signals of large amplitudes in the PPG signals. It is also reflected in the frequency domain and overlaps with the frequency range of breath or heart rate. The quality of the PPG signal depends on the location and the properties of the subject's skin at measurement, including the individual skin structure, the blood oxygen saturation, blood flow rate, skin temperatures and the measuring environment. These factors generate several types of additive artifact which may be contained within the PPG signals. The main artifacts in PPG are described as follows:

- *Powerline Interference* is a type of noise that could be due to the instrumentation amplifiers as for the ECG. Moreover, high frequency artifacts caused by mains power sources interference is induced onto the PPG recording probe or cable. This artifact introduces a sinusoidal component into the recording, clearly displayed as a spike.

- *Motion artifact* may be caused by poor contact to the fingertip photo sensor. Variations in temperature and bias in the instrumentation amplifiers can sometimes cause baseline drift as well. Usually the cause of motion artifacts is assumed to be due to vibrations or movement of the subject. The shape of the baseline disturbance caused by motion artifacts can be assumed to be a biphasic signal resembling one cycle of a sine wave.

- *Low amplitude PPG signal* takes into account the PPG waveform that is subject to sudden amplitude changes due to the automatic gain controller, which adjusts the gain of the amplifier automatically based on the amplitude of the input signal. This may cause amplitude saturation in the amplitude of the PPG waveform at a maximum or minimum value, or to rest at some random fixed value. However, the reduction of PPG amplitude can be directly attributable either to a loss of central blood pressure or to constriction of the arterioles perfusing the skin.

- *Premature ventricular contraction* is a kind of artifact in which these premature ventricular beats interrupt the normal heart rhythm and cause an irregular beat. This is often felt as a "missed beat". This type of arrhythmia will affect the main events detection accuracy in PPG signals [28].

2.2 Digital Filtering

There can be several ways to prevent the presence of artifacts, but it is impossible to completely eliminate the noise by just prevention. For this reason digital filter is one way to clean the signal from noise. Filters are used for two purposes: *signal separation*, used when a signal as been contamined with interference or noise, and *signal restoration*, used when a signal has been distorted in some way (for example an image acquired with an improperly focused lens). Digital filtering, which is nothing but a discrete-time LTI system, is a common part of preprocessing the recorded biosignals to attenuate out-of-band noises and artifacts. The out-of-band noise refers to unwanted signals or noise that exist outside the desired frequency band of a system.

There are two ways of obtaining a digital filter:

- By convolution of the impulse response of the filter with the input signal: $y[i] = \sum_{j=0}^{M-1} h[j]x[i-j]$, that normally means $y[i] = h_0x[i] + h_1x[i-1] + h_2x[i-2] + ...$

  Where: *y[i]* is the output signal, *h[j]* is the impulse response (coefficients of the filter), *x[i-j]* is the input signal and M is the quality of the filter, so the higher is M the better is the filtered signal.

  This kind of filters are called **Finite Impulse Response (FIR)**, they are the slowest but have the best performance and are defined by their filter kernel (impulse response),

- By recursion, considering in the summation also the previous values of the output signal:

  $y[i] = a_0x[i] + a_1x[i-1] + a_2x[i-2] + ... + b_1y[i-1] + b_2y[i-2] + ...$

  Where a and b are the recursive coefficients of the filter.

  These kind of filters are called **Infinite Impulse Response (IIR)**, since their impulse response is composed by sinusoids exponentially decaying. Theoretically these are infinite but can be truncated after it falls beyond noise level. They have worse performance but higher speed and are not defined by a filter kernel but by a set of recursive coefficients.

Both FIR and IIR filters have been found to be used in the pre-processing stage depending on the application and given specification. Depending on the frequency range of the artifact affecting the biosignal, there are several types of FIR or IIR filters that can be used:

- *Low-Pass Filter* if the desired bandwidth of the recorded signals is known and various biosignals are being considered, a standard FIR low-pass filter can be applied to eliminate high-frequency out-of-band disturbances and artifacts present in the raw recording. This type of filter is versatile and can be employed, for instance, to eliminate EMG noise from ECG signals or to avoid aliasing during EEG acquisition.

- *High-Pass Filter* can be used when is needed to attenuate the signal taken into account below a certain frequency range. For example, a standard techniques for removing the baseline wander in the ECG acquisition is to use a high-pass filter. Another case in which a high-pass filter may be used is when is needed to remove the dc offset (constant voltage that is added to an AC signal) which in turn is largely caused by the electrode/gel/body interface, if a patient stays still, the cut-off frequency of the filter can be chosen equal to 0.05 Hz.

- *Notch Filter* or also known as band-rejection filter, passes most frequencies unaltered, but attenuates those in a specific range. It is used in many applications where a specific frequency component needs to be eliminated. For example, it can be used to remove powerline interference (50 Hz). However, the problem with notch filtering is that it not only removes the removes the powerline interference at the fundamental frequency but also removes signal components at that notch frequency, so the cut-off frequency has to be determined in advance to design the filter.

- *Pass-Band Filter* is a filter that allows signals between two specific frequencies to pass, but that discriminates against signals at other frequencies. Instead of having a cut-off frequency there is a pass-band region in which the signal pass without any corruption, and outside the range the signal is attenuated. For example, a correct selection of the frequency filter pass-band ensures adequate isolation of the high frequency QRS complex of the ECG signal against the background of the low frequency P and T waves of the signal, low frequency noise, and 50Hz powerline artifacts. An optimal filter pass-band for ECG processing can be 8-20 Hz, in order to be effective in reducing noise and interference while preserving the QRS complex and other important low-frequency features of the signal.

Digital and analog filters both take out unwanted noise or signal components, but filters work differently in the analog and digital domains. Analog filters are circuits made of analog components such as resistors, capacitors, inductors and op amps. Digital filters are often embedded in a chip that operates on digital signals. According to the frequency response function achieved by the filter, the common filters in wearable sensors can be divided into three types: Butterworth filter, Chebyshev filter, and Elliptic filter, which can segment the target signal with different characteristics. It is worth noting that the three types are not fixed analog or digital filters, and the conversion between analog filters and the corresponding digital filters can be achieved through algorithms.

- *Butterworth Filter* has the characteristic that the frequency has good stability both inside and outside the pass frequency range, and the frequency band is maximally flat in the pass band. However, Butterworth filters have the disadvantage of a slow descent in the stop band, resulting in a long equivalent transition band. If the signal of interest happens to be within the transition band, it is prone to distortion. This disadvantage can be overcome as the filter order increases. The decay of the resistance band accelerates with increasing order, resulting in more accurate processing results. Based on the above characteristics, the Butterworth filter is suitable for cases where the passband and stopband ripples are small, and the requirements for the transition band signal are low. When choosing the Butterworth filter order, accuracy and complexity should be balanced. For example, in a wearable device for real-time detection of eye vergence in a virtual reality, a third-order Butterworth filter is used for band-pass filtering, or when performing in-ear continuous PPG monitoring, a fourth-order Butterworth low-pass filter with a cutoff frequenct of 10 Hz was selected as the low-pass filter [29].

- *Chebyshev Filter* is another type of filter that compared to the Butterworth filter, with the same order the Chebyshev filter drops faster in the stop band, but the response in the pass band fluctuates with respect to the response in the pass band region for a Butterworth filter. Chebyshev filters are further divided into two types: Chebyshev I filters which have equal ripple in the pass band and flat in the stop band; Chebyshev II filters which have flat in the pass band and equal ripple in the stop band. However, sometimes too rapid a drop can also have an adverse effect on the results. For example, the rapid drop in the stopband of a Chevyshev I filter gives it a narrower transition band than a Chevyshev II filter, which can adversely affect its ability to filter out anomalously correlated signals [29].

- *Elliptic filters*, are equiripple in both passband and stopband; it is different from the Butterworth filter with a flat passband and an equal ripple in the stopband, and the Chebyshev I filter with an equal ripple in the passband and a flat stopband. Comparing with the filters mentioned before, at the same order, Elliptic filters have the smallest passband and stopband fluctuations, as well as the narrowest transition band. Elliptic filters require only a lower order to achieve the same accuracy as Butterworth and Chebyshev filters. Elliptic filters have the significant features of a narrow transition band and fast attenuation [29]. In the field of wearable sensors, Elliptic filters are not as popular as the other two types of filters. Most studies on Elliptic filters focus on its ability to handle ECG signals, since the Elliptical filter has the best effect of low-pass filtering among all the filters in the case of high-frequency noise generated. Moreover, because the Elliptical filter transition band is so narrow, ECG sensors often use it precisely to eliminate 50 Hz power line interference [29].

2.3 Single Artifacts Removal Techniques

The variety of artifacts and their overlapping with signal of interest in both spectral and temporal domain, even sometimes in spatial domain, makes it difficult for simple signal preprocessing techniques such as typical digital filtering or amplitude thresholding to identify them from desired biosignals. Usually, the extra-physiological artifacts such as line interference or electrode noise can be removed by digital filtering techniques as there is a spectral separation. But careful attention is required to remove physiological artifacts as they co-exist within the same frequency range of the signals.

2.3.1 Empirical Mode Decomposition

Several techniques have been proposed in the literature for denoising biosignals based on *empirical mode decomposition*(EMD). This technique has been used for analyzing non stationary signals. The aim of EMD methods is to adaptively represent the signal as sums of zero-mean oscillating components, called the intrinsic mode function (IMFs) using a sifting process. In other words, the EMD algorithm decomposes the signal, x[n], into a set of components with amplitude-frequency modulated, b[n], called intrinsic mode functions (IMFs). In the whole data set and at every point, every IMF must satisfy that the number of extrema are the same with the number of zero crossings or differ at most by one, and the mean value of the evelope defined by the maxima and minima must be zero. The signal reconstruction process is achieved by total sum of IMFs and the residual. The major advantage of EMD is that the basic functions are derived from the signal which is different from the wavelet approach whose basis functions are fixed. The noise components of a noisy signal are centered on the first IMFs (high-frequency IMFs) and the useful information of the signal is often concentrated on the last IMDs (low-frequency IMFs). Thereby, the denoising method can be based on the partial construction of the signal using only the last relevant IMFs. This denoising method can be used for several artifacts such as: ocular artifacts in EEG, motion artifact in PPG. EMD has been adopted to analyze non-linear and non-stationary signals like EEG as explained in [30]. It has been demonstrated that EMD filtering technique can be very helpful to remove the eye blink and eye movement artifacts in a single channel EEG by partial reconstruction from the components of decomposition. In this method, the noisy IMFs are identified relying on entropy and reconstruct the signal using IMFs that comprises lower entropy [30]. EMD decomposes the signal into IMFs, which are extracted through an iterative sifting process. The algorithm for a general raw EEG signal x(t) is described as follows:

(a) Determine the peaks of raw EEG signal,

(b) Construct the lower envelope and upper envelope of the signal separately by using a cubic spline interpolation,

(c) Generate the first Intrinsic Mode Function (IMF1) by subtracting the mean evelope from x(t).

(d) The first residual component is obtained by subtracting IMF1 from x(t),

(e) Repeat the process described from step (a) to calculate the next IMF by considering the residual component as a new signal,

(f) Repeat the process until no more IMFs can be extracted. The original signal x(t) can be reconstructed from IMF

In the case of motion artifacts in PPG, with a measure from an accelerometer is possible to measure movement and link it to the part of the respiratory signal which is corrupted with motion artifacts, but traditional filtering methods may not work well since they cannot distinguish between the sought after signal and the movement noise. Nevertheless, there are some possibilities to alleviate this issue even if no acceleration signal is available. The disadvantage of these methods is that they only marks the faulty parts and cut them out entirely. To overcome this problem it is possible to generate a syntethic reference signal out of the corrupted PPG signal with the use of EMD. The idea is to generate a reference signal from the corrupted PPG signal using Complex Empirical Mode Decomposition (CEMD). To generate the reference noise signal the following steps have to be done:

- First, all the local minima and maxima of the originals signal $x(t) = d(t) = S(n) + N(n)$ need to be found. Where *x(t)* is the raw accelerometer signal, *S(n)* is the raw PPG signal and *N(n)* is the motion noise.

- The next step is to envelope all the maxima and all the minima, after all the envelope signals are generated, the mean value, m(t) is calculated. The value of the mean is then subtracted from the original signal: h(t)=x(t)-m(t) and the new signal h(t) is decomposed into IMF by sifting process until h(t) meets the IMF conditions.
- Therefore, the next step is to identify the quasi-residue function r(t)=x(t)-c. This loop has to be repeated until r(t) has only one extrema.
- Afterwards, the spectrum of each IMF based on the predefined frequency range has to be computed.
- The last step to generate the reference noise signal is identifying the desired signal portion range and eliminating IMF corresponding to the desired frequency components of the PPG [31].

It is possible to generate a reference signal out of a corrupted PPG signal with EMD, after the reference signal is generated, an adaptive filter can be used to remove motion artifacts [31].

In [32], an integrated EMD adaptive threshold denoising method (IEMD-ATD) is proposed, which is suitable for the reduction of noise in ECGs. This method contains four steps:

1. First, the ECG is decomposed through integrated EMD (IEMD) into a set of IMFs and one residual term,
2. Second, all of the IMFs are divided into three groups: high-frequency noise predominant IMFs, noise-free IMFs and IMFs with low frequency artefacts,
3. Third, high-frequency noise predominant IMF are denoised by the proposed peak filtering denoising method after the adaptive threshold is calculated.
4. Finally, the denoised ECG signal is reconstructed by summing the denoised IMFs and the noise-free IMFs and directly discarding the IMFs with low-frequency artefacts.

2.3.2 Wavelet Transform

A wide range of approaches have been developed to try to extract both time and frequency information from a waveform. Basically they can be divided into two groups: time-frequency methods and time-scale methods.

Wavelet transform is a time-frequency technique that was introduced to overcome the limitations in time and frequency resolution. Wavelet decomposes a signal into different frequency components and studies each component with a resolution matched to its scale. This property can be used for denoising purposes. The wavelet transform can be used as yet another way to describe the properties of a waveform that changes over time, but in this case the waveform is divided not into sections of time, but segments of scale. In numerical analysis and functional analysis, a discrete wavelet transform (DWT) is any

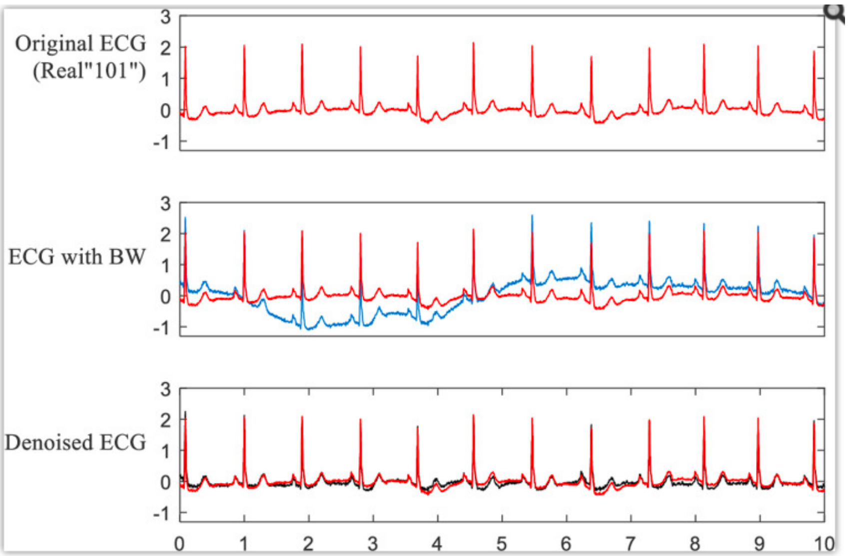


Figure 2.5: Denoised ECG by using EMD [32].

wavelet transform for which the wavelets are discretely sampled. As with other wavelet transforms, a key advantage it has over Fourier transforms is temporal resolution: it captures both frequency and location information. For example, wavelet denoising is applied on the raw PPG signal to eliminate high frequency noise, and then a methods based on wavelet transform combined with adaptive filter is applied to eliminate motion artifacts. The discrete wavelet transform (DWT) is not time-invariant. Due to this drawback, denoising via the DWT often suffers from additional artifacts like ringing effects in the vicinity of a discontinuity. To address the problem, the stationary wavelet transform (SWT) can be used , which is time-invariant and performs no downsampling. Consequently, the lenght of the sequences at each level is the same as that of the original sequence, which provides better sampling rates in the low frequency bands compared with standard DWT. For example, in [33], EDA data were decomposed into 8 levels. After high-pass filtering inside the SWT, the wavelet coefficients of SCL and SCRs will both have mean values around zero.

