Universidade de Aveiro
2021

CAROLINA FERNANDES ALVES

CLASSIFICAÇÃO FISIOLÓGICA DE EMOÇÕES: DO MODELO DISCRETO AO DIMENSIONAL

DISCRETE TO DIMENSIONAL PHYSIOLOGICAL EMOTION CLASSIFICATION

Universidade de Aveiro
2021

CAROLINA FERNANDES ALVES

CLASSIFICAÇÃO FISIOLÓGICA DE EMOÇÕES: DO MODELO DISCRETO AO DIMENSIONAL

DISCRETE TO DIMENSIONAL PHYSIOLOGICAL EMOTION CLASSIFICATION

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Engenharia Biomédica, realizada sob a orientação científica da Doutora Susana Manuela Martinho dos Santos Baía Brás, Investigadora no Instituto de Engenharia Electrónica e Informática de Aveiro (IEETA), Departamento de Electrónica, Telecomunicações e Informática (DETI) da Universidade de Aveiro, e da Doutora Sónia Cristina Alexandre Gouveia, Investigadora Auxiliar no Instituto de Engenharia Electrónica e Informática de Aveiro (IEETA), Departamento de Eletrónica, Telecomunicações e Informática (DETI) da Universidade de Aveiro.

Dedico este trabalho à minha família, amigos e a todos que fizeram parte deste meu percurso na Universidade de Aveiro.

“Happiness is only real when shared.” - Christopher McCandless

o júri

presidente

Professora Doutora Ana Luísa Monteiro Silva
Professora auxiliar do Departamento de Física da Universidade de Aveiro

vogais

Doutora Maria Luísa Morais e Sousa de Castro Guedes
Investigadora Doutorada CINTESIS, Faculdade de Medicina da Universidade do Porto (FMUP)

Doutora Susana Manuela Martinho dos Santos Baía Brás
Investigadora no Instituto de Engenharia Electrónica e Telemática de Aveiro (IEETA) da Universidade de Aveiro (orientadora)

agradecimentos

Este trabalho é o término do meu percurso na Universidade de Aveiro, no Mestrado Integrado em Engenharia Biomédica. Desta forma, não podia deixar de agradecer a todas as pessoas que me acompanharam e que tornaram tudo isto possível.

Mãe e Pai, um gigante obrigada. Obrigada por me apoiarem, por me darem força nos momentos mais difíceis e por me compreenderem melhor do que ninguém, às vezes até melhor do que eu própria! Obrigada por me terem proporcionado uma vida muito feliz e por me ensinarem a viver a vida da melhor forma. Admiro-vos muito, mais do que podem imaginar, e espero um dia ser como vocês. Quico, meu 'charabaneco', apesar de apareceres tarde, foste a melhor coisa que me aconteceu na vida. Foi contigo que aprendi a amar e és agora a minha maior fonte de felicidade! Vou para sempre encher-te de beijinhos e abracinhos, mesmo que reclames todas as vezes. Vais ser sempre o meu bebé! Avós, tios e primos, que nos últimos meses perguntaram todas as semanas se já tinha entregue a tese, finalmente acabei! Obrigada por terem sempre uma palavra amiga e de força, foi muito importante para mim.

Amiguinhos, de cá (Figueira da Foz) e de lá (Aveiro), vocês são incríveis. É verdade que nasci com uma família maravilhosa, mas também consegui arranjar outra igualmente boa! Às minhas meninas de Aveiro, Catarina, Andreia, Simone, Ana e Cátia, muito obrigada por me terem aturado e apoiado diariamente nos últimos anos! Foram várias as sessõs de estudo, os almoços, os lanchinhos e as gargalhadas nos intervalos que partilhámos. Conhecer-vos foi a cereja no topo do bolo de estar em Aveiro e fico muito feliz por ter partilhado estes anos convosco! À grupeta da Figueira, Ana, Bia, Mariana, Diana, Pagano, Sérgio e Diogo, que continuemos a nos encontrar aos fins-de-semana para cafés e jantares! Que existam mais acampamentos com massas por cozer e dunas para subir, e que apesar de cada um de nós estar num lugar diferente, que consigamos arranjar sempre um tempinho uns para os outros. Aprendi muito com cada um de vocês e sei que posso contar sempre convosco. Laura, minha amiga de cá e de lá, foi contigo que partilhei mais momentos nestes últimos 5 anos. Temos muitas histórias e espero que ainda arranjemos muitas mais porque é isso que fica na memória! Obrigada por tudo.

Professora Susana e Professora Sónia, muito obrigada. O último ano foi um desafio e não podia ter contado com um melhor apoio. Por todas as vezes em que ficava bloqueada, em que tinha dúvidas e que não sabia por onde ir, muito obrigada pela vossa disponibilidade em me ajudar e esclarecer. Foram incansáveis e espero ter feito juz ao que era esperado deste trabalho.

Deixo ainda um agradecimento especial à Filipa Barros pela disponibilidade e ajuda dada na área da Psicologia e estimulação emocional, e a todos os participantes que participatam no estudo deste trabalho, permitindo assim a sua realização.

palavras-chave

Emoção; Elicitação emocional; Sinais fisiológicos; Classificação; Modelo Discreto; Modelo Dimensional

resumo

As emoções desempenham um papel muito importante na vida humana. A maneira como nós comunicamos e interagimos uns com os outros, as nossas ações e pensamentos, são todos influenciados por elas, seja de uma forma positiva ou negativa. Infelizmente, existe uma variedade de doenças mentais, como a ansiedade e depressão, que são caracterizadas por uma prevalência de emoções negativas e nas quais as pessoas tendem a ter uma maior dificuldade em compreender o seu estado emocional. Consequentemente, é muito importante que cada um de nós seja capaz de identificar as nossas emoções, de forma a garantir que as conseguimos controlar e que o efeito contrário não ocorra. Os sistemas de reconhecimento de emoções podem ser uma das soluções para ajudar as pessoas a identificar as suas emoções, levando a uma melhoria do seu bem-estar e saúde.