A typical histogram of the wavelet coefficients of an Skin Conductance (SC) signal is shown in Figure 2.6, with a fitted model of two mixed Gaussians superimposed. The Gaussian with smaller variance corresponds to the wavelet coefficients of SCL, while the Gaussian with larger variance corresponds to the wavelet coefficients of SCRs. Apart from the two distributions, there are a few very large coefficients in the histogram, which are motion artifacts that need to be removed [33].

In [34] study for ECG, a discrete wavelet transform using the eight-order symlet wavelet was applied to a single channel recording to create an eight-level wavelet decomposition of the raw signal. After calculating the wavelet coefficients, nonlinear thresholding in the wavelet domain was used for artifact removal, where the absolute value of the coefficients greater than the threshold was set to zero. The inverse wavelet transform was then implemented using the new coefficients to obtain the clean ECG signal as shown in 2.7.

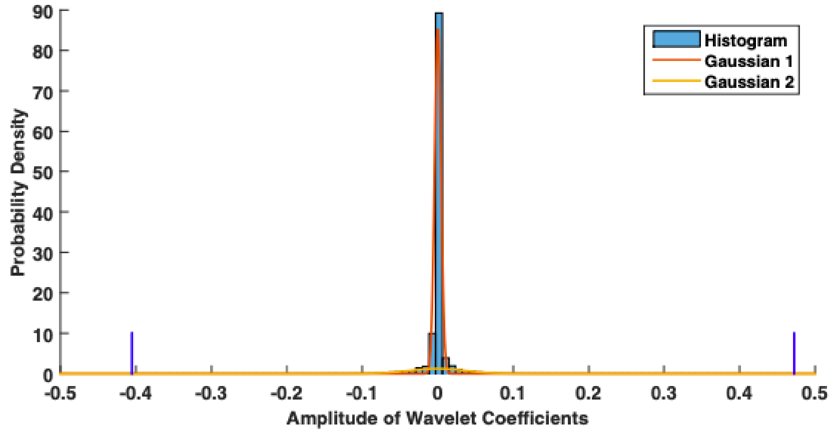


Figure 2.6: Histogram of the wavelet coefficients of an SC signal with a fitted model of two mixed Gaussians superimposed

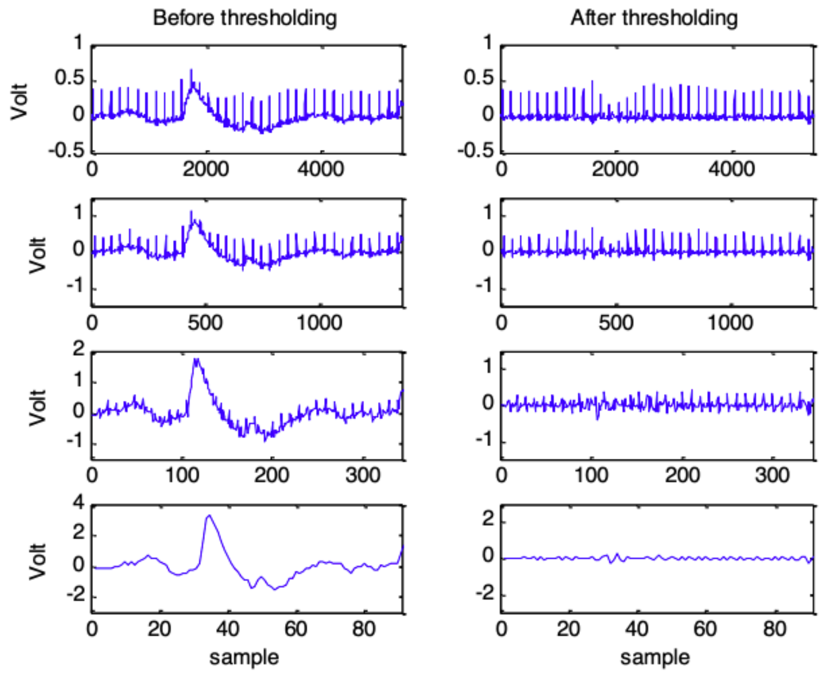


Figure 2.7: Four levels of wavelet decomposition of the ECG signal before and after thresholding [34].

2.3.3 Independent Component Analysis (ICA)

Independent component analysis (ICA) is a statistical and computational technique for revealing hidden factors that underlie sets of random variables, measurements, or signals. ICA defines a generative model for the observed multivariate data, which is typically given as a large database of samples. In the model, the data variables are assumed to be linear mixtures of some unknown latent variables, and the mixing system is also unknown. The latent variables are assumed non-Gaussian and mutually independent, and they are called the independent components of the observed data. These independent components, also called sources or factors, can be found by ICA. For the ICA model, n linear mixtures

$$x_1, x_2, .., x_n$$

of n independent components combined together in a mixture are denoted as:

$$x_j = a_{j1}S_1 + a_{j2} + \ldots + a_{jn}S_n \qquad \forall j \tag{2.1}$$

Whereas, X is a random vector whose elements are a mixture of

$$x_1, x_2, .., x_n$$

, S is the random vector with components

$$s_1, s_2, .., s_n$$

and

$$a_{j1}, a_{j2}, .., a_{jn}$$

are the mixing coefficients.

This equation model can be rewritten as the generalized form:

$$X = AS \tag{2.2}$$

The above model is called Independent Component Analysis or the ICA model. This model is solved with the assumption that components of S are statistically independent, independent components follow non-Gaussian distribution, and the mixing matrix A is square. After estimating the matrix A, is easy to find its inverse transformation (I):

$$S = IX$$

This equation calculates each independent component of S from the mixture of signals [35]. ICA has been used to extract independent components from ECG signal which is synthetically mixed with the random noise in [35]. The concept of two independent components, i.e. one is ECG pure signal and the other as noise, has been applied.

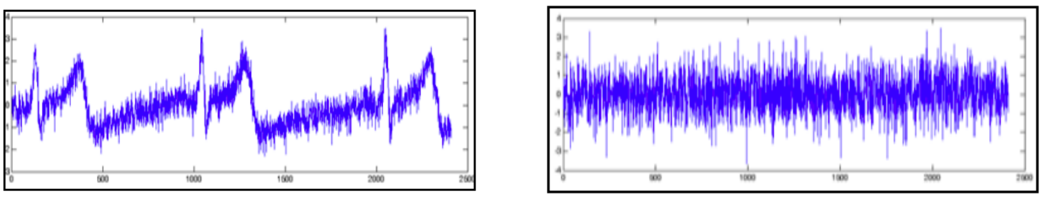


Figure 2.8: Left: Contaminated ECG component. Right: Extracted noise from ECG signal using ICA

In [3] an ICA has been used to reduce intrinsic signal artifacts in EEG, the raw EEG was decomposed into different components and therefore the artifactual components were identified. To calculate different components in

EEG signals the Extended Infomax method was applied to decompose the original EEG recording across 14 different electrodes into 14 components. The, three components that represented the most common intrinsic signal artifacts were removed.

2.3.4 Principal Component Analysis (PCA)

PCA is one of the Factor Analysis methods. It is aimed at the reduction of a wide range of random variables to a smaller set. PCA is a mathematical technique that can be used to identify and remove sources of noise or variation that are not of interest in a given dataset. In the context of biosignals, this can include various types of artifacts such as electrical noise, motion artifacts, or other sources of interference that can affect the quality of the signal. The goal of using PCA in this context is to extract the underlying signal of interest while minimizing the impact of the artifacts. This is typically done by identifying and removing the principal components of the signal that are most strongly correlated with the artifacts. PCA is a method based on matrix calculus, one of the first steps is to calculate and to select the eigenvalues of the co-variation matrix. The greatest eigenvalues should be selected from the calculated numbers. On the basis of the selected values the analysis is carried out in order to ascertain the most significant principal components. The greater an eigenvalue is, the less information is going to be lost. The goal is to minimise information loss. For the selected eigenvalues the eigenvectors are computed by means of a system of linear equations. The last step of the algorithm is the estimation of the point in the new space, the point corresponding to the given observation vector [23]. In general, the PCA method allows to separate some of the factors, components, which contain artefacts. It also allows to decrease, for example, the volume of an EEG signal and as a result, an EEG signal decreased by the artefacts is obtained. However, eliminating too much information from an EEG signal can render the signal useless. It is the fragments which are essential for the description of the EEG signal content that would be extracted in that case.

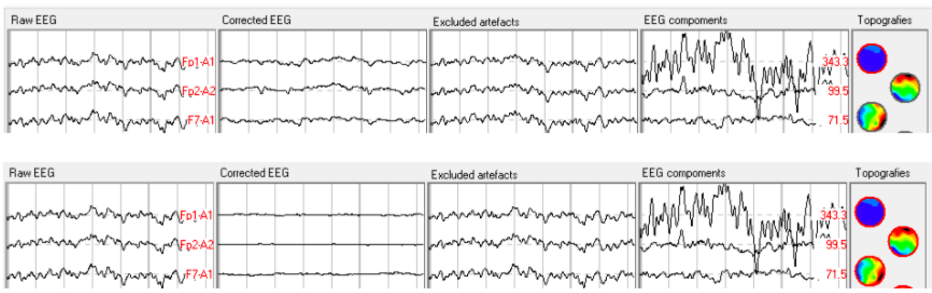


Figure 2.9: PCA methods applied for artefact correction

The first picture presents elimination of one EEG signal component. Along with that, the second one presents how a signal looks after elimination of 12 components. That resulted in disappearing of the EEG signal [23].

2.3.5 Adaptive Filter

In general, an adaptive filter is a type of digital filter that is capable of adjusting its parameters in response to changes in the input signal. It is used in various signal processing applications, such as noise cancellation, equalization, and system identification. The adaptive filter works by minimizing the difference between the desired output and the actual output of the filter. This is achieved by adjusting the filter coefficients in real-time based on the error signal, which is the difference between the desired output and the actual output of the filter.

Adaptive filters have been shown to be useful in motion artifact reduction in some studies, for example Tong et al. [36] used an anisotropic magnetoresistive (AMR) sensor and an accelerometer as the source of reference input, and their results indicate that adaptive filtering can reduce the amount of motion artifact present in the ECG, and the accelerometer-based sensor outperforms the AMR sensor.

In [36], the adaptive filtering method is used because it can adaptively track the signal under non-stationary conditions and can adjust its impulse response to filter out the noise in the input with little or no prior knowledge of the signal and noise characteristics. The adaptive filter works generally for the adaptation of signal-changing environments, the spectral overlap between noise and signal, and unknown or time-varying noise. It has the capability of adaptively tracking the signal under non-stationary conditions and can be used for different purposes, such as system identification, prediction, and noise cancellation. Figure 2.10 shows the working principle of the adaptive filter in noise cancellation. In adaptive noise cancellation systems, the objective is to produce a system output *X'(n)* that is the best fit in the least squares sense to the signal *X(n)* by feeding the system output back to the adaptive filter and adjusting the filter through an adaptive algorithm to minimize total system output power. The adaptive filter contains a digital filter with adjustable coefficients and the adaptive algorithm to modify the values of coefficients for filtering each sample. The coefficients of the digital filter are continuously changed according to the chosen adaptive algorithm so as to minimize the mean squared value of the error signal *e(n)*. The reference signal *S(n)* is fed into a digital filter to produce an output *N'(n)*, which is as close as possible to the replica of the noise *N(n)*. Subsequently, this filtered signal output *y(n)* is subtracted from the primary input *d(n)* to obtain the estimated desired signal *X'(n)*.

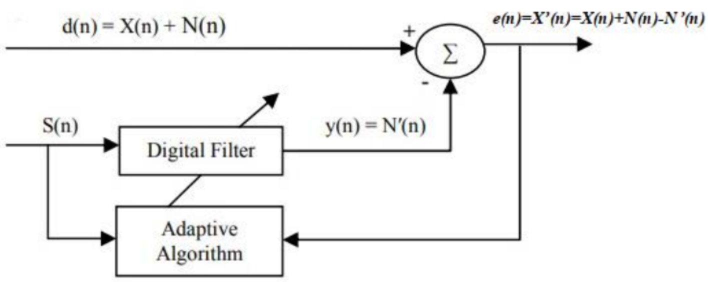


Figure 2.10: Principle of the adaptive filter in noise cancellation [36].

2.3.6 Sampling Nyquist Theorem

The Sampling Nyquist Theorem is defined as follows: in order for a band-limited (i.e., one with a zero power spectrum for frequencies $f > F$), baseband ($f > 0$) signal to be reconstructed fully, it must be sampled at a rate

$$f_s > 2 * F \tag{2.3}$$

.

A signal sampled at $f_s = 2F$ is said to be Nyquist-sampled, and

$$f_s$$

is called the Nyquist frequency. No information is lost if a signal is sampled at the Nyquist frequency, and no additional information is gained by sampling faster than this rate. The Theorem relates the frequency bandwidth F of the signal subjected to sampling to the minimal required sampling frequency *fs*. This relationship is described as follows:

$$f_s \geq 2fm \tag{2.4}$$

where *fm* is the frequency of the highest significant frequency component emerging above the noise floor of the entire system.

Many educational resources identify fm as the highest frequency component in the signal, and in a noise free environment this definition would be the simplest. However, biomedical processes and signals are routinely recorded in dynamic noisy environments, and the identification of this component is difficult, if not impossible. Therefore, for the purposes of biomedical signal processing, it would be prudent to modify the traditional definition of *fm* from the highest frequency component in the signal to the highest significant frequency component in the signal emerging above the noise floor. The determination of this component can be made objective if one considers the nonidealities and artifacts determining the noise floor of the analog instrument prior to the digitization, as well as the resolution of the analog-to-digital conversion process used by the data acquisition system .

Chapter 3

Machine Learning Algorithms for Stress Detection

3.1 Machine Learning Techniques

Machine Learning is a system of computer algorithms that can learn from examples on their own without being explicitly coded by anyone and automatically improve their performance through experience. In this way the system is capable of improving its actions and make better decisions, since it becomes more and more accurate in predicting the outcome.

A learning paradigm is a particular pattern on which a system learns. A machine, when given some data, also has a pattern approach that dictates the learning process. There are three main paradigms when it comes to Machine learning, and they can be divided into supervised, unsupervised, semi-supervised, and reinforcement learning (RL) [37].

- *Supervised learning* is an approach where a computer algorithm is trained on input data that has been labeled for a particular output. This kind of method assumes that labeled training data is available. The various algorithms generate a function for mapping inputs to desired outputs. By comparing its output with the intended one on the training data, it can find and correct mistakes, modifying the model and achieving a greater accuracy [37]. This type of paradigm is based on training and good at both classification and regression problems.

- *Unsupervised learning*, here models are not supervised using a training dataset, which means that the computer is provided with unclassified and unlabelled data. The models find the hidden patterns themselves and understand from given data. The main goal is to find the fundamental structure of the dataset and group that data according to similarities and finally signify that dataset in a compressed format.

- *Semi-Supervised learning* are problems where you have a large amount of input data (X) and only some of the data is labeled (Y). These kind of problems sit in between both supervised and unsupervised learning.

- *Reinforcement learning* involves training agents to take actions in an environment to maximize a reward signal. It has the main aim to, through interactions with the environment, to produce actions that maximize the system's performance. In this way, software agents can discover and model the ideal behaviour within a specific context.

  The method used is a trial and error search, where the system learns which actions are favorable, and therefore gets a reward, and which ones are disadvantageous, getting a penalization. This reinforcement signal is essential for the machine to learn which actions are best.

Supervised learning problems can be further grouped into regression and classification problems as shown in Figure 3.1:

- **Classification**: a classification problem is when the output variable is a category, such as "stress" or "no stress"

- **Regression**: a regression problem is when the output variable is a real value, such as "euros" or "height".

Unsupervised learning problems can be mostly grouped in one category [37]:

- **Clustering**: is a technique that groups data points. That is, given a set of data points, a clustering technique can be used in order to classify each data point into a particular group. A clustering problem is where you want to discover the inherent groupings in the data, such as grouping customers by purchasing behavior.

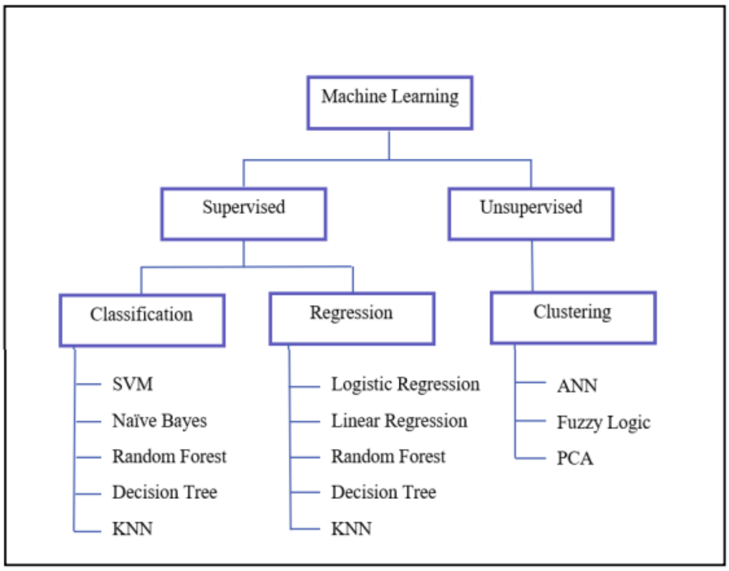


Figure 3.1: Machine Learning Techniques [38]

Some of the most popular algorithms in supervised learning are briefly presented below:

- *Support Vector Machine* (SVM), this algorithm includes two concepts: hyperplane and feature space. In a binary classification, the hyperplane is a decision boundary that separates the two classes. In a multi-class classification problem, the hyperplane is a higher-dimensional surface that separates the data points of different classes. The goal is to find the hyperplane that maximizes the margin between the classes, i.e., the distance between the hyperplane and the closest data points of each class. Depending on the number of features and their dimensions, sometimes a transformation in vector space is required to achieve the best possible separation hyperplane. The search for the separation hyperplane in the transformed spaces, usually of very high dimension, is based on the kernel function. This algorithm simultaneously minimize the classification error, while maximizing the geometric margin, which is the distance between the hyperplane and the nearest training points.

  SVM works well with unstructured (data that has no pre-defined structure or format) and semi-structured data (data that has some structure or format, but is not fully structured). SVM can solve any complex problem with an appropriate kernel function, can scale high dimensional data, risk of overfitting is low, comparatively memory efficient. The disadvantages is that it requires a long time for training large datasets, lack of transparency of results and doesn't perform very well in noise. The main applications of this algorithm can be for example handwriting and text recognition, facial expression classification, speech recognition, cancer diagnosis and prognosis [38].

- *Naïve Bayes* is defined as a type of probabilistic classifier that aims to process and categorize data. The operation of this classifier is simple, it is essentially a technique for assigning probability theory to classify data. Naive Bayes classification algorithms utilise the Bayes theorem. The central idea is that the probability of an event may be adjusted as new data are entered. This classifier is not a single algorithm, but a family of automatic learning algorithms that makes use of statistical independence. This method is faster, solves multi-class prediction problems, is more suitable for categorical input variables than numerical variables, requires much less training data when its assumptions of the independence of features hold true. The disadvantages are that when it faces the zero-frequency problem, which are situations where one or more categories or classes in the training dataset have no instances or examples, the estimation can be wrong in some cases which makes its probability outputs less reliable and it assumes that all predictors are independent which happens rarely in real life. The main applications of this algorithm can be text classification and sentiment analysis [38].

- *Decision Trees* are described as algorithms that perform repeated splits in the dataset to provide maximum data separation, the resulting structure is similar to a tree. Each node represents a feature and each branch is a decision that needs to be made according to the value of that feature. One of the most frequently used criterion to split into branch is using the information gain, this implies that the entropy reduction caused by dataset division is maximized in every split. It is a simple and easy method to understand, interpret and visualize, the output can be easily interpreted by humans. It is used for both classification and regression problems and can handle continuous and categorical variables. Another advantage is that it handles missing values and outliers automatically and needs less training period. The disadvantages are that overfitting is present, is not suitable for large datasets, small noise can make it unstable leading it to wrong predictions, chances of high variance in the outputs which leads to many errors in the final estimation. The main applications may be the use of demographic data to find prospective clients, energy consumption and healthcare management [38].

- *Random Forest* is the most popular Decision Tree algorithms use, it is defined as a classifier that uses ensembles of trees in order to achieve a better classification performance. It provides high accuracy through cross-validation, reduces overfitting in decision trees, works fine with categorical and continuous values, used for both classification and regression problems, automatically handles missing values present in the data, uses a rule-based approach that does not require normalizing data, and feature scaling. The disadvantages are that it takes more computational power and resources required to build numerous trees to combine their outputs, it also requires more time for training, suffers interpretability due to the ensemble of decision trees, fail to determine the consequence of each variable. The main applications may be on bank industry, healthcare sectors, stock market and E-commerce [38].

- *Linear Regression* is a statistical method used to model the relationship between a dependent variable and one or more independent variables. When there is only one independent variable and the relationship between the independent and dependent variable is linear, the regression is simple and linear. However, when there are multiple input variables, it is a multiple linear regression. The aim is to fit a line between the input and output variables, known as the regression line. The best fitting line or equation is the one that minimizes the sum of the squared errors between the predicted values and the actual values of the dependent variable.

- *Logistic Regression* is a classification algorithm used to predict the probabilities of a targeted variable. The objective or dependent variable is dichotomous, meaning that there may only be two possible classes but it may be expanded also to a multi-class setting. It is easier to implement and interpret, make no assumptions about distributions of classes in feature space, it can easily extend to multiple classes, classify unknown records fast, it performs well on linearly separable datasets, is less prone to overfitting and can consider regularization to avoid it. It provides also great training efficiency in some cases with low computation power, it may be updated easily to reflect new data. The disadvantages are that it constructs linear boundaries, it may lead to overfitting when the number of observations are less than the number of features. Is also important to ensure that there is little to no multicollinearity, when two or more predictor variables in a regression model are highly correlated with each other, between the independent variables. Logistic Regression is not suitable for solving non-linear problems due to its linear decision surface. Additionally, capturing complex relationships may be challenging, and the model can be sensitive to outliers. The main applications may be for example: online credit card transaction, text editing, gaming [38].

Some of the most popular algorithms in unsupervised learning are briefly presented below:

- *Artificial Neural Networks (ANNs)* are widely used computer models in prediction. ANNs consist of a set of units, called artificial neurons, connected together to transmit signals. The input information passes through the neural network, where it undergoes various operations, producing output values. Each neuron is connected to the others through links. As a general rule, they have one or more layers of interconnection, called hidden layers, which are responsible for connecting the neurons to each other. Between each of the neurons, a gate called the activation function which allows each neuron to be activated or deactivated according to a certain activation function is usually added. ANNs may be used for example in image and speech recognition, in robotics and automation environment and in natural language processing.

3.1.1 Performance Metrics for Classifiers

Performance metrics are useful in machine learning because they provide a way to quantify the effectiveness and accuracy of a model. They allow practitioners to compare different models, evaluate the performance of a model over time, and make informed decisions about which model to use in a given situation.

Performance metrics also provide a way to evaluate the performance of a model on unseen data. This is important because it helps practitioners to understand how well the model will perform on new data that it has not seen before.

Additionally, performance metrics can also be used to identify areas where the model is weak and needs improvement. For example, if a model has a low recall, it may not be identifying all positive cases, and this can indicate that the model needs to be adjusted or retrained.

Every machine learning task can be broken down to either *Regression* or *Classification*, just like the performance metrics. In regression, it can be used, for instance, the average prediction error (on the test set) to evaluate the performance of a particular prediction algorithm. In classification, the counterpart of the average prediction error would be the so called **accuracy**, which simply counts the relative number of correct predictions:

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions\ made}$$

The accuracy is the proportion of correct predictions made by the model out of all predictions made. It is commonly represented as a percentage, with a value between 0 and 100. A model with high accuracy is able to correctly identify the majority of the samples in the dataset. It should be noted that accuracy alone may not always be the best metric to evaluate a model's performance, as it does not take into account false positives or false negatives. However, accuracy hardly tells the full story about a classification algorithm's performance.

In classification, is also important to be aware about the type of error that may occur. In the context of machine learning, Type I error is related to false positive rate **(FPR)** and Type II error is related to false negative rate **(FNR)**. A type I error in this context would be when individuals are classified as positive, while they negative in reality. A type II error is when the individuals that are classified as negative, while they are positive in reality.

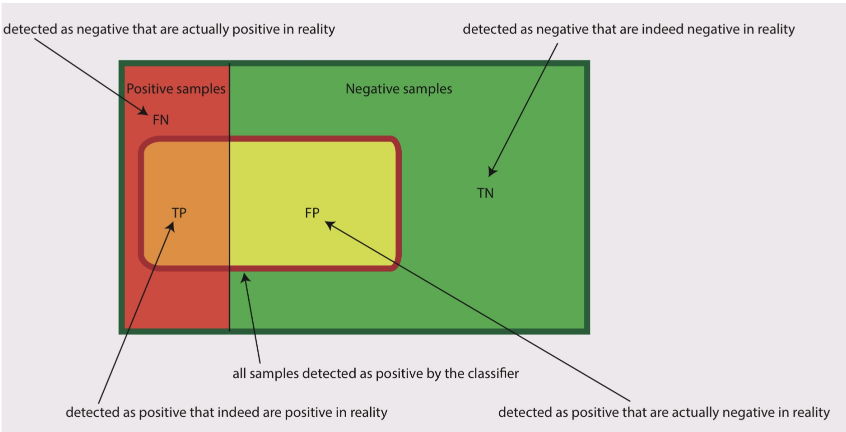


Figure 3.2: False Positives (FP) and False Negatives (FN). Source: *Optimization for Machine Learning Course*

A confusion matrix is a table that is used to define the performance of a classification model, typically a binary classifier, as shown in Figure 3.3. It is a table with two rows and two columns that reports the number of true positives, false positives, true negatives, and false negatives for a binary classifier. The columns of the matrix represent the predicted class, while the rows represent the actual class. The entries in the matrix are the number of observations that fall into each combination of predicted and actual class. The four important measures that can be derived from a confusion matrix are:

- True Positive (TP): the number of correct positive predictions,
- False Positive (FP): the number of positive predictions that are actually negative,
- True Negative (TN): the number of correct negative predictions,
- False Negative (FN): the number of negative predictions that are actually.

The confusion matrix can be used to calculate various metrics such as accuracy, precision, recall, F1 score, and AUC-ROC. Together these provide a

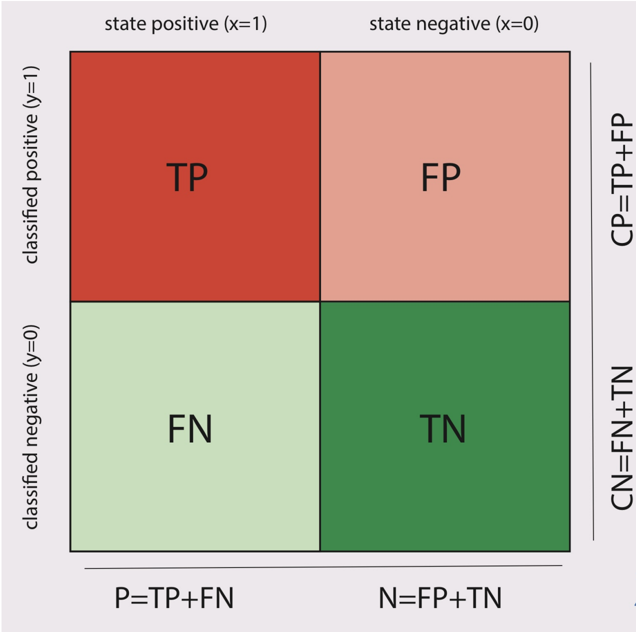


Figure 3.3: 2x2 confusion matrix. Source: *Optimization for Machine Learning Course*

comprehensive understanding of model performance. By building the confusion matrix on test sample data, we can obtain estimates of the main classification performance criteria:

- *Precision* p: the number of TP divided by the number of all classified positive results,
- *Recall* r: the number of TP divided by the number of total actual positives,
- *Specificity* s: the number of TN divided by the number of total actual negatives
- *F1* score: is a metric that combines precision and recall to measure the performance of a binary classifier, it is the harmonic mean of precision and recall. It is particularly useful when the classes in the dataset are imbalanced.

In binary classification, the class prediction for each instance is often made based on a continuous random variable X, which is a "score" computed for the instance (e.g., estimated probability in logistic regression). Given a threshold parameter T, the instance is classified as "positive" if X>T, and "negative" otherwise. The ROC curve is created by plotting the true positive rate (recall, TPR) against the false positive rate (FPR) at various threshold settings. In other words, the ROC curve shows the ratios between true alarms and false alarms. The Area Under the ROC Curve (AUC) is an evaluation metric for checking a classification model's performance, it represents the degree or measure of separability and has a shape as shown in Figure 3.4. It tells how much the model is capable of distinguishing between classes, so higher the AUC, better the model is at predicting 0s as 0s and 1s as 1s. A perfect classifier will have an AUC of 1, while a random classifier will have an AUC of 0.5. An AUC of 1 means that the model has perfect discrimination and it is able to separate the positive and negative classes perfectly. AUC values between 0.5 and 1 are considered to be good models. AUC is a useful metric as it is insensitive to the imbalance of the dataset and it does not depend on the threshold setting of the model. It's a robust metric that can be used in various applications such as medical diagnosis, fraud detection, and natural language processing.

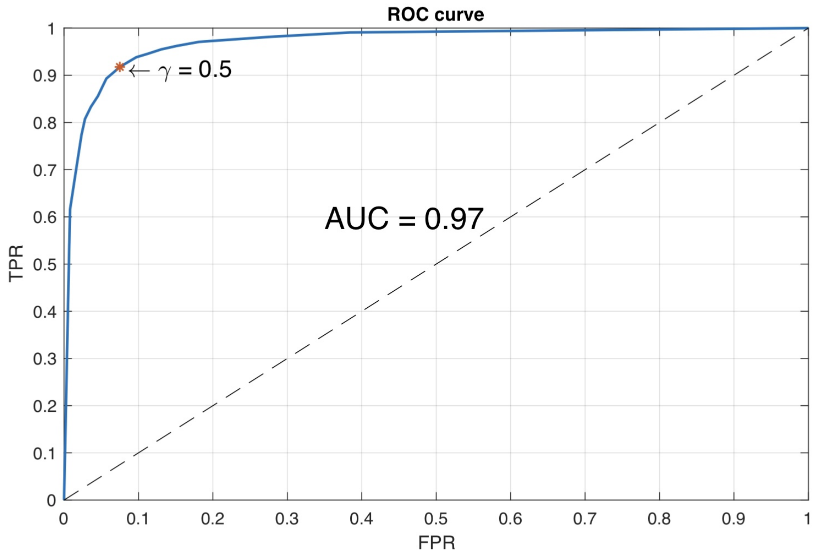


Figure 3.4: Example of a ROC curve and its AUC

However, in the case of highly imbalanced set, the ROC curve is not the best metric to use to evaluate the performance of a method.

In the case of highly imbalanced set, is preferred to use the precision-recall (PR) curve. In an imbalanced dataset, the number of examples in one class (the minority class) is much smaller than the number of examples in the other class (the majority class). Therefore, PR curves can be used in this case to evaluate the model's performance based on its ability to correctly identify examples from the minority class.

The PR curve shows the trade-off between precision and recall for different threshold as shown in Figure 3.5. A high area under the curve represents both high recall and high precision, where high precision relates to a low false positive rate, and high recall relates to a low false negative rate.