Os estudos nesta área têm explorado diferentes tópicos que vão desde o tipo de sinais e características, ao método de seleção de características e de classificação emocional. Além disso, também começaram a divergir na abordagem para descrever as emoções, que pode ser discreta (e.g. alegria, medo) ou dimensional (e.g. nível de agradabilidade, ativação). Neste trabalho, as duas abordagens foram estudadas de forma a compreender o impacto da descrição emocional no processo de classificação e, assim, concluir sobre a abordagem mais adequada para identificar as emoções. Para tal, foi criada uma base de dados constituída pelos sinais fisiológicos: eletrocardiograma, atividade eletrodérmica e eletromiograma dos músculos medial frontal e trapézio. A análise exploratória destes dados permitiu descrever as emoções do ponto de vista da resposta fisiológica. O eletrocardiograma e a atividade eletrodérmica apresentaram-se como sendo os sinais que melhor discriminam a atividade emocional (têm o maior número de características que distinguem os estados emocionais). Numa análise multivariável, verificou-se que a informação do eletromiograma também era uma fonte discriminatória, uma vez que as suas características eram sistematicamente selecionadas pelo classificador. A abordagem inicialmente estudada assentou sobre o modelo discreto de emoções, contudo a classificação errada de algumas observações levou a ponderar a hipótese de testar um modelo dimensional (agradabilidade/ativação). Este modelo revelou-se mais robusto que o anterior, o que levou a concluir que se adapta melhor quer à resposta da emoção, quer à resposta individual de cada pessoa ao estímulo. Comprovando assim a sua melhor descrição da emoção.

keywords

Emotion; Emotion elicitation; Physiological signals; Classification; Discrete Model; Dimensional Model

abstract

Emotions play a very important role in human life. The way we communicate and interact with others, our actions, thoughts, are all influenced by them, whether in a positive or negative way. Unfortunately, there is a variety of mental diseases, like anxiety and depression, that are characterized by a prevalence of negative emotions, and in which people tend to have a higher difficulty in understanding their emotional state. Consequently, the importance of each one of us being able to identify their emotional state is crucial to guarantee a healthy control over emotions. The emotion recognition systems can be one of the solutions to help people identify and interpret their emotions, hence increasing their well-being and health.

Studies in this area have explored different topics ranging from the type of signals and features to the method of feature selection and emotional classification. Furthermore, they also began to diverge in the approach of describing emotions, which can be discrete or dimensional. In this work, the two approaches were studied to understand the impact of the emotional description on the classification process and to conclude on the most adequate approach to identify emotions. To this end, it was created a database of the physiological signals: electrocardiogram, electrodermal activity and electromyogram of the medial frontalis and trapezius muscles. An exploratory analysis was performed with these data revealing that the electrocardiogram and electrodermal activity represent the most informative in emotion discrimination. Nevertheless, in a multivariable approach, the features from electromyogram reveal to be useful on emotion classification. The approach initially studied was based on a discrete model of emotions, however, misclassification of some observations led to considering the hypothesis of testing a dimensional model (valence/arousal). This model proved to be more robust than the previous one, which led to the conclusion that it is better adapted to both emotional response and the individual response to the stimulus, confirming its best description of the emotion.

Contents

List of Figures

List of Tables

List of Acronyms

ADF	Augmented Dickey-Fuller
ANS	Autonomic Nervous System
CNN	Convolutional Neural Network
CWT	Continuous Wavelet Transform
DE	Differential Entropy
DWT	Discrete Wavelet Transform
ECG	Electrocardiogram
EDA	Electrodermal Activty
EEG	Electroencephalogram
EMD	Empirical Mode Decomposition
EMG	Electromyogram
EMOTE	Emotion Multimodal Database
FFT	Fast Fourier Transform
GSR	Galvanics Skin Response
HF	High Frequency
HOS	High Order Statistics
HRV	Heart Rate Variability
IMF	Intrinsic Mode Function
KNN	K-Nearest Neighbor
KS	Kolmogorov-Smirnov
LF	Low Frequency
MLP	Multilayer Perceptron
PSD	Power Spectral Density
RM ANOVA	Repeated Measures ANOVA
RSP	Respiration
SAE	Sparse auto-encoder
SCR	Skin Conductance Response
STFT	Short Time Fourier Transform
SVM	Support Vector Machine
UN	United Nations
VMD	Variational Mode Decomposition
WHO	World Health Organization

1. Introduction

1.1 Context

Over the last year and looking at what's happening in the world, with the coronavirus disease 2019 pandemic, the social distancing, the new health care, the fact that sometimes we see each other thinking of what could happen to us, to our family, friends, country, are causing a lot of fear, stress, and anxiety in adults and children. These types of emotions are mostly negative and are affecting our way of life, in the simplest ways [1].

Emotions play a very important role in human life, affecting both human physiological and psychological state [2]. Consequently, our actions and thoughts are influenced by them, whether positively or negatively. Given their role, each one of us should be able to identify them so that their contribution is controlled according to the situations in which we find ourselves.

The field of psychology has been studying emotions for long and several definitions have been proposed. In general, emotions are considered as responses to certain stimuli that will induce corporal changes, which can be physiological or behavioral. According to Plutchik (2001) emotions can also be considered as chains of events, in which cognition is usually found at the beginning of the sequence [3]. Thus, emotional experience encompasses cognitive, behavioral, and physiological components.

If emotions influence our perception, rational thinking, and decision-making, then just as those around us must be able to identify our emotional state, affective computers should also be able to recognize them and respond intelligently according to the detected emotion. Affective computing, which encompasses computer science, psychology, and cognitive science, is the area that has been studying how to integrate emotions in the computer-user interaction [4].

More than a century ago, Walter Cannon (1915) studied the physiology of emotion, trying to understand the possible relationship between the autonomic nervous system and emotions. Nowadays, there is still no consensus if some kinds of emotions have unequivocal and invariant autonomic responses, however, it is known that the physiological changes of emotions are related to the nervous system [5]. In particular, emotions are related to the autonomic nervous system (ANS) that is responsible for releasing neurotransmitters that will reach the target organ and induce the physiological response. It's dependent on the central nervous system and its activation is involuntary, which means that it cannot be easily controlled [2], [6].

The corporal changes in an emotional experience can be distinguished in human physical signals and physiological signals. Physical signals like facial expression, voice, gesture, and posture can be easily detected, however, they can be controlled and hidden by the person, which can lead to emotional misinterpretation. On the other hand, physiological signals like body temperature, electroencephalogram (EEG), electrocardiogram (ECG), electromyogram (EMG), electrodermal activity (EDA), are dependent on the autonomic nervous system and, therefore, are involuntary and not easily controlled by the person. Consequently, the evaluation of physiological signals can be more reliable to detect emotions rather than the physical responses [2], [6].