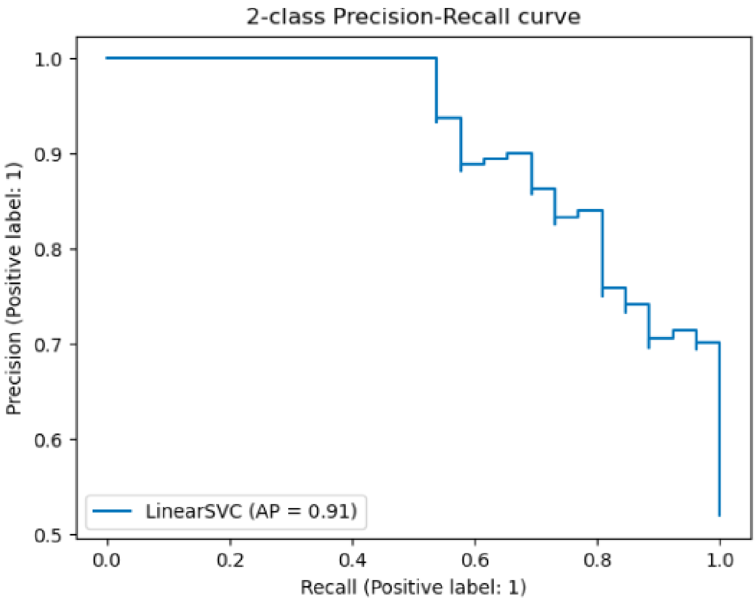


Figure 3.5: Example of a PR curve [39].

Chapter 4

Proposed Framework

4.1 General Pipeline of the work

The proposed work is divided into the following steps as shown in the general scheme in Figure 4.1:

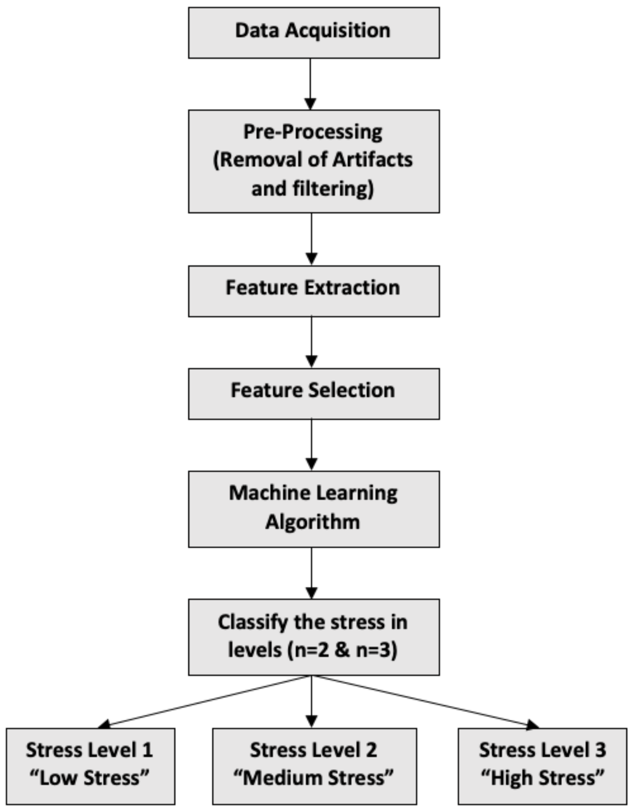


Figure 4.1: General Scheme for analysis of mental stress

- *Step 1*: Data Acquisition;
- *Step 2*: Pre-processing;
- *Step 3*: Feature Extraction;

- *Step 4*: Feature Selection;
- *Step 5*: Machine Learning Algorithm;
- *Step 6*: Classify the stress in levels (n=2 and n=3).

4.1.1 Wearable Sensors Used

The sensors used in this work belongs to the *Biosignalsplux* company, which is a company that specializes in the development and production of hardware and software solutions for biosignal monitoring and analysis.

Biosignalsplux's biosensors can measure a variety of physiological signals, such as electrocardiogram (ECG), electromyogram (EMG), electrodermal activity (EDA), respiratory signals, among others. In the proposed work the biosensors are used to measure in particular ECG, EDA and Respiration Signal.

The whole used system is composed by four components: a 4-channel hub, a ECG sensor, a EDA sensor and a Piezo-Electric Respiration (PZT) sensor.

Figure 4.2: 4-channel biosignalsplux hub.

The 4-channel hub enables the usage of up to 4 sensors simultaneously. The standard version of this hub does not have an internal memory and a digital port which allows the usage of *biosignalsplux* accessories. However, these features can be added to a 4-channel device. The communication with *biosignalsplux* devices is done via Bluetooth or USB.

Three different signals (ECG, EDA, Respiration Signal) have been acquired through *biosignalsplux*, all of them with a sampling frequency equal to 500 Hz and a resolution of 16 bit.

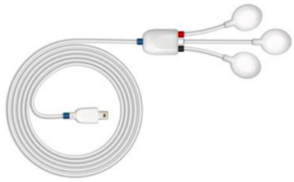


Figure 4.3: Standard biosignalsplux ECG sensor

The low-noise ECG local differential triode configuration, figure 4.3 enables fast application and unobtrusive data acquisition. The state-of-the-art design of the analog frontend on this sensor is specifically targeted at analyzing minutiae in the data and provides medical grade raw sensor data. In Figure 2.2 is shown how the electrodes should be placed in the human body during the acquisition.

This sensor can be used to extract heart rate data and other ECG features, for this reason can be used in a lot of research fields such as biomedical, sports, biofeedback and so on.

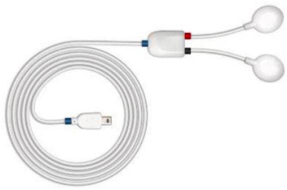


Figure 4.4: Standard biosignalsplux EDA sensor

The *biosignalsplux* EDA sensor is capable of accurately measuring the electrical properties of the skin which changes. The low-noise signal conditioning and amplification circuit design provide optimal performance in the detection of even the most feeble electrodermal skin response events.

The biosignalsplux EDA sensor is designed to acquire the change of skin activity such as sweat with two measuring electrodes. One example is the placement of the electrodes on the anterior side of the hand on two adjacent fingers of interest as is shown in Figure 2.4. Another placement could be on the palm of the hand.

Figure 4.5: Standard biosignalsplux Respiration(PZT) sensor

The *biosignalsplux* Piezoelectric Respiration (PZT) sensor is an entry-level solution for basic respiration data acquisition. This sensor consists of a wearable chest-belt with an integrated localized sensing element that measures displacement variations caused by the volume changes of the thorax or abdomen during respiratory cycles (inhaling/exhaling).

Typical applications of this sensor include respiration monitoring to determine respiration cycles, rates, relative amplitudes, and other features.

4.1.2 Data Acquisition

Forty healthy subjects recruited participated in this study. The subjects were asked to sit comfortably and stay still.

All the data were collected in a controlled laboratory setting using wearable sensors through the *biosignalsplux* platform. In total, 5 active AgCl (silver chloride) electrodes were used with an elastic band for the respiration signal.

All the signals were acquired with a sampling frequency of 500 Hz and a resolution of 16 bit.

In the Data Acquisition process, the first step to be done is the attachment of the sensors to the subject. This step is critical in the data acquisition process, as it ensures that the data collected is accurate and reliable. For this reason, is important to ensure that the subject is comfortable and relaxed during the attachment process to minimize any potential sources of stress that could affect the results.

Once the sensors are attached to the subject, the second step is the application of stressors, which refers to the process of inducing a stressor in the subject to elicit a physiological response that can be measured and analyzed.

In the proposed thesis, the application of stressors is made by asking to the subjects to perform a stress test, which is composed by different tasks as shown in Figure 4.6.

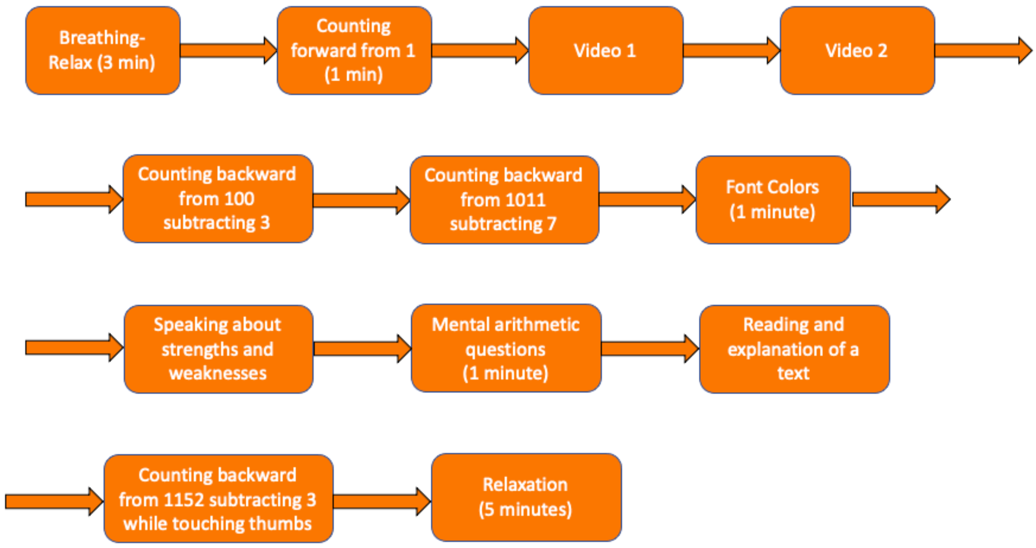


Figure 4.6: Stress Test Pipeline

In particular, the whole test presented in our work takes approximately 45 minutes in which different kind of stressors are applied in a controlled environment.

Both the first and last task are used just as relaxation task, in both tasks the subject is typically asked to sit quietly and relax for a short period of time, while their physiological responses are measured using the sensors that were attached in the data acquisition process.

In particular, the first relaxation task is used as a baseline that provides a reference point for the subject's physiological responses in a relaxed state, against which the responses to the stressor can be compared.

Once the baseline task is done, the stressors are applied to the subject.

At first, it was asked to the subject to watch two different types of several minutes video clips coming from two different movies. The first video was designed to induce positive emotions, such as excitement or fun. The second video, instead, was designed to induce negative emotions, such as fear, anxiety or disgust.

Four arithmetical questions were proposed in this test, three counting backward and one mental arithmetical task. Particularly, the difficulty of the counting backward task increases moving forward with the test. In the mental arithmetical task, it was asked to the subject to solve as much as possible mathematical problems in one minute, the problems may be designed to be challenging or difficult, such as long division or complex multiplication, to induce stress in the subject.

With the arithmetical questions also the Stroop Color task is proposed, which procedure is presented in the subsection 1.4.3. This task can be stressful because it requires the subject to suppress a prepotent response (reading the word) and instead respond based on a different dimension (color).

At the end, a speaking and reading tasks were also proposed. In the speaking part, it was asked to the subject to describe which are his strength and weaknesses in one minute. In the reading task, the subject is asked to read a passage of text under time pressure and once the time is over, is also asked to the subject to explain the text in detail in one minute.

After each task four questions were proposed to all the subjects to establish the value of the person's perceived stress as expressed in self-report ratings (e.g. from 0 to 10). The ground truth was formed as a combination of ratings coming from 4 questions:

1. How *relaxed* were you?
2. How *stressed* were you?
3. How did you *feel*?
4. How *involved* were you?

The first two questions are related to the level of relax and stress during the tasks and the last two questions are related to the valence and arousal measures.

In the presented work, two different classification were made: a binary classification and a three-class classification. Once all the self-assessment coming from the subject were taken, these were mapped into labels using a precise method.

- For the binary classification, the self-assessment rating can be mapped to a binary label, where a rating of stress between 0 and 5 is labeled as "no stress" (label 0) and a rating bigger than 5 is labeled as "stress" (label 1).
- For the three-class classification, the self-assessment rating can be mapped to a three-class label, where a rating of stress between 0 and 4 is labeled as "no stress" (label 0), a rating between 5 and 6 is labeled as " medium stress" (label 1) and a rating bigger than 6 is labeled as "high stress" (label 2).

4.1.3 Pre-Processing

The first step of the pre-processing part consist in the Unit Conversion of the acquired signals. In scientific terms it is always recommended the use of specific units, like electric tension (V) or electric current (A). Each sensor that *Biosignalsplux* commercialized has a datasheet where a transfer function is mentioned for unit conversion be done.

For the ECG, the transfer function is the following and the signal range goes from -1.5 mV to 1.5 mV:

$$ECG(V) = \frac{(\frac{ADC}{2^n} - \frac{1}{2}) \cdot (VCC)}{G_{ECG}} \tag{4.1}$$

$$ECG(mV) = \frac{ECG(V)}{1000} \quad (4.2)$$

In which:

- VCC=3V (operating voltage),
- $G_{ECG} = 1000$ (sensor gain)
- ADC is value sampled from the channel,
- n is the number of bits of the channel.

For the EDA, the transfer function is the following and the signal range goes from $0\,\mu S$ to $25\,\mu$S :

$$EDA(\mu S) = \frac{(\frac{ADC}{2^n}) \cdot (VCC)}{0.12} \quad (4.3)$$

$$EDA(S) = EDA(\mu S) \cdot (10^{-6}) \quad (4.4)$$

Where:

- VCC=3V (operating voltage),
- ADC is the value sampled from the channel,
- n number of bits of the channel.

For the respiration signal, the transfer function is the following and the signal range goes from -50% to 50%:

$$PZT(\%) = (\frac{ADC}{2^n} - \frac{1}{2}) \cdot (100\%) \quad (4.5)$$

Where:

- PZT(%) is the displacement value in percentage (%) of full scale,
- ADC is the value sampled from the channel
- n is the number of bits of the channel

The second step in the pre-processing part consist in applying the denoising method to the acquired signal to remove the noise (acquisition component that is not relevant for the study process).

Filtering of Heart Rate Signal

In the case of ECG, there may be many source of noise due to different factors as presented in section 2.1.1. There are different ways to suppress the noise in the Heart Rate signal. The first way to reduce the noise can be made by using two different digital filters: Notch filter and High-Pass Butterworth Filter. The Notch filter is used to suppress the Powerline Interference at 50Hz with a quality factor equal to 0.5. The High-Pass Butterworth Filter with cut-off frequency of 0.5Hz, instead, is used to eliminate the Baseline Wander (low frequency noise), EMG noise and electrodes motion noise.

In the proposed work, noise reduction is made by using the band pass filter consisting of a low pass filter cascaded by a high pass filter with a filtered band between 0.5 Hz and 40 Hz and the filtered signal has the shape as shown in Figure 4.7.

The purpose of low pass filter is to suppress high frequency noise, filter design using digital filters having integer coefficients allows real time processing speeds.

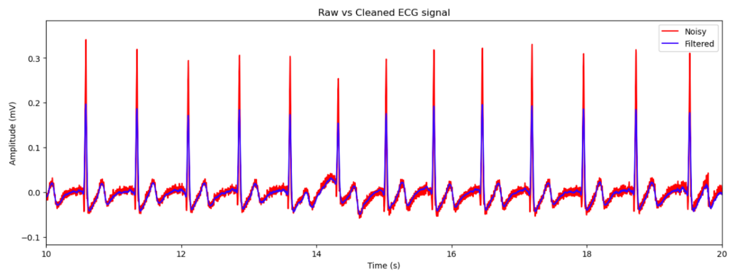


Figure 4.7: Example of Raw vs Cleaned ECG

Filtering of the EDA signal

Noise reduction in the case of EDA signal can be made in different ways. A first method that can be applied is made by: a low-pass filter with cut-off frequency of 0.25 Hz and smoothing techniques such as exponential smoothing and moving average filter have been widely applied to mitigate noises with higher frequency range that desired EDA signals, which are caused by electromagnetic fields or instability of electrode contacts. Also an high pass filter with 0.05 Hz has proven effective to suppress lower frequency noises introduced by variability in electrode impedance, humidity and temperature on skin.

In the presented thesis, noise reduction is made by using a Low-Pass Butterworth Filter with cut-off frequency of 3 Hz and order 4. This filter is mainly used to eliminate the high-frequency noise due to the electrode motion artifact and the filtered output is shown in Figure 4.8.

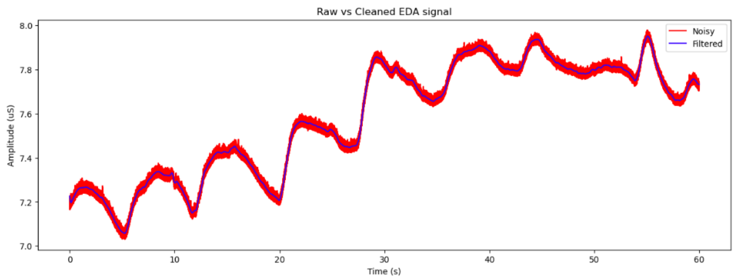


Figure 4.8: Example of Raw vs Cleaned EDA

Filtering of the Respiration Signal

For the Respiration Signal, there may be several approach to clean from noise sources. Wavelet denoising can be used to remove noise from the respiration signal while preserving its temporal and spectral characteristics, this technique decomposes the signal into different frequency subbands and applies different levels of smoothing to each subband, based on the amount of noise present. Adaptive filtering methods can also e used to remove noise from the respiration signal, which adapt to the changing characteristics of the signal and noise. These methods are particularly useful for removing noise that changes rapidly over time or is non-stationary.

In the proposed thesis, the approach used is to use filtering techniques such as a bandpass filter, which can remove noise outside of the frequency range of the respiration signal. In particular, a Pass-Band Butterworth filter with [0.05,3] Hz band-pass region, has been used to remove low-frequency noise and high-frequency noise, respectively, and the filtered output is shown in Figure 4.9.

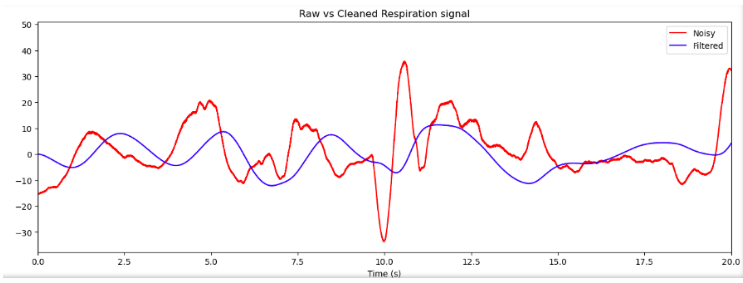


Figure 4.9: Example of Raw vs Cleaned Respiration Signal

It is important to consider the type of noise and characteristics of the respiration signal when selecting the appropriate method for cleaning the signal.

Signal To Noise Ratio (SNR)

A common evaluation metric used to evaluate the performance of digital filtering is SNR.

SNR stands for Signal-to-Noise Ration, which is a commonly used measure to evaluate the quality of a signal in the presence of noise. In the context of biosignals, noise can arise from various sources, SNR is defined as the ration of the power of the signal to the power of the noise. Mathematically, it can be expressed as:

$$SNR = 10 \cdot \log_{10}(\frac{P_{signal}}{P_{noise}}) \tag{4.6}$$

where P_{signal} is the power of the signal, and P_{noise} is the power of the noise. The SNR is usually expressed in decibels (dB), which is a logarithmic unit that allows for easy comparison of ratios. A higher SNR indicates that the signal is stronger relative to the noise, and thus the signal can be more easily distinguished from the noise. A low SNR, on the other hand, indicates that the noise is more dominant and can make it difficult to accurately measure the signal.

In summary, SNR is an important metric in the analysis of biosignals as it provides a measure of the quality of the signal in the presence of noise. Specifically, the signal is higher than the noise if the SNR is higher than 1 (in dB higher than 0) and inversely, if it is below 1 (in dB below 0) the influence of the noise is higher than the influence of the signal and, thus, it might be impossible to recover the signal.

Though signal to noise ratio is important for every type of signal, the process for calculating it can be difficult. In this work an example of SNR values is shown in Table 4.1 with the value of a raw ECG and the cleaned version of the signal with different types of digital filter:

Type of Filter	SNR	SNR in dB
RAW	33.435	30.484
Pass-Band Filter 1	39.648	31.964
Pass-Band Filter 2	38.173	31.635
Notch Filter	35.801	31.078
Notch+High-Pass Filter	32.438	30.221

Table 4.1: SNR values for an ECG signal

Where:

- The first Pass-Band Filter is made by using the *HeartPy* package with cut-off frequency [0.01, 40] Hz and order 3,
- The second Pass-Band Filter is made by using the *Butterworth* Filter with cut-off frequency [0.01, 40] Hz and order 3,
- The third case is made by using a simple Notch Filter with cut-off frequency of 50 Hz and quality factor equal to 1,
- The fourth filter is made by the combination of a Notch Filter and a *Butterworth* High-Pass Filter with cut-off frequency of 0.5 Hz and order 5.

HeartPy is an open-source Python package that provides tools for heart rate variability (HRV) analysis. In this context, it has been used for signal processing.

4.1.4 Feature Extraction

A feature of a biosignal refers to a measurable characteristic or attribute of a physiological signal that can be used to extract useful information about the underlying physiological process or state.

In general, features of biosignals can be extracted using some signal processing techniques, and can provide information that can be used to diagnose or monitor physiological conditions, such as stress, or also to develop predictive models in the field of stress detection.

There may be different kind of features of biosignals:

- Time-domain features: derived from the amplitude and timing of the signal, these type of features show how a signal changes with time. Examples include: Mean Heart Rate (HR), R peaks, Mean RR intervals, Tonic component of SC (SCL), Phasic component of SC (SCR), Frequency of SCR peaks, extraction of inhalation peaks (RSP peaks), Breath rate, Mean BB interval.

- Frequency-domain features: derived from the frequency content of the signal, these type of features show how much of the signal lies within each given frequency band over a range of frequencies. Examples include: Total power of the signal, Absolute power of the low frequency (LF) band (that ranges in the 0.04-0.15 Hz band) , Absolute power of the high frequency (HF) band (that ranges in the 0.15-0.4 Hz band) , LF to HF power ratio (LF/HF).

- Nonlinear features: derived from the nonlinear dynamics of the signal, these features describe complex and often non-repetitive patterns in the signal. Examples include: index of short time variability of BB intervals (SD1), index of long time variability of BB intervals (SD2), approximate entropy of heart rate variability (ApEn).

- Statistical features: derived through statistical analysis of the signal, these features may help to identify data trends and patterns. Examples include: mean value of the signal, variance of the signal (Var), standard deviation of the signal (SD), kurtosis of the signal (Ku), skewness of the signal (Sk).

It should be noticed that these features are used to characterize the signals accurately.

Signal	Time-Domain Features	Frequency-Domain Features	Non-linear Features	Statistical Features
ECG	*meanHR, minHR, maxHR, sdHR, modeHR, nNN, meanNN, SDSD, CVNN, SDNN, pNN50, pNN20, RMSSD, medianNN, q20NN, q80NN, minNN, maxNN, triHRV*	*totalpower, LF, HF, ULF, VLF, VHF, LF/HF, rLF, rHF, peakLF, peakHF*	*SD1, SD2, SD1SD2, ApEn, SampEn*	*min, max, mean, sd, sk, ku, median, q1, q3, q05, q95*
EDA	*minSCR, maxSCR, meanSCR, sdSCR, maxpeakSCR, minpeakSCR, meanpeakSCR, maxdiff, mindiff, meandiff*	*totalpower, LF, HF, ULF, VLF, VHF, LF/HF, rLF, rHF, peakLF, peakHF*	*SD1, SD2, SD1SD2*	*min, max, mean, sd, sk, ku, median, q1, q3, q05, q95*
RESP	*meanRR, minRR, maxRR, sdRR, modeRR, nBB, meanBB, SDSDb, CVNNb, SDNNb, RMSSDb, medianBB, q20BB, q80BB, minBB, maxBB, meant, SDTT, medianTT, SDTT, medianTT, q20TT, q20TT, minTT, maxTT, meanBA, SDBA, medianBA, q20BA, q80BA, minBA, maxBA, meanBW, SDBW, medianBW, q20BW, q80BW, minBW, maxBW*	*totalpower, LF, HF, ULF, VLF, VHF, LF/HF, rLF, rHF, peakLF, peakHF*	*SD1, SD2, SD1SD2, ApEn, SampEn*	*min, max, mean, sd, sk, ku, median, q1, q3, q05, q95*

Figure 4.10: Extracted Features From Biosignals

In Figure 4.10 is shown which are the features that are selected and extracted from each biosignal take into accont. In particular, for both ECG, EDA and Respiration Signal the Frequency-Domain, Non-linear and Statistical Features exctraced are all the same. These features are defined as follows:

- For the Frequency-Domain Features: *LF* is the spectral power of the low frequency band (0.04-0.15 Hz), *HF* is the power of the low frequency band (0.15-0.4 Hz), *ULF* is the power of the ultra low frequency band (0.0-0.0033 Hz), *VLF* is the power of the very low frequency band (0.0033-0.4 Hz), *VHF* is the power of the very high frequency band (0.4-0.5 Hz), *LF/HF* is the LF to HF power ratio, *rLF* is the ratio of low frequency, *rHF* is the ratio of high frequency, *peakLF* is the peak frequency of the low frequency band, *peakHF* is the peak frequency of the high frequency band.

- For the Non-Linear Features: *SD1* is the index of short time variability of RR or BB intervals, *SD2* is the index of long time variability of RR or BB intervals, *SD1SD2* is the ratio between short term and long term variability of RR or BB intervals.

- For the Statistical Features: *sd* is the standard deviation, *sk* is the skewness, *ku* is the kurtosis, *q1* is the first quantile of the signal (25th percentile), *q3* is the third quantile of the signal (75th percentile), *q05* is the 5th percentile of the signal and *q95* is the 95th percentile of the signal.

Instead, for the Time-Domain Features there are different features for each particular signal:

- For the **ECG** signal: *meanHR* is the average heart rate, *minHR* is the minimum value of heart rate, *maxHR* is the maximum value of heart rate,*sdHR* is the standard deviation of HR, *modeHR* is the difference between maxHr and minHr, *nNN* is the number of RR intervals, *meanNN* is the average of RR intervals, *SDSD* is the standard deviation of the successive differences between RR intervals, *CVNN* is the standard deviation of RR intervals, *SDNN* is the standard deviation of RR intervals divided by the mean of the RR intervals, *pNN50* is the proportion of RR intervals greater than 50ms, out of the total number of RR intervals, *pNN20* is the proportion of RR intervals greater than 20ms, out of the total number of RR intervals, *RMSSD* is the root mean square of successive RR interval differences, *medianNN* is the median of the RR intervals, *q20NN* is the 20th percentile of the RR intervals, *q80NN* is the 80th percentile of the RR intervals, *minNN* is the minimum of the RR interval, *maxNN* is the maximum of the RR interval and *triHRV* is the HRV triangular index, integral of the intensity of the RR interval histogram divided by its height.

- For the **EDA** signal: *min* is the minimum value of SCR, *max* is the maximum value of SCR, *mean* is the average value of SCR, *sd* is the standard deviation of SCR, *maxdiff* is the maximum difference between peaks and onsets, *mindiff* is the minimum difference between peaks and onsets, *meandiff* is the average difference between peaks and onsets.

- For the **RESP** signal: *meanRR* is the average respiration rate (RR), *minRR* is the minimum value of RR, *maxRR* is the maximum value of RR, *sdRR* standard deviation of RR, *modeRR* is the difference between maxRR and minRR, *nBB* is the number of Breath to Breath, *meanBB* is the average of BB intervals, *SDSDb* is the standard deviation of the successive differences between BB intervals, *CVNNb* is the standard deviation of BB intervals, *SDNNb* is the standard deviation of BB intervals divided by the mean of the BB intervals, *RMSSDb* is the root mean square of successive BB interval differences, *medianBB* is the median of the BB intervals, *q20BB* is the 20th percentile of the BB intervals, *q80BB* is the 80th percentile of the BB intervals, *meanTT* is the average time between successive Troughs (TT), *SDTT* is the standard deviation of TT intervals, *q20TT* is the 20th percentile of the TT intervals, *q80TT* is the 80th percentile of the TT intervals, *meanBA* is the average valuea of breath amplitudes, *meanBW* is the average value of breath widths (time between peak and through).

All the features were extracted manually by constructing proper functions in the Python environment made for that purpose. There are also many algorithms that can be used for feature extraction from biosignals, depending on the type of signal and the specific features of interest.

Examples of commonly used algorithms for feature extraction are:

- Fourier Transform, technique used to converts a signal from the time domain to the frequency domain, it can be useful for identifying characteristic frequencies in the signal,

- Wavelet Transform, technique used to identify patterns in signals that vary in both frequency and time, it can be useful for identifying transient events in the signal,

- Time-Frequency Analysis: such as the spectogram, can provide information about the changes in frequency content over time, it can be useful for identifying changes in frequency content associated with different physiological processes.

4.1.5 Features Selection

A feature selection algorithm is important for several reasons in the environment of machine learning pipeline for biosignal analysis.

It helps first of all to improve the model performance by selecting only relevant features that are strictly correlated to the work taken into account, in this case correlated to the stress detection field, because irrelevant or redundant features can cause overfitting or may decrease accuracy. It may also reduces the computational complexity of the model by reducing the number of features that need to be analyzed, resulting in faster training and testing times and reduced memory usage.

Since the total number of feature extracted are approximately 131, a feature selection method is needed.

The most important aspect of a feature selection algorithm is that it also enhances interpretability by identifying the most important features that contribute to the model's predictions, helping to understand the underlying physiological processes or conditions that are being analyzed.

The general steps to follow when is needed to implement a feature selection algorithm are the following:

1. Define the problem: determine the objective of the feature selection process,

2. Choose a metric: is important to determine an evaluation metric to measure the performance of the model. In the case of a stress detection algorithm, which is a classification problem, it might be useful to choose the accuracy or the precision as the evaluation metric,

3. Choose a feature selection method: there are various methods for selection features, including filter methods, wrapper methods, and embedded methods. Each method has its strengths and weaknesses, and the choice of method depends on the dataset and the problem at hand,

4. Implement the feature selection algorithm: apply the chosen method to the dataset to select the most relevant features,

5. Evaluate the performance: evaluate the performance of the model with the selected features using the chosen evaluation metric. Compare the perdormance to the baseline model using all features to determine if the feature selection process improved the model's performance.

In the presented work the chosen feature selection algorithm is the Recursive Feature Elimination (RFE) that works as shown in Figure 4.11, which uses a machine learning model to iteratively select the most important features from a dataset, in this particular case, the machine learning model selected is the Logistic Regression model.

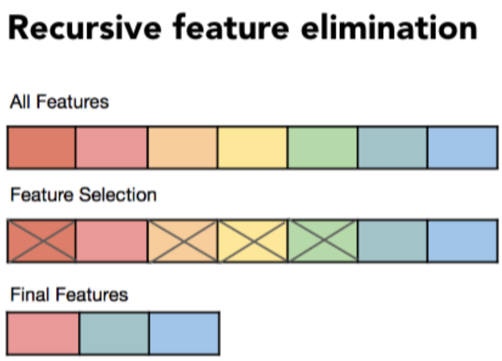


Figure 4.11: How a Recursive Features Elimination works

In the context of logistic regression, RFE involves training a logistic regression model using all the available features in the dataset. After that, the algorithm ranks the importance of each feature based on their coefficients, which indicate how much each feature contributes to the prediction of the target variable. The least important feature is then removed from the dataset, and the logistic regression model is retrained using the remaining features. The process is repeated until we have a predefined number of features or until the performance of the model stops improving.

In the proposed thesis, after a trial and error the final number of feature selected is chosen equal to 20, the combination of this number of features with the different machine learning algorithms taken into account gave the best accuracy achieved.

Other Feature Selection strategies

There are several other feature selection algorithms that can be used with a dataframe and labels to identify the most relevant features for a binary and multiclassification problem. Some other examples may be:

- SelectKBest: this algorithm uses statistical tests (such as for example ANOVA) to select a fixed number of features that have the highest score,
- Lasso (Least Absolute Shrinkage and Selection Operator): this regularization technique can be used for linear models to select features by shrinking the coefficients of less important features to zero,
- Random Forest: this algorithm uses decision trees to identify the most important features by calculating the feature importance measure.