In the late 19th and early 20th centuries, the appearance of electronic devices able to detect most of the physiological signals gained a significant role in the healthcare field. These smart wearables by monitoring signals easily, in real-time, and for a long duration have allowed the development of new diagnostic and therapeutic techniques [7], [8]. In particular, for emotion recognition, where it is important to assess the subjects' emotion in their usual context, these types of devices, that are less

intrusive, have been facilitating the acquisition of emotional responses in a real context, ensuring the ecological validity of the emotion evaluation [3]. In addition, studies in this area were revolutionized, and different approaches have been developed using different types of physiological signals and creating different systems with an emphasis on feature extraction, feature selection, and classifier designs [9].

In the evolution of the concept of emotion, two distinct models have emerged that characterize emotions, the discrete and the dimensional. In the discrete model, emotions are defined by a list, and it is considered that there is a limited number of them, whereas, in the dimensional model, emotions are defined by dimensions in which their intensity is considered. Likewise, in the studies done so far in emotion recognition, these two approaches have been explored and in general, the most promising results involve emotional characterization in a dimensional space [2].

The investigation of emotion recognition systems is a growing area, and its success will benefit a variety of areas such as psychology, assisting in the diagnosis and treatment of psychological illnesses, as medicine, helping patients with communication and socialization difficulties, and as in education, through the adaptation of teaching methodologies. According to the World Health Organization (WHO), depression affects more than 261 million people worldwide and autism spectrum disorder affects 160 children worldwide [10], [11]. These numbers refer only to two of the many diseases and disorders that may benefit from the evolution of emotional identification, thus justifying the importance of studies in this area. The recognition of emotions is one of the ways to improve the well-being and health of humans and the development of these systems can be one of the approaches to fulfill one of the sustainable goals of the UN, which is directly related to 'good health and well-being' [12].

1.2 Work Objective

The recognition of emotions encompasses several aspects ranging from the choice/acquisition of the physiological signals used, to the implementation of the systems. First of all, it is crucial to use signals that represent real situations, that is, physiological signals of people experiencing real emotions. Creating a new database allows controlling not only a variety of parameters that influence the elicitation of emotions but also the signals acquisition, which leads to the first objective of the work:

1. Development of an experimental protocol for the acquisition of physiological signals with emotion elicitation.

Understanding how emotions are manifested in the human body, through physiological signals, optimize their acquisition only to those who present more significant changes to emotions. In addition, determining which features have the highest amount of relevant information allows optimizing the process of extracting them and, thus, reduce the computational cost. To accomplish this, the second objective is:

2. Development of an exploratory data analysis framework to understand which signals and features best describe emotions.

Regarding the implementation of emotion recognition systems, several techniques and algorithms can be used to identify emotions from features extracted from signals. However, the challenge is to find the best combination that maximizes the ability of the system to correctly identify the emotion, so this work aims the:

3. Construction and optimization of an emotional model using machine learning techniques.

Finally, the growing interest in the dimensional models to describe emotions, led studies to start using dimensions in emotion recognition systems. Thus, in this work is intended to:

4. Understand the impact of each model in the emotional classification.

1.3 Work Outline

To achieve the objectives proposed in the previous section, the work presented is divided into 6 chapters. In **Chapter 1**, the theme of the study is contextualized and the objectives to achieve are presented. Next, in **Chapter 2**, it is described the evolution of the concept of emotion and are presented the two models for describing emotions and a summary of the physiological signals and methods used in studies in the field of emotion recognition systems. **Chapter 3**, fulfills the first objective of the work, where the data collection protocol is described and the database of the physiological signals obtained is characterized. The last three objectives of the work are explored in Chapter 4 and Chapter 5. In **Chapter 4**, the methodology and implementation used to build the emotional model are described, and in **Chapter 5** are presented the results from the exploratory data analysis and classification by the emotional model. **Chapter 6**, summarizes the results and conclusions obtained previously, contextualization of applications of the proposed emotional model, and improvements for future work.

2. Background

2.1 Definition of Emotion

The beginning of the evolution of the emotion concept was in 1872 with the *Expression of the Emotions in Man and Animals* by Charles Darwin. Darwin defined emotions as feelings caused by external states and not caused by bodily states, and this was a starting point to many other theorists that later preferred to adopt behavioral and physiological definitions of emotion [13]. Besides that, he identified similarities between emotional responses in human and non-human animals, which later led to emotions being seen as responses to individual survival [14].

In 1884 William James suggested that the experience of certain emotions is the result of particular changes in behavioral and physiological responses [15]. Some of these responses are easily identifiable by humans, such as the variation of heart rate, sweating of the hands, relaxation or tension in the muscles, smiling, etc., and James defined them as “coarser emotions: fear, rage, grief, love, in which everyone recognizes a strong organic reverberation”. However, he also mentioned that there are more subtle emotional responses, which human beings cannot easily identify [6].

Supporting James’ theory, in a biological view, Lange defined emotion explicitly as a cardiovascular response “We owe … the emotional side of our mental life, our joys and sorrows, our happy and unhappy hours, to our vasomotor system.” [6] and Dewey stated “Emotion is a mode of behavior, which is purposive, or has an intellectual content. […] Certain movements, formerly useful, are reduced to tendencies to action or to attitudes, and when instinctively aroused into action, serve as means for realizing ends.” [14].

Afterward, Walter Cannon declared that physiological responses were the consequence, not the cause, of emotional processes, and that physiological responses were too undifferentiated to account for the variety of distinct emotional feeling states [15].

Some theorists believe that emotions are processes and actions that are used intentionally by the body in cases of survival. In this field stands out the concept of motivation and James had concluded some statements about the difference between emotion and motivation. On the one hand, motivation, which included instincts, comprises mandatory actions triggered by stimuli from the environment, defined by limbic circuits, and can go beyond bodily changes including relationships with objects and other animals. On the other hand, even though any stimulus that induces an instinct also induces an emotion, the emotional response ends in the subject's body [6].

Despite the many definitions and theories around this concept, emotions are considered to be related to the nervous system and that in emotional situations the body reacts, undergoing some changes, which may be behavioral, physiological, or even conscious experience.

2.1.1 Discrete and Dimensional Models

Besides the different definitions of emotion, there is also a conflict in its nature. The two theories that try to explain it are the discrete and dimensional models.

The discrete model assumes that it is only possible to experience one emotion at a time, there is a limited number of them, and that basic emotions were formed throughout human evolution and are born with us [16].

Over the years, the list of discrete emotions has changed according to the theorist who defined them. Descartes defined a list of emotions including wonder, love, joy, desire, hate, and sadness, however, later, Ekman defined six basic emotions, happy, sad, anger, fear, surprise, and disgust. In 1980, Plutchik proposed a wheel of discrete emotions (Figure 1) with eight basic emotions in the center (joy, trust, fear, surprise, sadness, disgust, anger, and anticipation), represented with a darker tone, thus the most intense ones; and in the flower blooms, emotions are illustrated with a lighter tone meaning they are weaker [2], [6].