Is also important to notice that it is not mandatory to use the same algorithm for feature selection and classification. Is possible to use any feature selection method (such as RFE or SelectKBest). The purpose of feature selection is to identify the most informative features in your data that are relevant to the target variable, regardless of the classification algorithm you will use. In fact, using different algorithms for feature selection and classification can sometimes be beneficial.

For example, you can use a feature selection method that is less computationally expensive, such as SelectKBest, in combination with a more computationally expensive classification algorithm such as a deep neural network. This can save you time and computational resources without sacrificing performance.

Another method that may be taken into account besides features selection is SHAP (SHapley Additive exPlanations). It is a method for explaining the output of a machine learning model by assigning feature importance scores to the input features. SHAP values provide a unified framework for feature attribution in machine learning models, and can be used to explain the output of any black-box model. The SHAP values represent the contribution of each feature to the model output, they also take into account the interaction between features and the correlation between features. These values can also be used to generate feature importance plots that show the relative importance of each feature in the model as shown in Figure 4.12.

To calculate SHAP values, first of all is needed to train the model on the data, then a set of background data is selected to represent the distribution of

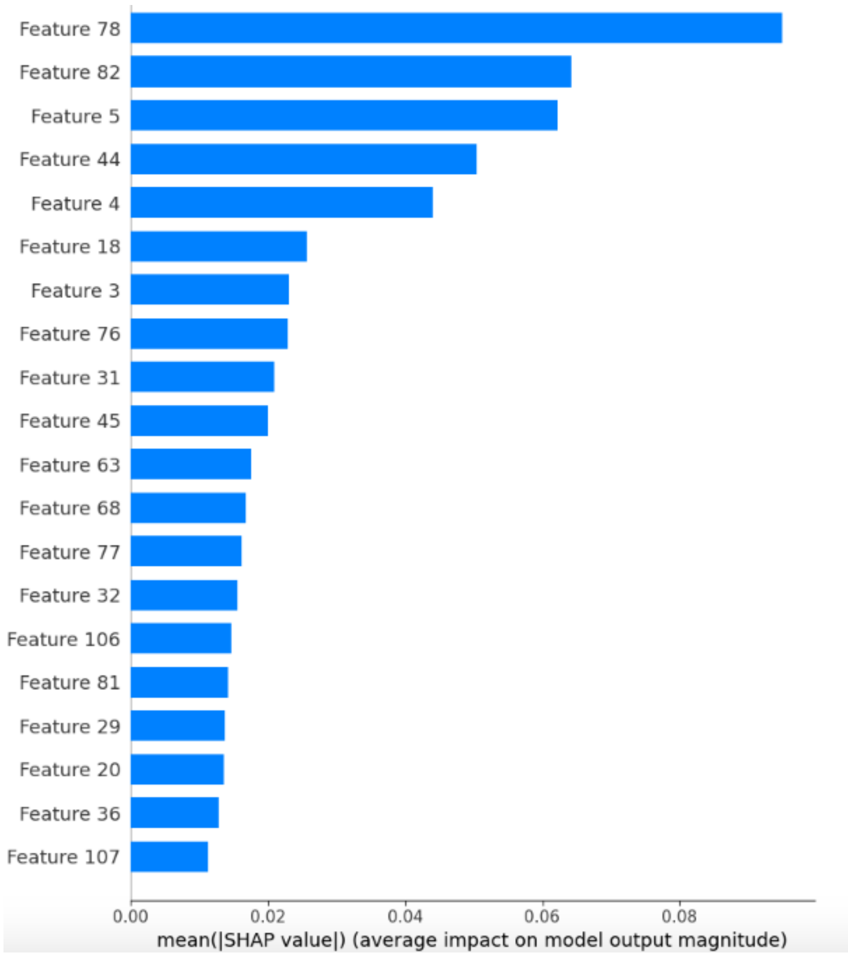


Figure 4.12: Features Importance using SHAP values

the data. Then the SHAP values are calculated for each data point by changing the features and observing oh the model output changes.

Once the values are calculated, they can be used to generate feature importance plots that show the relative importance of each feature in the model, and can be used to identify the most important features for stress detection.

4.1.6 Machine Learning Algorithm

Once the relevant features have been selected, the next step in the machine learning part of a stress detection algorithm is to train a model using the selected features.

The first step to be done is the splitting of the labeled biosignal data into a training set and a testing set, it is an important step since it enables the evaluation of the model's performance on new, unseen data.

The data are usually split randomly, which means that each data point has an equal chance of being included in either the training or testing set. The size of the training set is typically larger than the size of the testing set, with a common split being 80% for training and 20% for testing. However, the optimal split will depend on the size and complexity of the dataset, as well as the performance of the model on the testing data.

In addition to splitting the data into a training and testing set, cross-validation can be also used to further assess the performance of the model. In cross-validation, the data is split into multiple folds, and the model is trained and evaluated on each fold separately.

In the proposed work the data are split randomly with the 80% of the data for training and 20% for test with random state equal to 42.

The third step is to choose a machine learning algorithm, since there are many different algorithms that can be used for stress detection, including decision trees, support vector machines, random forest, neural network, linear discriminant analysis and XGBoost, the choice of the proper algorithm will depend on the specific characteristics of the data and the goals of the project.

The first task taken into account is the binary classification, which involves label the data into two categories or classes, that means to classify biosignals as either indicative of stress or not indicative of stress. The dataset was labeled with binary values (0 or 1) indicating whether the individual was experiencing stress or not during the recording of the biosignals.

First Approach with Leave-One-Out (LOO) Cross-Validation

Several techniques were used to be able to do a comparison and choose the most suitable machine learning algorithm for this stress detection study. The first technique used was the Leave-One-Out (LOO) classification, which can be used for evaluating the performance of a classification model.

In LOO classification, each data point in the dataset is used as a test instance once, while the rest of the data points are used as training instances to train the model. To perform LOO classification, the model is trained on all data points in the dataset except for one, which is used as the test instance. The model then predicts the class of the test instance based on its features, and the predicted class is compared to the actual class label of the test instance. This process is repeated for every data point in the dataset, with each data point being used once as a test instance.

The first classification algorithm used with LOO technique is Random Forest, which builds an ensemble of decision trees, where each tree is trained on a random subset of the features and data points. For this reason, combining LOO cross-validation with Random Forest can be a powerful approach for building a stress detection algorithm. This approach led to an accuracy of 69%.

The second classification algorithm used with LOO technique is Linear Discriminant Analysis, which is a statistical technique used to find a linear combination of features that can best separate two or more classes of data. In this context, LDA can be used to find a linear boundary that separates stress and non-stress states based on the features extracted. Combining LOO cross-validation with LDA can be useful to build a stress detection algorithm, since LOO can be evaluated on a diverse range of data points and LDA can provide a linear boundary that separates stress and non-stress states. This approach led to an accuracy of 50%.

The third classification algorithm used with LOO technique is Support Vector Machine (SVM), which is useful to find an hyperplane that best separates two classes of data, with the aim of maximizing the margin between the hyperplane and the data points closest to it. Furthermore, SVM can provide a hyperplane that separates stress and non-stress states based on the features extracted. This approach led to an accuracy of 58%.

The fourth classification algorithm used with LOO technique is Penalized regression with Elastic Net , which is a statistical method that combines both L1 and L2 regularization penalties to perform variable selection and avoid overfitting in regression models. Penalized logistic regression with Elastic Net regularization can be used for classification tasks, where it can find a set of predictor variables that are most important for predicting the target variable. Furthermore, Elastic net can provide a sparse set of important features that are most useful for predicting stress, while also avoiding overfitting. This approach led to an accuracy of 62%.

The last classification algorithm used with LOO technique is Light Gradient Boosting Machine (LGBM), which is a gradient boosting framework that uses tree-based learning algorithms. This type of method is often used for multiclass classification problems and can handle high-dimensional datasets with ease. The use of both LOO and LGBM can provide accurate and efficient predictions for multiclass classification problems, while also handling high-dimensional datasets. This approach led to an accuracy of 63%.

In Figure 4.13 the previous explained results are summarized:

Machine Learning Algorithm	Accuracy
Random Forest	*69%*
Support Vector Machine	*58%*
Linear Discriminant Analysis	*50%*
Penalized Regression with Elastic Net	*62%*
Light Gradient Boosting Machine	*63%*

Figure 4.13: Results of the Machine Learning Algorithms used with LOO technique.

The algorithm that present the best performance in the binary classification using LOO technique is Random Forest, which brings to a confusion matrix of this shape:

Second Approach with Recursive Feature Elimination (RFE)

In order to achieve better performance, to improve the interpretability of the model and to reduce the computational time, before the machine learning algorithm is applied the RFE technique.

The first algorithm used with RFE technique is XGBoost. XGBoost is an algorithm used for binary classification problems, it builds a series of decision trees, where each tree tries to correct the mistakes of the previous tree. During the training part, the algorithm assigns weights to the training examples based on their importance, and uses gradient descent optimization to find the optimal weight for each decision tree.

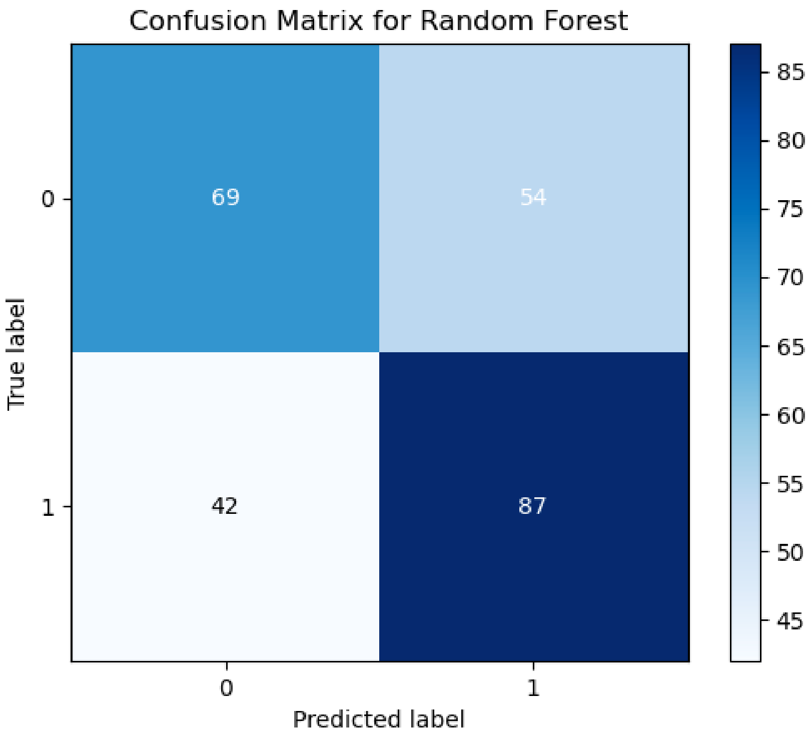


Figure 4.14: Confusion Matrix of Binary Classification with Random Forest

Once all the decision trees are trained, their predictions are combined using a weighted average, and the final prediction is made by comparing the weighted average to a threshold value. This approach led to an accuracy of 84%.

The second algorithm used with RFE technique is Linear Discriminant Analysis, which leads to better performance, in particular leads to an accuracy of 75%. However, it's important to note that the optimal approach may vary depending on the specific dataset and classification task. In some cases, it may be more effective to apply LDA before RFE or use other features selection methods altogether.

The third algorithm used with RFE technique is Neural Network. After selecting the relevant features, you would train the neural network on the selected features using the training set. The neural network can have multiple hidden layers and use activation functions to map the input features to the output labels. In particular, two different dense layer were used: the first one by using a Rectified Linear Unit (ReLU) activation function, which is a non-linear function or piecewise linear function that will output the input directly if it is positive, otherwise, it will output zero, the second hidden layer use as activation function a sigmoid function, which is a non-linear function that guarantee that the output of this unit will always be between 0 and 1. The last two parameters to tune are the loss function and the optimizer. The chosen loss function is the binary crossentropy, which is a loss function that is commonly used in binary classification problems, which measures the difference between the predicted probability distribution and the true probability distribution. Specifically, it calculates the cross-entropy loss between the predicted output and the true output. Instead, the chosen optimizer is the Adam optimizer, which is commonly used to update the weights and biases of a neural network during training. It is an adaptive learning rate optimization algorithm that dynamically adjusts the learning rate of each weight based on the first and second moments of the gradient. This can help to speed up the convergence of the neural network during training. The following approach led to an accuracy of 79%.

Machine Learning Algorithm	Accuracy
XGBoost	*84%*
Linear Discriminant Analysis	*75%*
Neural Network	*79%*

Figure 4.15: Machine Learning Algorithms used with RFE.

The most performing algorithm for binary classification is XGBoost with the previous use of an RFE algorithm for features selection. This method leads to the following confusion matrix:

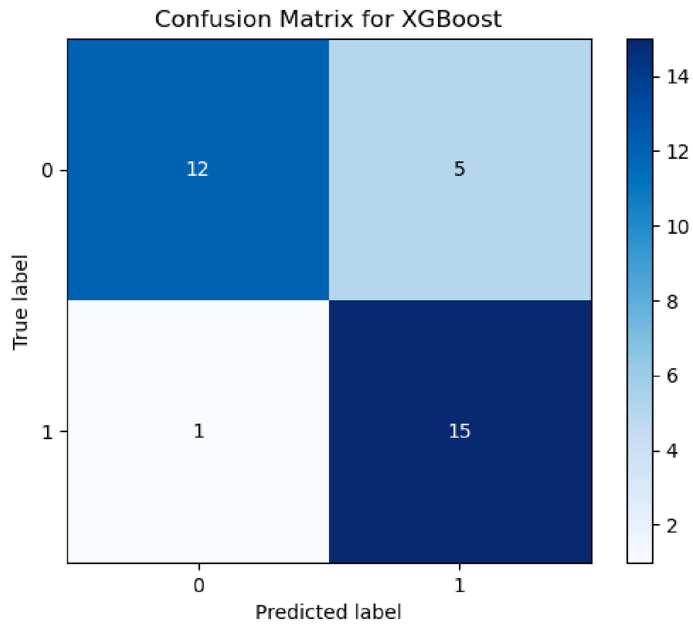


Figure 4.16: Confusion Matrix for XGBoost with RFE method.

The XGBoost model has a ROC curve of this shape:

This model presents also an AUC=0.82 of the ROC curve. This result means that the model's performance is reasonably good at distinguishing between the two classes in the binary classification problem. In pratical terms, this means that if you use this model to classify new instances, you can expect it to correctly classify positive instances as positive about 82% of the time, while minimizing the number of false positives.

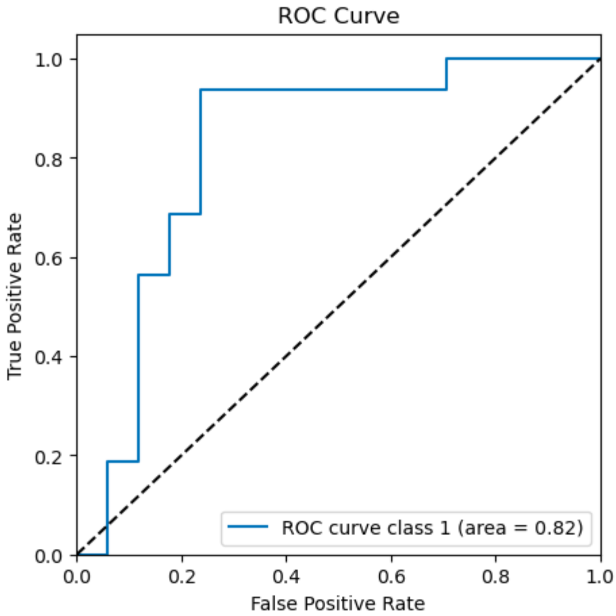


Figure 4.17: ROC curve of XGBoost algorithm

Multi-Classification (3 Classes)

The second machine learning algorithm is for the Multiclassification case (3 classes): labeled as "no stress" (label 0), "medium stress" (level 1) and "high stress" (level 2). This algorithm presents worst performance with respect to the Binary Classification problem.

This may due to several reasons, one problem is related to the higher level of complexity of the problem. With three classes, the algorithm has to learn to distinguish between three different categories instead of just two. This can make it harder for the algorithm to find the decision boundaries between the different classes and lead to worse performance.

Another problem is the potential for imbalanced classes. For example, one or more classes may have significantly fewer data points that the others, making it harder for the algorithm to learn the characteristics of those classes and leading to poorer performances. In the presented work, in the binary classification, the dataset is balanced having 85 samples labeled as "no stress" and 77 samples labeled as "stress"; for the 3-class classification, instead, there are 65 samples labeled as "no stress", 50 samples labeled as "medium stress" and 47 samples labeled as "high stress". This may lead to a imbalanced dataset in a low-level.

Additionally, a 3-class classification problem may be more prone to overfitting than binary problem. This is because there are more parameters for the algorithm to learn, which can make it more likely to memorize the training data instead of learning general patterns.

4.1.7 Related works

Stress detection using wearable sensors

Nowadays sensors play an important role in medical science and related applications. These are generally used for the detection and measurement of various diseases and their levels. The devices which use one or more sensors such as HR, ST, GSE, RR, ACC, BP sensors are considered as wearable sensors. Stress can be monitored using only one physiological signal also, but the results could be inappropriate. *Stress Detection using ECG* HRV significantly contributes to stress detection due to its close relationship with the autonomic nervous system. With the help of the combination of different HRV characteristics, it is possible to distinguish between rest, physical and mental conditions, as HRV is sensitive to any change in the mental or physical state [38]. Also, the reactivity and recovery from mental and physical stress are strongly correlated with the HRV parameters associated with parasympathetic activity. In [40] a small and lightweight sensor named RF-ECG was used to record the real-time ECG signals with 204 Hz sampling rate. Stress detection was used here to address the confusion issues of facial recognition to activate the relaxation service. Negative emotions and stress were recognized with 83.33% of accuracy by combining emotion recognition and stress detection.

As there are many techniques available for the estimation of stress from physiological signals, its fine-grained assessment is still a challenge. Tania Pereira et al. studied various HRV metrics for stress level assessment using a short-time window, where a sub-set of HRV metrics namely AVNN, rMSSD, SNDD and pNN20 showed consistent differences between stress, and non-stress phases.

A publicly available dataset "WESAD" [41] was used and in particular ACC, ECG, BVP, BT, respiration, EMG and EDA data were used. In this study, some machine learning algorithms have been used such as KNN, Linear Discriminant Analysis, Random Forest, Decision tree, kernel SVM classifiers, and achieved an accuracy of up to 81.65% and 95.20% for 3-class and binary class respectively. They have also applied a simple feed-forward deep learning technique which increased the accuracy up to 84.32% and 95.21% for that respective classes which showed that deep learning is better than the traditional machine learning classifiers and the generalization is possible with the Leave-one-subject-out evaluation scheme. In [42] ECG, HR and GSR were used during a Stroop Test. A fuzzy logic algorithm was developed and the data was trained in the Adapative Neuro-Fuzzy Interface System (AVFIS) and the C programming language was used which has support for detectiong data from serial ports used by Arduino boards. An accuracy of 72% was calculated for preditcing stress levels, by conducting a one-tailed Spearman's Rank Correlation coefficient test, it results that this system is very capable of precisely predicting the user's level of mental stress with a high level of consistency. The advantages are that this system can be used to increase awareness of stress among mouse users, but the limitations are that as the sensors were self-made, they have to first tested and validated by experts before using in the experiment.

In [2] an overview of multimodal analysis studies, along study population, stimuli used, biosignals recorded, classification scheme and best accuracy received is summarized in Figure 4.2.

Study	Population	Stimuli	Biosignal Used	Classification System	Best Accuracy Achieved
Kim et al. [43]	Videos, images, sounds	EDA, SKT, ECG	SVM	78.4%	
Healey et al. (2005) [44]	24 subjects	Driving Task	ECG, EMG, EDA, RSP	LDA	97.30%
Giakoumis et al. [45]	21 subjects	Puzzle, Memory Tasks	EDA, ECG, Body activity	LDA	96.60%
Setz et al. (2010) [46]	33 subjects	MIST	EDA, ECG, RSP	LDA, SVM	82.80%
Al-shargie et al. (2015) [47]	12 subjects	MIST	EEG	SVM	94.00%
Xia et al. (2018) [48]	22 subjects	Mental Arithmetic Task	EEG, ECG	PCA, SVM	79.54%
Cheema et al. (2019) [49]	30 subjects	Institute examination	PCG, ECG	LS-SVM	96.67%

Table 4.2: Overview of Stress Detection Algorithms

Chapter 5

Discussion and Results

5.1 Discussion

There may be several limitations to a stress detection algorithm that uses biosignals from wearable sensors and self-assessment of the subjects during a stress test.

First of all, all the stress tests have been performed in a laboratory or a well-controlled environments, by simulating through tasks stress conditions. The induced stressors were usually intense in order to achieve a prominent and measurable amount of acute stress, as in as in [2]. However, in real life conditions, stressors are usually complex procedures that involve many aspects of human personality or multiple stressors occur due to the complexity of the way of living. Different sessions of an experiment would be employed to cover different stressors but their simultaneous application is not always possible.

The effectiveness of stress inducing tasks is also a question under investigation. Their effectiveness are subject to originality and habituation. Other important dimensions are its duration and associated involved processes (e.g., habituation during continuous/repeated exposure, competition with opposing external stimuli and self-regulation). In [2] the most effective stressful effects are observed at the beginning of the experimental procedure. The engagement of the participant wanes as the experiment progresses, and its maintenance can be partially achieved with relaxation intervals between tasks. Furthermore, the stress detection algorithm may only be able to detect certain types of stress, such as physiological stress, but may not be able to detect other types of stress, such as psychological stress or stress related to specific tasks or events.

However, both simulated stressors and real stressors have their own advantages and disadvantages, and the choice depends on the specific context and goals of the algorithm. Simulated stressors can provide controlled and repeatable conditions for testing and training the algorithm, they can also be easily manipulated to create different levels and types of stressors. This allows for standardized and consistent evaluation of the algorithm's performance.

Real stressors, instead, can provide a more reliable setting for evaluating the algorithm. They can capture the complexity and variability of real-life stressors and the dynamic changes in physiological and psychological responses over time, but they can be also unpredictable and difficult to replicate.

In the case of developing an algorithm for real-time stress monitoring in daily life, using real stressors may be appropriate. Alternatively, if the goal is just to compare the performance of different algorithms in a controlled environment, simulated stressors may be more appropriate.

The second limitation is about the self-assessment of the subjects. For example, in [3], two datasets have been used to show the versatility of their work on predicting individuals' stress both in a controlled laboratory setting while acquiring brainwaves using a traditional wired-EEG and at actual job sites while recording workers' brainwaves using a wearable EEG device.

For the first dataset, the participants rated each video regarding the levels of arousal, valence, like/dislike, dominance and familiarity. The authors used the subjective rating to label signals as high stress and low-stress conditions.

In the second dataset, the authors collected EEG signals using a wearable EEG device from three real construction sites and EEG signals were collected at least two hours after workers started their work to eliminate possible individual biases from varying baseline stress levels among different subjects. In order to assess subjects' stress levels and label the signals, subjects' cortisol samples were collected from their saliva after each session; high level of cortisol is associated with higher stress.

Considering the higher EEG recording quality in the first dataset, since the signals were recorded in a controlled laboratory, compared to the second dataset, it was expected that the first one would lead to a higher classification accuracy. In Figure 5.1 is shown the algorithms prediction accuracy for the two different datasets.

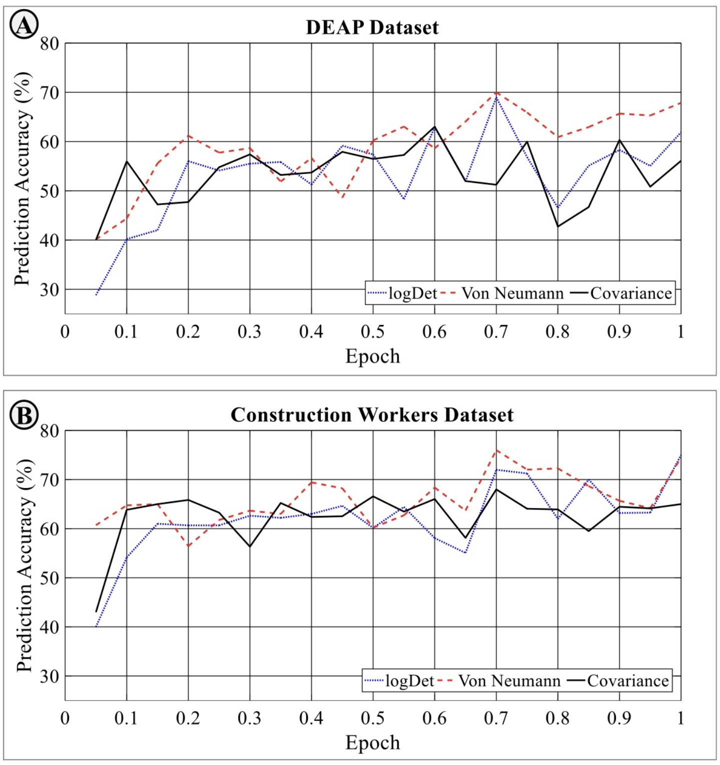


Figure 5.1: Algorithms prediction accuracy.

One possible explanation for this result is the selected stressors to induce stress in the tested datasets; real job sites stressors have been used to induce stress in the second dataset, while in the first one virtual stimuli are used. The stressors in actual jobs site induce higher stress to compare to a virtual stimulus such as watching the video. One other reason for the lower classification accuracy of the suggested framework can be related to the labeling process. In the first dataset, a subjective survey was used to label the individual's stress. However, in the second dataset, the authors labeled the data by measuring individuals' stress by using the cortisol level. There exists higher uncertainty while labeling the data using subjective methods as compared with labeling based on stress hormone.

Another issues, as explained in [2], is that most biosignals are susceptible to noise or artifacts due to individual's body parts movements or activities.

Signal denoising includes techniques such as low-, band- and high-pass filtering, notch filtering, the Least Mean Squares (LSM) or Recursive LSM, wavelet denoising, Principal Component Analysis (PCA), Independent Component Analysis (ICA), and their variations. In the presented work, the mainly used denoising method is digital filtering. This may lead to several problems:

- Loss of information about the signal: digital filtering is a linear process that removes frequency components that are above or below a certain threshold. Is important to set the threshold in such a way that important frequency components about the signal are not removed, otherwise it may lead to a loss of signal information,

- Introduction of artifacts: digital filtering can introduce artifacts in the signal, in the case in which the filter settings are not optimized for the particular signal being processed,

- Inability to remove all types of noise: digital filtering is effective at removing certain types of noise, such as high-frequency noise, such as high-frequency noise, but it may not be effective at removing other types of noise, such as baseline wander or motion artifacts,

Other methods, such as wavelet transforms, EMD, and ICA, can be useful in cases where the noise is non-stationary or has a complex frequency spectrum. These methods can also be more effective at separating the signal of interest from noise that is mixed in with the signal. However, they may be more computationally intensive and require more expertise to implement effectively. Overall, digital filtering is a useful method for denoising biosignals, but it is important to consider the specific characteristics of the signal and the noise when deciding on the best approach.

Another limitation is related to the limited dataset available for the stress detection algorithm. When developing a stress detection algorithm, the quality and amount of data used to train and test the model is crucial to determine how reliable is the model trained.

In this thesis, a dataset of 35 healthy subjects were recruited to participate in this study. The subjects consisting of 28 men and 7 women, are aged between 20 and 44, having a 80% of the dataset composed by men and just the 20% with women, this may already lead to an unbalanced dataset.

When using a limited dataset there may be different issues: the first one is related to the fact that the algorithm may not have sufficient training data to learn the necessary patterns and features that are essential for detecting stress. With a smaller sample size, the algorithm may not generalize in a proper way and may also produce less reliable results when presented with new data.

Another problem that may arise in the field of using a limited dataset is overfitting. When there is a limited and unbalanced amount of data, the algorithm may focus on specific features and patterns within the data rather than generalizing to broader trends. This may lead to the problem that the algorithm may not be able to detect stress in new data that differs significantly from the original dataset. Therefore, if the dataset is not diverse enough, there may be the problem of having biased results and this may also lead to the algorithm only detecting certain types of stress and not others.

Finally, a limited dataset can result in the lack of variability in the training data, this may lead to the algorithm not being able to handle the range of stress levels that occur in real-world situations. This may also be one of the reason why the algorithm performs better in the binary classification case and worst in the three-classification case.

In summary, a limited dataset for a stress detection algorithm can result in insufficient training data, overfitting, biased results and lack of variability in the algorithm's performance. For this reason, is essential to ensure that the dataset used for training is diverse and large enough to produce accurate and reliable results when detection stress.

Another problem about acquiring of the data from people can be related to the quality of the data taking into account. The quality of the dataset collected can affect the performance and reliability of the algorithm, and the quality may be influenced by some factors, including:

- Movement artifacts: any kind of movement of the subject may lead to a motion artifact in each signal that sometimes can reduce the quality,

- Signal acquisition hardware: the quality can be affected by the quality of the hardware used to acquire it, including amplifier, electrode lead and other equipment. For example, in some subjects it was possible to find a saturation during the acquisition of the EDA signal, in that case, the information in that period of time during the saturation is lost,

- Patient-specific factors: individual variability in physiological parameters such as heart rate or breathing rate may change depending on the specific subject.

Poor data quality can result in inaccurate or unreliable results, which can limit the usefulness of the algorithm in practical applications.

5.2 Results

Stress detection, self-assessment and analysis in humans are significant processes in order to confront this phenomenon.

From the above discussed approaches, it is clear that physiological sensor signals can be used to detect stress level of the individual where physiological sensing devices are used to collect these signals. To apply a stress detection method some steps are required in order to obtain reliable results.

The first important point to note is that the classification process must be kept in mind from the moment the signals are obtained (during the acquisition process). Without a robust acquisition process the signals become useless for a correct later classification. All operation carried out in later stages will be useless if the first step of the whole process is not carried out correctly. For this reason, the acquisition process is crucial and is also important to prepare the subjects in order to avoid voluntary movements.

In addition, a correct application of the different filters during the pre-processing stage will be crucial for the following phases. For instance, if the signal is not filtered appropriately, noise can be introduced into the data and it may lead to some mistakes in the feature extraction algorithm.

Once the signals have been pre-processed, the next important step is to obtain the different features that allow us to quantify them. This will be done in a further process which is the classification, but the crucial previous step is the feature selection. There is no evidence in the literature on the number of variables or the minimum number of functions that should be used. In this context, is important to do a trial and error approach, which is an approach that involves testing different feature selection algorithms until the desired outcome is achieved. This approach typically involves trying multiple solutions or methods, and then evaluating their effectiveness based on the results.

This approach is often used in situations where the problem is complex and there is no clear or obvious solution, and where it is not possible to simply apply a proven formula or method.

Furthermore, using more physiological signals (such as ECG, EDA and Respiration Signal) has the advantage that it enables to monitor several systems and, as a consequence, several types of responses provide a better mapping of the physical, psychological and cognitive state of a subject. Nonetheless, a disadvantage is that the use of different signals makes the system more complex, more difficult to maintain and it has a higher computational cost when using a classifier.

According to the analysis performed previously, there are specific biosignals that present consistent pattern so as to be efficient and specific in discriminating stress conditions. For instance, Heart Rate (HR) related features are the most prominent features which increases significantly during stress, however, this may be attributed only to the arousal dimension. Also Skin Conductance Response (SCR) and Skin Conductance Level (SCL) appear also to be consistent measures being typically increased during stress.

Another important issue to concern is the substantial intra- and inter-individual variability of the stress response [2]. For a given person, the same stimulus or condition may elicit a strong or weaker stress response depending on varying social (e.g., high vs. low peer pressure), contextual (e.g., high reward vs. low reward), and cognitive-emotional parameters (e.g, instruction to engage in cognitive appraisal of the situation). Different persons may develop very disparate stress responses to the same stressor. Moreover, different stressor types may be more appropriate and specific to particular types of stress.

Various machine learning algorithms were applied to build classification model. The most performing models in both binary and three-class classification were XGBoost and Random Forest. These methods are particularly effective for a stress detection task because they can handle complex time-series data and are robust to noise, which may be very common when it comes to biosignals.