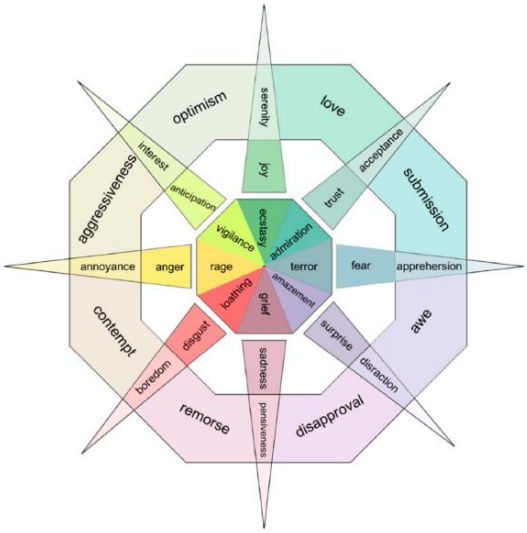

Figure 1 *- Plutchik's wheel of emotions [2].*

Overall, the discrete model describes emotions qualitatively. However, for more complex emotions or even for situations with mixed emotions this description is not enough. To overcome this difficulty the dimensional model describes them quantitatively. In the dimensional model, emotions are seen as combinations of fundamental dimensions and cognitive processes [16]. Some of these dimensions are valence, arousal, which in some theories is divided into tension arousal and energy arousal, and approach-avoidance [6]. In the context of emotions, arousal refers to the strength of the emotional experience, therefore, related to excitement-calmness in energy arousal, and tension-relaxation in tension arousal. If no discrimination is made, arousal ranges from passive to active [2]. Valence refers to the degree of pleasure (positive) and displeasure (negative) associated with emotion, and approach-avoidance is related to the motivation to approach or avoid certain stimuli [16], [17].

Although there is no consensus on the number of dimensions to define emotions, all theories agree that their number is limited and the most common dimensions in this model are valence and arousal [18]. This model also encompasses theories that believe that some types of emotions are the result of construction processes such as perception, attention, and memory, which combined and under the influence of social and linguistic factors, give meaning to emotions [16].

Recently, a combination of the two models was proposed, by assigning specific dimensions to each discrete emotion, and graphically representing the discrete emotions in the corresponding dimensional space (Figure 2) [16]. For example, sadness has negative valence and low arousal (passive) whereas, anger has negative valence and high arousal (active) [2].

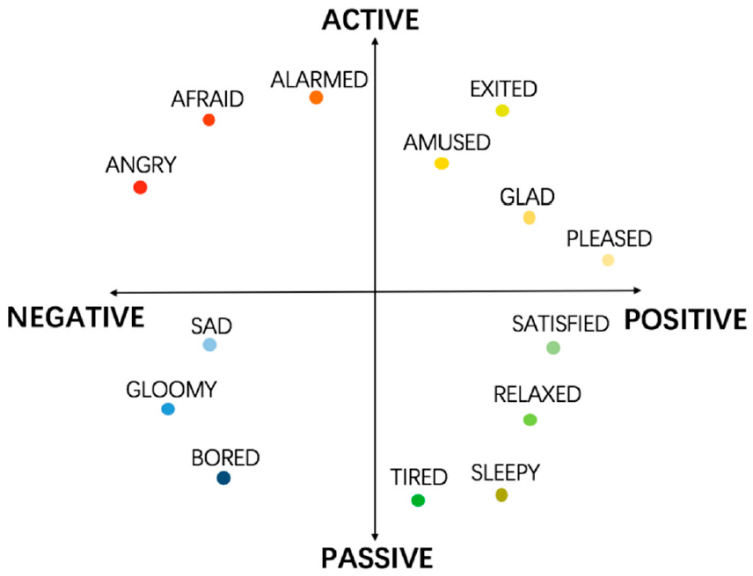

Figure 2 *- 2D emotion space model [2].*

2.2 Emotion Recognition

In our daily life, in a social environment, identifying the emotions of those around us doesn't require great steps and methods, it is simpler, we only use our capacity of perception. However, in emotion recognition, which aims to automatically classify the human's emotional state [19], there is a set of steps to perform so that it is possible from a signal to identify the subject's emotional state. Beginning with data collection until emotion identification, relevant features are extracted, data is selected if the dimension is large, and classifiers are implemented [20]. The classification, being the final step of emotion recognition, uses the selected features and the correspondent emotion to learn and, later, predict emotions of other feature data.

Initially, the studies were based on the discrete model and the emotions mostly used were happiness, sadness, fear and anger. However, with the evolution of dimensional models, studies also started to investigate this approach. The main difference, when compared to discrete, is that there is a focus on the quantification of emotions, and for that reason are used dimensions such as valence, arousal, liking, dominance, and approach-avoidance. According to the studies done so far, the dimensions that best define emotions are valence and arousal, hence they are the most used in these studies. The other dimensions only encompass small variations in emotion hence they are less significant.

In both approaches, it is common to use classifiers to identify emotions, however, using different prediction approaches. In the discrete model, the classifier predicts a discrete emotion like sadness or joy, whereas, in the dimensional model, it predicts a label related to the dimension, like negative valence and low arousal (characteristics of e.g. sadness).

In addition to the divergence in the approach of describing emotions, studies can adopt two different strategies: subject dependent or subject independent. On subject dependent, it is assumed that each person feels an emotion differently, so that the classification method should be adaptative in order to find individual dependencies. On subject independent, it is assumed that the main emotional process is similar between subjects, so the algorithm is general and tries to find the key characteristics that describe the emotional process [2].

2.2.1 Physiological Signals

The performance of the emotion recognition systems is highly correlated to the ability of the classifiers to correctly identify the emotion, and it depends on several factors that go from the type of data to the chosen classification method. Consequently, it is crucial to understand the data and to define the processing strategy that may better describe the emotional physiological process.

Different types of data can be collected to proceed to the identification of the subject's emotions. One is related to emotions self-assessment by filling questionnaires and the other to signals of the human body [21]. In the field of signals, two types stand out, one referring to physical signals and the other to physiological signals.

Some examples of physical signals are facial expressions, voice, posture, gestures, among others. Although they are easily collected and subsequently identified, stand out the disadvantage of being easily controlled by the human being, which leads to misrepresentation of the true emotional state [2]. Despite being the first signals used in this area, due to its disadvantage, they started to be used as ground truth and confirm the emotion identified by the computational method. Self-assessment questionnaires also serve as guidance, since they share the disadvantage of physical signals and other two that are related to the difficulty of some people to identify their emotional state and to the social influence that may induce the subject to unconsciously answer to what the society is expected to answer.