In addition, both methods provide feature importance measures, which can help identify the most important features that contribute to stress detection. This can provide insights into the underlying physiological mechanisms of stress and make the model more easy to be interpreted.

5.3 Conclusion

The present study employed diverse signal processing and machine learning techniques to create a stress detection framework that utilizes physiological signals from multiple modalities. This framework takes into account the variations in biosignal patterns across different individuals when exposed to various stressors.

Based on the results obtained in this study, it can be concluded that a stress detection algorithm using biosignals acquired from wearable sensors can achieve good performance in terms of accuracy. This suggests that the use of such technology has the potential to provide reliable and accurate stress detection in various settings.

The application of a combination of Recursive Feature Elimination (RFE) and XGBoost algorithms showed good results, especially in the case of binary classification for stress detection. The performance of this combination algorithm was noteworthy, achieving an overall accuracy of 84%.

Furthermore, many wearable devices are available in the market which can be used in physiological signal data collection. These devices are user-friendly and give less error and noise. Hence, these can be used to monitor and measure stress levels without affecting the user's daily functioning.

However, further research is necessary to improve the performance of the algorithm and to investigate its applicability in real-world scenarios.

Bibliography

[1] American Psychological Association. «APA Dictionary of Psychology». In: (2015) (cit. on p. 2).

[2] Giorgos Giannakakis, Dimitris Grigoriadis, Katerina Giannakaki, Olympia Simantiraki, Alexandros Roniotis, and Manolis Tsiknakis. «Review on Psychological Stress Detection Using Biosignals». In: *IEEE Transactions on Affective Computing* 13.1 (2022), pp. 440–460. DOI: `10.1109/TAFFC.2019.2927337` (cit. on pp. 2–4, 6–13, 63, 65, 67, 69).

[3] Houtan Jebelli, Mohammad Mahdi Khalili, and SangHyun Lee. «A Continuously Updated, Computationally Efficient Stress Recognition Framework Using Electroencephalogram (EEG) by Applying Online Multitask Learning Algorithms (OMTL)». In: *IEEE Journal of Biomedical and Health Informatics* 23.5 (2019), pp. 1928–1939. DOI: `10.1109/JBHI.2018.2870963` (cit. on pp. 2, 3, 30, 66).

[4] Patricia E G Bestelmeyer, Sonja A Kotz, and Pascal Belin. «Effects of emotional valence and arousal on the voice perception network». en. In: *Soc. Cogn. Affect. Neurosci.* 12.8 (Aug. 2017), pp. 1351–1358 (cit. on p. 3).

[5] Thi-Dung Tran, Junghee Kim, Ngoc-Huynh Ho, Hyung-Jeong Yang, Sudarshan Pant, Soo-Hyung Kim, and Guee-Sang Lee. «Stress Analysis with dimensions of valence and arousal in the Wild». en. In: *Appl. Sci. (Basel)* 11.11 (June 2021), p. 5194 (cit. on p. 3).

[6] Wikipedia. *Elettrocardiogramma — Wikipedia, L'enciclopedia libera.* [Online; in data 14-marzo-2023]. 2023. URL: `http://it.wikipedia.org/w/index.php?title=Elettrocardiogramma&oldid=132172573` (cit. on p. 5).

[7] Larisa G Tereshchenko and Mark E Josephson. «Frequency content and characteristics of ventricular conduction». In: *Journal of electrocardiology* 48.6 (2015), pp. 933–937 (cit. on p. 4).

[8] Robert B Aird and Y Gastaut. «Occipital and posterior electroencephalographic ryhthms». en. In: *Electroencephalogr. Clin. Neurophysiol.* 11.4 (Nov. 1959), pp. 637–656 (cit. on p. 7).

[9] biosignalsplux. *EEG biosignalsplux.* `https://support.pluxbiosignals.com/wp-content/uploads/2021/11/Electroencephalography_EEG_Datasheet.pdf` [Accessed: (Use the date of access)]. 2022 (cit. on p. 8).

[10] Almudena Bartolomé-Tomás, Roberto Sánchez-Reolid, Alicia Fernández-Sotos, José Miguel Latorre, and Antonio Fernández-Caballero. «Arousal Detection in Elderly People from Electrodermal Activity Using Musical Stimuli». In: *Sensors* 20.17 (2020). ISSN: 1424-8220. DOI: 10.3390/s20174788. URL: https://www.mdpi.com/1424-8220/20/17/4788 (cit. on p. 9).

[11] biosignalsplux. *EDA biosignalsplux*. https://support.pluxbiosignals.com/wp-content/uploads/2021/11/Electrodermal_Activity_EDA_Datasheet.pdf [Accessed: (Use the date of access)]. 2022 (cit. on p. 10).

[12] Keiko Ikeda, Kiyoshi Kawakami, Hiroshi Onimaru, Yasumasa Okada, Shigefumi Yokota, Naohiro Koshiya, Yoshitaka Oku, Makito Iizuka, and Hidehiko Koizumi. «The respiratory control mechanisms in the brainstem and spinal cord: integrative views of the neuroanatomy and neurophysiology». In: *The journal of physiological sciences : JPS* 67.1 (Jan. 2017), pp. 45–62 (cit. on p. 10).

[13] Aamir Arsalan, Syed Muhammad Anwar, and Muhammad Majid. «Mental Stress Detection using Data from Wearable and Non-wearable Sensors: A Review». In: *CoRR* abs/2202.03033 (2022). arXiv: 2202.03033. URL: https://arxiv.org/abs/2202.03033 (cit. on pp. 11, 15).

[14] biosignalsplux. *RESP biosignalsplux*. https://support.pluxbiosignals.com/wp-content/uploads/2021/11/Respiration_PZT_Datasheet.pdf [Accessed: (Use the date of access)]. 2022 (cit. on p. 11).

[15] Denisse Castaneda, Aibhlin Esparza, Mohammad Ghamari, Cinna Soltanpur, and Homer Nazeran. «A review on wearable photoplethysmography sensors and their potential future applications in health care». In: *International journal of biosensors & bioelectronics* 4.4 (2018), pp. 195–202 (cit. on p. 12).

[16] Noor Halim, Khairul Sidek, and H. Mansor. «Stress Recognition Using Photoplethysmogram Signal». In: *Indonesian Journal of Electrical Engineering and Computer Science* 8 (Nov. 2017), pp. 495–501. DOI: 10.11591/ijeecs.v8.i2.pp495-501 (cit. on p. 13).

[17] Anjana Bali and Amteshwar Singh Jaggi. «Clinical experimental stress studies: methods and assessment». In: 26.5 (2015), pp. 555–579 (cit. on pp. 13, 14).

[18] Reg Arthur Williams, Bonnie M Hagerty, and Geneal Brooks. «Trier Social Stress Test: a method for use in nursing research». en. In: *Nurs Res* 53.4 (July 2004), pp. 277–280 (cit. on p. 14).

[19] Federica Scarpina and Sofia Tagini. «The Stroop Color and Word Test». In: *Frontiers in Psychology* 8 (2017). ISSN: 1664-1078. DOI: 10.3389/fpsyg.2017.00557 (cit. on p. 14).

[20] Katarina Dedovic, Robert Renwick, Najmeh Khalili Mahani, Veronika Engert, Sonia J Lupien, and Jens C Pruessner. «The Montreal Imaging Stress Task: using functional imaging to investigate the effects of perceiving and processing psychosocial stress in the human brain». In: *Journal of psychiatry & neuroscience : JPN* 30.5 (Sept. 2005), pp. 319–325 (cit. on p. 15).

[21] Alexandra L. Shilton, Robin Laycock, and Sheila G. Crewther. «The Maastricht Acute Stress Test (MAST): Physiological and Subjective Responses in Anticipation, and Post-stress». In: *Frontiers in Psychology* 8 (2017). ISSN: 1664-1078. DOI: 10.3389/fpsyg.2017.00567. URL: https://www.frontiersin.org/articles/10.3389/fpsyg.2017.00567 (cit. on p. 15).

[22] Wikipedia contributors. *Electromyography — Wikipedia, The Free Encyclopedia.* [Online; accessed 23-March-2023]. 2023. URL: https://en.wikipedia.org/w/index.php?title=Electromyography&oldid=1135070884 (cit. on p. 16).

[23] Thion Ming Chieng, Yuan Wen Hau, and Zaid Omar. «The Study and Comparison between Various Digital Filters for ECG De-noising». In: *2018 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES)*. 2018, pp. 226–232. DOI: 10.1109/IECBES.2018.8626661 (cit. on pp. 17, 31).

[24] Rahul Kher. «Signal Processing Techniques for Removing Noise from ECG Signals». In: *J Biomed Eng* 1 (2019), pp. 1–9 (cit. on p. 18).

[25] Jose Antonio Urigüen and Begoña Garcia-Zapirain. «EEG artifact removal—state-of-the-art and guidelines». In: *Journal of Neural Engineering* 12.3 (Apr. 2015), p. 031001. DOI: 10.1088/1741-2560/12/3/031001. URL: https://dx.doi.org/10.1088/1741-2560/12/3/031001 (cit. on pp. 18, 19).

[26] Shkurta Gashi, Elena Di Lascio, Bianca Stancu, Vedant Das Swain, Varun Mishra, Martin Gjoreski, and Silvia Santini. «Detection of Artifacts in Ambulatory Electrodermal Activity Data». In: *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4.2 (June 2020). DOI: 10.1145/3397316. URL: https://doi.org/10.1145/3397316 (cit. on p. 20).

[27] Gaang Lee, Byungjoo Choi, Houtan Jebelli, changbum ahn changbum, and SangHyun Lee. «Noise Reference Signal-based Denoising Method for EDA Collected by a Multimodal Wearable Biosensor from Fields». In: *Journal of Computing in Civil Engineering* (July 2020). DOI: 10.1061/(ASCE)CP.1943-5487.0000927 (cit. on p. 21).

[28] Mohamed Elgendi. «On the Analysis of Fingertip Photoplethysmogram Signals». In: *Current cardiology reviews* 8 (Feb. 2012), pp. 14–25. DOI: 10.2174/157340312801215782 (cit. on p. 22).

[29] Zixuan Liu, Jingjing Kong, Menglong Qu, Guangxin Zhao, and Cheng Zhang. «Progress in Data Acquisition of Wearable Sensors». In: *Biosensors* 12.10 (2022). ISSN: 2079-6374. URL: https://www.mdpi.com/2079-6374/12/10/889 (cit. on pp. 24, 25).

[30] Vijayasankar Anumala and Rajesh Pullakura. «Removal of Ocular Artifacts in Single Channel EEG by EMD, EEMD and CEEMD Methods Inspired by Wavelet Thresholding». In: *International Journal of Image, Graphics and Signal Processing* 10 (May 2018), pp. 45–55. DOI: 10.5815/ijigsp.2018.05.05 (cit. on p. 26).

[31] David Pollreisz and Nima TaheriNejad. «Detection and Removal of Motion Artifacts in PPG Signals». In: *Mobile Networks and Applications* 27.2 (Apr. 2022), pp. 728–738 (cit. on p. 27).

[32] Miao Zhang and Guo Wei. «An integrated EMD adaptive threshold denoising method for reduction of noise in ECG». en. In: *PLoS One* 15.7 (July 2020), e0235330 (cit. on pp. 27, 28).

[33] Weixuan Chen, Natasha Jaques, Sara Taylor, Akane Sano, Szymon Fedor, and Rosalind W Picard. «Wavelet-based motion artifact removal for electrodermal activity». en. In: *Annu Int Conf IEEE Eng Med Biol Soc* 2015 (2015), pp. 6223–6226 (cit. on p. 28).

[34] S Abbaspour, H Gholamhosseini, and M Linden. «Evaluation of Wavelet Based Methods in Removing Motion Artifact from ECG Signal». In: *16th Nordic-Baltic Conference on Biomedical Engineering*. Springer International Publishing, 2015, pp. 1–4 (cit. on pp. 28, 29).

[35] Puneet Mishra and Sunil Kumar Singla. «Artifact removal from biosignal using fixed point ICA algorithm for pre-processing in biometric recognition». In: *Meas. Sci. Rev.* 13.1 (Jan. 2013) (cit. on p. 30).

[36] Xiang An, Yanzhong Liu, Yixin Zhao, Sichao Lu, George K. Stylios, and Qiang Liu. «Adaptive Motion Artifact Reduction in Wearable ECG Measurements Using Impedance Pneumography Signal». In: *Sensors* 22.15 (2022). ISSN: 1424-8220. DOI: 10.3390/s22155493. URL: https://www.mdpi.com/1424-8220/22/15/5493 (cit. on p. 32).

[37] Machado de Brito. *Applying Machine Learning Algorithms to Medical Knowledge*. Minho: Journal Not Specified, 2019 (cit. on pp. 34, 35).

[38] Shruti Gedam and Sanchita Paul. «A Review on Mental Stress Detection Using Wearable Sensors and Machine Learning Techniques». In: *IEEE Access* 9 (2021), pp. 84045–84066. DOI: 10.1109/ACCESS.2021.3085502 (cit. on pp. 35–38, 63).

[39] F. Pedregosa et al. «Scikit-learn: Machine Learning in Python». In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830 (cit. on p. 42).

[40] Zhenya Zang, Dong Xiao, Quan Wang, Ziao Jiao, Zinuo Li, Yu Chen, and David Day-Uei Li. «Hardware inspired neural network for efficient time-resolved biomedical imaging». en. In: *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* 2022 (July 2022), pp. 1883–1886 (cit. on p. 63).

[41] Philip Schmidt, Attila Reiss, Robert Duerichen, and Kristof Van Laerhoven. «Wearable affect and stress recognition: A review». In: (Nov. 2018). arXiv: 1811.08854 [cs.HC] (cit. on p. 63).

[42] Angie M Ceniza and Anthonette D Cantara. «Stress Sensor Prototype: Determining the Stress Level in using a Computer through Validated Self-Made Heart Rate (HR) and Galvanic Skin Response (GSR) Sensors and Fuzzy Logic Algorithm». In: *INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT)* 05.03 (2016) (cit. on p. 63).

[43] K H Kim, S W Bang, and S R Kim. «Emotion recognition system using short-term monitoring of physiological signals». In: *Medical and Biological Engineering and Computing* 42.3 (May 2004), pp. 419–427 (cit. on p. 64).

[44] J.A. Healey and R.W. Picard. «Detecting stress during real-world driving tasks using physiological sensors». In: *IEEE Transactions on Intelligent Transportation Systems* 6.2 (2005), pp. 156–166. DOI: 10.1109/TITS.2005.848368 (cit. on p. 64).

[45] Dimitris Giakoumis, Anastasios Drosou, Pietro Cipresso, Dimitrios Tzovaras, George Hassapis, Andrea Gaggioli, and Giuseppe Riva. «Using activity-related behavioural features towards more effective automatic stress detection». en. In: *PLoS One* 7.9 (Sept. 2012), e43571 (cit. on p. 64).

[46] Cornelia Setz, Bert Arnrich, Johannes Schumm, Roberto La Marca, Gerhard Tröster, and Ulrike Ehlert. «Discriminating stress from cognitive load using a wearable EDA device». en. In: *IEEE Trans. Inf. Technol. Biomed.* 14.2 (Mar. 2010), pp. 410–417 (cit. on p. 64).

[47] Rateb Katmah, Fares Al-Shargie, Usman Tariq, Fabio Babiloni, Fadwa Al-Mughairbi, and Hasan Al-Nashash. «A review on mental stress assessment methods using EEG signals». en. In: *Sensors (Basel)* 21.15 (July 2021), p. 5043 (cit. on p. 64).

[48] Likun Xia, Aamir Malik, and Ahmad Subhani. «A physiological signal-based method for early mental-stress detection». In: *Biomedical Signal Processing and Control* 46 (Sept. 2018), pp. 18–32. DOI: 10.1016/j.bspc.2018.06.004 (cit. on p. 64).

[49] Amandeep Cheema and Mandeep Singh. «An application of phonocardiography signals for psychological stress detection using non-linear entropy based features in empirical mode decomposition domain». In: *Applied Soft Computing* 77 (Apr. 2019). DOI: 10.1016/j.asoc.2019.01.006 (cit. on p. 64).