On the other hand, physiological signals, that are a result of the (involuntary) autonomic nervous system, are difficult to control and for that reason are more reliable. Although in the past, the acquisition of these signals could be intrusive due to certain sensors and electrodes, with the great evolution of wearable technologies in the 90s, the intrusiveness was diminished and nowadays the acquisition of physiological signals is easier and more comfortable to the subject [22]. The most used physiological signals are electroencephalogram (EEG), electrocardiogram (ECG), electrodermal activity (EDA), electromyogram (EMG), and respiration (RSP), where the target position of the sensors is represented in Figure 3 [21]. The following sections describe these physiological signals and some of the features and methods that are used in emotion recognition studies. The tables that summarize the emotion recognition studies' information are available online[1].

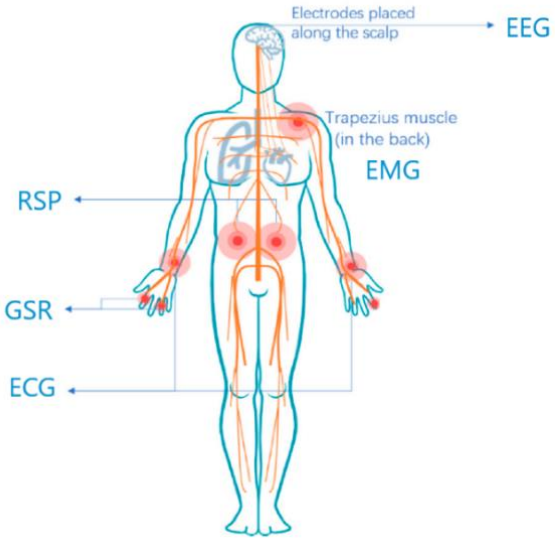

Figure 3 *- Physiological signals: electroencephalogram (EEG), electromyogram (EMG), respiration (RSP), galvanic skin response (GSR), and electrocardiogram (ECG). (Adapted from [2])*

[1] https://github.com/carolinafalves/Repository_CarolinaAlves.git

Electroencephalogram (EEG)

Electroencephalography is an electrophysiological noninvasive technique to record the electrical activity of the human brain. The EEG signal is composed of five frequency ranges namely: alpha, beta, delta, gamma, and theta, that are highly related to human emotional states [21], [23]. The use of EEG to identify emotions began in 1924 with the psychiatrist Hans Berger [21], and since then, several studies have been using different features and methods to obtain them.

According to [2], the most common features extracted from the EEG, in emotional recognition, are the power spectral density (PSD) [24], [25], the differential entropy (DE) [23], and intrinsic mode functions (IMF) [26], [27]. Briefly, the PSD is usually computed using the Short Time (STFT) or Fast (FFT) Fourier Transform, the DE is equivalent to the logarithmic power spectral density for a fixed length EEG sequence, and the IMF's may be computed using the Empirical (EMD) or Variational (VMD) mode decomposition.

There are a variety of methods to process and analyze the EEG signals that can be implemented according to the purpose of the study. To evaluate the level of valence and arousal, a Fast Fourier Transform or latency test can be used. To evaluate a specific emotion statistical methods or machine learning techniques can be considered [21]. Regarding the classifiers, most studies use in the machine learning approaches, the Support Vector Machine (SVM), and in deep learning, different types of neural networks.

Some of the advantages of using the EEG signals are related to its strong objectivity and to the easy acquisition using wearable EEG headset. However, it is also reported that is difficult to separate these signals from the background noise.

Electrocardiogram (ECG)

The electrocardiography is a noninvasive technique to record the electrical activity of the heart. The ECG signal is characterized by the presence of waves, namely: P, Q, R, S, T, and U, and each one of them is related to specific heart activity.

The QRS complex, the part of the ECG comprised by the Q, R, and S waves, defines the activation of the heart related to the human emotional state and has variant sensitivity to some specific emotions. For instance, it was reported that it is easier to recognize sadness than joy [21]. Some of the features related to this complex and used in emotion recognition are the R wave duration, QRS p-p amplitude, R wave amplitude, QRS wave area, and RR intervals [20], [28], [29].

The statistical features include energy, mean, standard deviation, maximum, minimum, kurtosis, skewness, and entropy. Part of studies use these features to analyze the heart rate variability (HRV), and also frequency domain features like the LF power, HF power, and the ratio of LF/HF [28]. The most common methods to obtain frequency domain features are the Continuous (CWT) and Discrete (DWT) Wavelet Transform, and the most used classifiers are the K-Nearest Neighbors (KNN), the SVM, and neural networks.

Most of the studies found using the ECG signal, relied on the discrete approach, except for [30], which explored the dimensional approach and used the coefficients obtained through the matching pursuit (MP) method as features, and deep learning methods to identify emotions.

Electrodermal activity (EDA)

The EDA is a measurement of the electrical parameters of the human skin. The emotional changes induce sweat reactions, mostly on hands, and this variation leads to the change of electrical resistance of the skin [21].

The EDA raw signal contains two types of activity: tonic and phasic. The tonic activity depends on the skin hydration level, dryness, and this is what leads to the changes in conductivity. The phasic activity is related to short-term peaks of the galvanic skin response (GSR) and is highly dependent on the sympathetic nervous system [21].

The most common features extracted are in the time and frequency domains, such as mean, median, standard deviation, average amplitude, average duration, maximum amplitude, and also in some studies are computed the wavelet coefficients [31]–[33]. The methods of obtaining them are similar to those mentioned above for the ECG, as they are the similar features of different signals. Regarding the classifiers, the review studies used the SVM, in the discrete approach, and the Multilayer Perceptron (MLP) and the Convolutional Neural Networks (CNN), in the dimensional approach.

The main advantages when using the EDA signal are that the data collection equipment is simpler and cheaper, the number of electrodes is reduced, the acquisition is easy, and that the signal is easily processed, not requiring much computational power.

Respiration (RSP)

The RSP signal allows obtaining the respiratory rate, which contains very useful information about emotional states [21]. According to [34], using the dimensional approach, high arousal is related to high respiration frequency and low arousal to low frequency.

Some of the features that can be extracted are the mean, the standard deviation, the maximum and minimum of the respiratory rate, and the PSD [35]. The reviewed literature, when using the features mentioned, was based on the discrete approach; while the study that based on the dimensional approach, used a deep learning method, the Sparse Auto-Encoder (SAE) to compute the features, and the logistic regression to classify the emotional states [34].

One of the major disadvantages of breathing is that although it depends on the emotional state it also depends on many other factors such as the movement of the human body, the person's fatigue, the temperature and humidity of the environment, among others, which leads to this signal be more sensitive and lead to misidentifications, hence less used [21].

Electromyogram (EMG)

The electromyography is a technique to record the electrical potential generated by muscle cells. The most frequent analysis is made to facial expressions because according to Ekman and Friese it may exist a dependency between simple emotions and facial expressions. The most often target facial muscles are *frontalis*, *corrugator supercilii*, *levator labii superioris*, *zygomaticus major*, and *orbicularis oculi*.

The features extracted from the EMG signals are those of higher-order statistics (HOS), including skewness and kurtosis, and statistical parameters such as normalized signals, the standard deviation of the raw signal, and mean of the absolute value of the first or the second difference [2], [36]. Many other studies were based on the wavelet coefficients obtained using different types of the Wavelet Transform (WT) method [37]–[39].

According to [21], the EMG is a good choice to detect strong emotions and is less suitable to identify small modifications of intensity within one emotion.

2.3 Applications

The identification of emotions has become very relevant in the most diverse areas such as robotics, marketing, entertainment, psychology, education, and medicine [21], [40].

Focusing on education, it is reported that students' emotions play an important role in the learning environment both in the classroom and e-learning [41], hence the need to create and adapt the methodologies that best suit the student to achieve higher academic performance. Parameters like excitement, disturbance, and the movement of eyes and head, may be used to infer meaningful information of the emotional state of the student. This analysis allows understanding, for example, whether the environment or the learning methodology is bothering the student, and if this happens, the teacher can adapt the methods or even the topics to make the whole learning method more appealing [41].

In the field of psychology, numerous mental disorders can be favored by emotion recognition systems including bipolar disorder, borderline personality disorder, generalized anxiety disorder, schizophrenia, and depression [42]. According to the World Health Organization (WHO) [10], depression affects more than 261 million people worldwide. It is characterized by persistent sadness and a lack of interest or pleasure in previously rewarding or enjoyable activities. In this sense, the identification of emotions may be able to assist in early diagnosis (e.g. if there is a prevalence of negative emotions), and in the treatment in which, over time, the patient is aware of his/her true mental state and can see the own evolution.

Another type of disorder where emotion recognition systems can help is the autism spectrum disorder (ASD), which translates into a developmental disability that can cause significant social, communication, and behavioral challenges [43]. According to the WHO [11], one in 160 children worldwide has an ASD. People diagnosed with ASD often have problems in communicating with others and consequently, they have some difficulty in understanding other people’s feelings or talking about their feelings. With the help of emotion recognition systems, people with ASD might have the chance to understand their and others’ emotional states, which is very important to work towards overcoming communication problems. For example, these systems can be incorporated into patients' ongoing therapies, or even into their daily lives.

Incorporating emotion recognition systems in classrooms, therapies, or even in people’s daily lives, can improve people’s emotional health and well-being. These systems can incorporate the solution to fulfill one of the sustainable goals of the United Nations (UN), which aims to ensure a healthy life and promote the well-being of all people ('good health and well-being') [12].

3. EMOTE data collection

The success of the emotion recognition models relies on the ability to identify the user's emotional state with good performance [44]. In addition to choosing the proper approach in the construction of the emotional model, it is of utmost importance to gather a large set of data, with high-quality signals, so that the model has sufficient and representative data for the development and validation processes [44], [45]. In this work, a dedicated experimental protocol was designed which lead to the creation of a multimodal emotion database of physiological signals using movies as a stimulus.

This chapter presents the different components of the experimental protocol designed to address the need to create a database of physiological signals for distinct emotional states. Section 3.1 presents the case study, including the objective and some factors that influenced the protocol's design. Section 3.2 presents the step of recruitment of participants and the inclusion criteria, and Section 3.3 describes the stimuli and questionnaires, including the main characteristics of the stimulation method and the description of the used questionnaires. Section 3.4 presents the setup, with all the equipment used, and Section 3.5 details the experimental protocol, including the procedure before, during, and after data collection. Finally, Section 3.6 presents the multimodal emotion database, describing the participants' characteristics, an evaluation of the questionnaires' responses, and global considerations regarding the acquired physiological signals.

The proposed protocol was approved by the *Conselho de Ética e Deontologia da Universidade de Aveiro* and by the delegate for the data protection affairs, with evaluation approval 12-CED/2020.

3.1 Case Study

The main goal of the study was to record physiological signals while the participant was induced to experience different emotional states by the visualization of movie excerpts. To guarantee a high-quality database, with which it is possible to carry out several analyses, the protocol's design was conceptualized to maximize the entire process by extracting as much information as possible and answering three relevant questions:

- What is the physiological description of an emotion?

Our physiological response varies with the emotional state [2]. Certain emotions have more impact on some physiological signals than others. Understanding which physiological signals are associated with each emotion or each dimension is essential to adapt the acquisition signal to the intended goal.

- How is the transition between emotions?

Throughout the day, the user's emotional state may vary [46]. Understand how the emotional transitions behave in physiological signals may have diverse applications, such as the early identification of emotions.

- Is the emotional response similar between days?

External factors can lead to different physiological responses to the same emotion on distinct days [46]. For the computational models to be able to recognize emotions on different days, it is necessary to understand how the emotional response fluctuates in physiological signals.

To accomplish these questions, the data acquisition of physiological signals was planned to be collected in two sessions, with a minimum time separation of one week. The movie excerpts used were validated in a previous systematic study, where each excerpt was rated according to its

intensity. The participant watched three sequential movie types, effective to induce three different emotions: fear (F), happiness (H), and neutral (N). The definition of a neutral condition is essential to describe the emotional “zero” condition, while fear and happiness are associated with negative and positive emotions, respectively.

In each session, the ECG, the EDA, and the EMG of the frontalis (EMG MF) and trapezius (EMG TR) muscles of the participant were continuously recorded. The selection of the physiological signals to acquire took into consideration the literature results concerning the ability of the signals (1) to vary according to an emotional state, (2) being used in a classification scheme exhibiting high performance, and (3) to be easily recorded to avoid an uncomfortable acquisition experience to the participant. Concerning the EMG, several options could have been considered. For example, the EMG acquired on a zygomatic muscle, was initially considered to be acquired. However, it was left out of the protocol due to the mandatory use of a facial mask (following the COVID-19 restrictions) and the possibility of conflict with the facial sensors just below the cheekbone of the participant.

The ecological validity implies collecting data in the usual or standard participant environment. However, there is not still a clear description of the correct procedure to evaluate emotional state. So, in the present protocol, lab conditions were used. To guarantee a more comfortable and real reaction of the participants, nonintrusive sensors were selected. An emotional response is directly related to sensorial input, so that all sensorial inputs should be elicited to maximize the emotional response [47]. Since this is not feasible in an experimental setup, in this study a visual and auditory sensorial input was elicited, by movie presentation.

3.2 Recruitment of participants and inclusion criteria

The recruitment of participants was carried out via invitation and divulgation of the experiment by email and personally. A given participant was included in the study if fulfilling the following criteria:

- Age between 18-35 years old;
- Normal or corrected-to-normal vision;
- No psychological or psychiatric diagnosed disorder;
- No physical condition that could impact the experiment (e.g. cardiac arrhythmia);
- No medication that could impact the experiment or that may be indicative of a psychological or psychiatric disorder (e.g. anti-depressives).

To access this information, the participants were asked to answer a sociodemographic questionnaire (described in Section 3.3), and according to their responses, the participants' selection was performed.

The participants included in the study were contacted and again informed of their voluntary participation and of the possibility to quit the experiment at any time. Participants were also asked to avoid caffeinated beverages or experience intense physical exercise one hour before the experiment, thus avoiding some external factors that could influence the physiological responses.

3.3 Stimuli and Questionnaires

Movie excerpts from horror, comedy, and documentary movies were used to elicit fear (F), happiness (H), and the neutral (N) state, respectively. It is known that each participant has its variability of physiological response [48] and so every experimental session began with the baseline assessment while the participant was presented with a neutral documentary of 5 minutes duration. In the first session (Figure 4), each emotion was elicited with an emotional set of 3 or more movie excerpts (average duration of 3 min each), totalizing 10 min of stimulation. The movie excerpts within each emotional set were ordered ascendingly according to their emotional intensity so that the emotional experience was increasing over time.

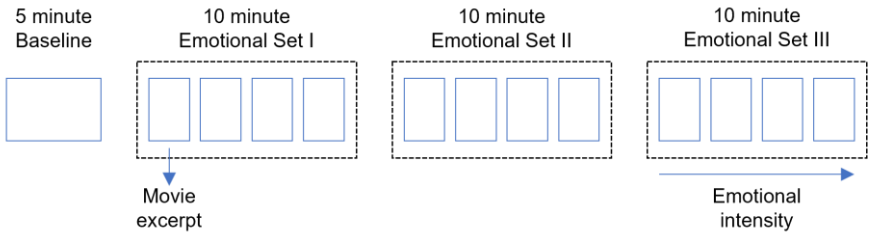

Figure 4 *- Sequence of movie excerpts for session 1.*

The first session had a video sequence totalizing 35 minutes of duration (Figure 4), contrasting with the 50 minutes for the second session (Figure 5). The second session was designed to record emotional responses to the stimulus presented in the first session as well as to the new stimulus. As shown in Figure 5, each emotional set totalized 15 minutes divided into 5 minutes of repeated excerpts (from the first session) followed by 10 minutes of new excerpts. The repeated videos allow the evaluation of the emotional response to the same stimulus on different days while the new videos evaluate the emotional response to the different stimuli of the same emotion. Moreover, the transition from the repeated to the new video excerpts adds the element of surprise to the experiment of each participant.

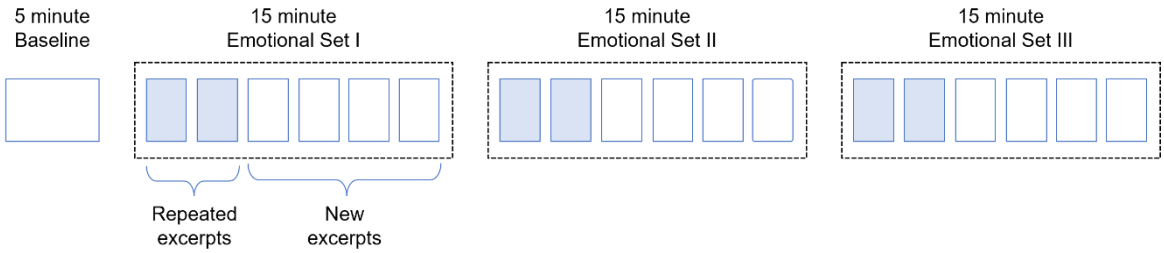

Figure 5 *- Sequence of movie excerpts for session 2, highlighting the excerpts repeated from the first session (blue) and the new excerpts (white).*

The minimum number of participants in the study was set as that allowing an equal counting of participants per emotional sequence. With the three emotional states (F, H and N) being ordered in different manners, six emotional sequences were considered (FHN, FNH, HFN, HNF, NFH, and NHF), each one with at least 5 participants, setting 30 as the minimum for this study. For each participant the order of the sequence was set the same for both sessions, to allow their comparison.

To assess the participants' characteristics and emotional states, 4 types of questionnaires (available online[2]) were applied based on sociodemographic characteristics, on the Toronto Alexithymia Scale (TAS-20), on the State-Trait Inventory for Cognitive and Somatic Anxiety (STICSA), and on the Visual

[2] https://github.com/carolinafalves/Repository_CarolinaAlves.git

Analog Scale (VAS). Each questionnaire was used for a specific function and at a particular time in the study, as shown in Figure 6.

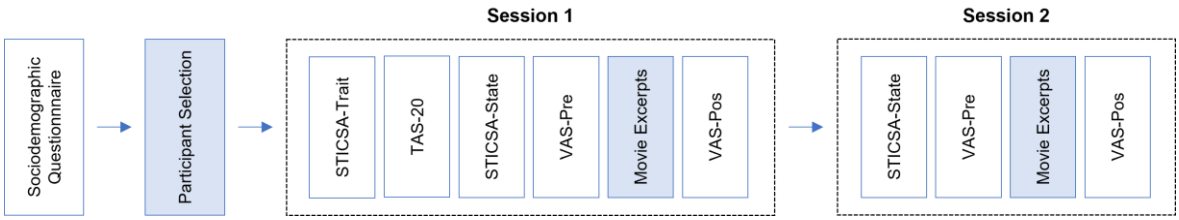

Figure 6 *- Sequence of questionnaires throughout the study.*

The sociodemographic questionnaire was used to select the participants, that is why it was the first to be answered. It includes multiple variables, such as age, gender, nationality, education, questions related to the health restrictions presented in the inclusion criteria of the participants (Section 3.2), and related to personal habits like drinking coffee and playing exercise.

The TAS-20 is a self-assessment instrument for assessing the alexithymia construct in its three dimensions: difficulty identifying feelings, difficulty describing feelings to others, and outwardly oriented thinking style [49]. The questionnaire consists of 20 statements to which the participant must respond with a degree of agreement with them, between 1 and 5 (1-Totally Disagree e 5-Fully Agree). This participant's report has the main goal of detecting participants that have difficulty in identifying their emotions, which can compromise the self-assessment of their emotional state during sessions. This questionnaire was answered only in the first session.

The STICSA is a self-report instrument to assess cognitive and somatic dimensions of trait and state anxiety [50]. The questionnaire is composed of several different affirmations regarding symptoms associated with anxiety, to which the participants should indicate how they feel on a scale between 1 to 4 (1-Never and 4-A lot). To assess the anxiety trait, the participant must answer how they usually feel, however, for the state they must report how they feel at a specific moment. The state is particularly interesting and informative for days when data is collected. Understanding how the participant feels at the beginning of the session may indicate or justified the answers in VAS questionnaires, and the physiological response of emotions. Thus, the STICSA-Trait questionnaire was answered only once at the beginning of the first session, but the STICSA-State questionnaire was answered at the beginning of both sessions.

The last questionnaire, based on the VAS, was designed to the participant self-assess their emotional state before (VAS-Pre) and after (VAS-Pos) viewing the video excerpts in both sessions. The purpose of the VAS-Pre was to detect the initial emotional state and, like the STICSA-State, to help explain specific characteristics of physiological signals. The VAS-Pos attended to assess the final participants' emotional state. Also, it was used to assess how the participants feel in the different movie sets, by recall memory. The questionnaire is divided into two parts. Firstly, it presents a scale from 0 to 100% regarding anxiety, happiness, fear, and stress. Then, the valence and arousal dimensions are introduced, in which the participant must answer on a scale of -5 to 5. Considering that not all participants may be familiar with the valence arousal terms, the scheme in Figure 7 was used to help clarifying these two concepts, in which low arousal and negative valence correspond to the minimum of the scale (-5), and high arousal and positive valence correspond to the maximum (5). The major goal of the VAS-Pos questionnaire was to understand the effectiveness of emotional induction.

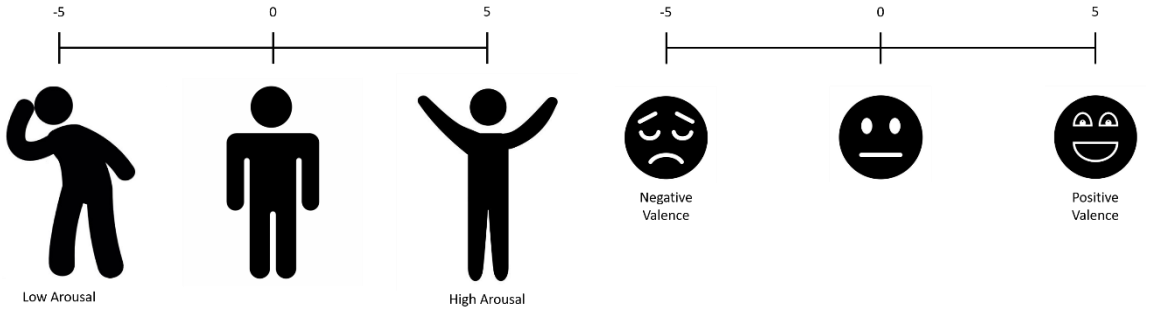

Figure 7 *- Scheme of arousal (left) and valence (right) dimensions used in the explanation of these concepts to the participants, on a -5 to 5 scale used in the VAS questionnaire.*

3.4 Materials and Setup

The physiological signals were acquired using the 4-channel biosignalsplux, from the Biosignalsplux Explorer Research Kit. This device collects the signals using sensors and transmits them via Bluetooth to the computer where they are visualized and stored [51]. As illustrated in Figure 8, the biosignalsplux hub acquires 4 signals making use of 5 sensors, namely 2 EMG sensors (one for EMG MF and the other to EMG TR), 1 EDA sensor, 1 ECG sensor, and an additional lead ground cable to serve as a reference to the sensors. Typically, each sensor has two electrodes, one red and another black, corresponding to the positive and negative electrodes, respectively. There are also reference electrodes, identified in Figure 8 in white.

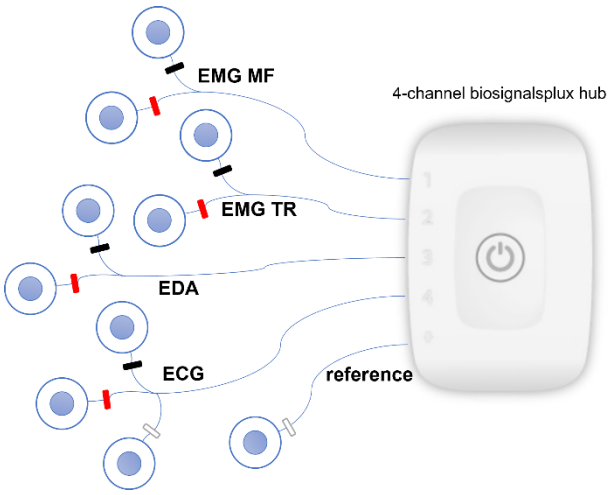

Figure 8 *- Connection diagram of acquisition sensors to the 4-channel biosignalsplux hub.*

To ensure that the collected data had no variations associated with the placement of the electrodes, it was particularly important to define their exact position and reproduce this protocol in all experiments. Based on BiosignalsPlux Guidelines and other studies that used the same signals, the position of each electrode followed the following instructions:

- EMG MF: the red electrode placed in the medial zone of the frontalis muscle and the black electrode in the lateral zone, as indicated in Figure 9.A);
- EMG TR: both electrodes placed in the middle trapezius, with the red electrode closer to the upper trapezius, as indicated in Figure 9.B);

- EDA: the red electrode placed above the thenar eminence and the black electrode above the distal palmar, as indicated in Figure 9.C);
- ECG: the ground truth electrode placed around 2 centimeters above the pelvis bone, and the red and black electrodes below the ribs, with the red one, shifted to the right and the black one to the left, as indicated in Figure 9.D);
- Reference: the single electrode placed above the clavicle on the left side of the body, as indicated in Figure 9.D).

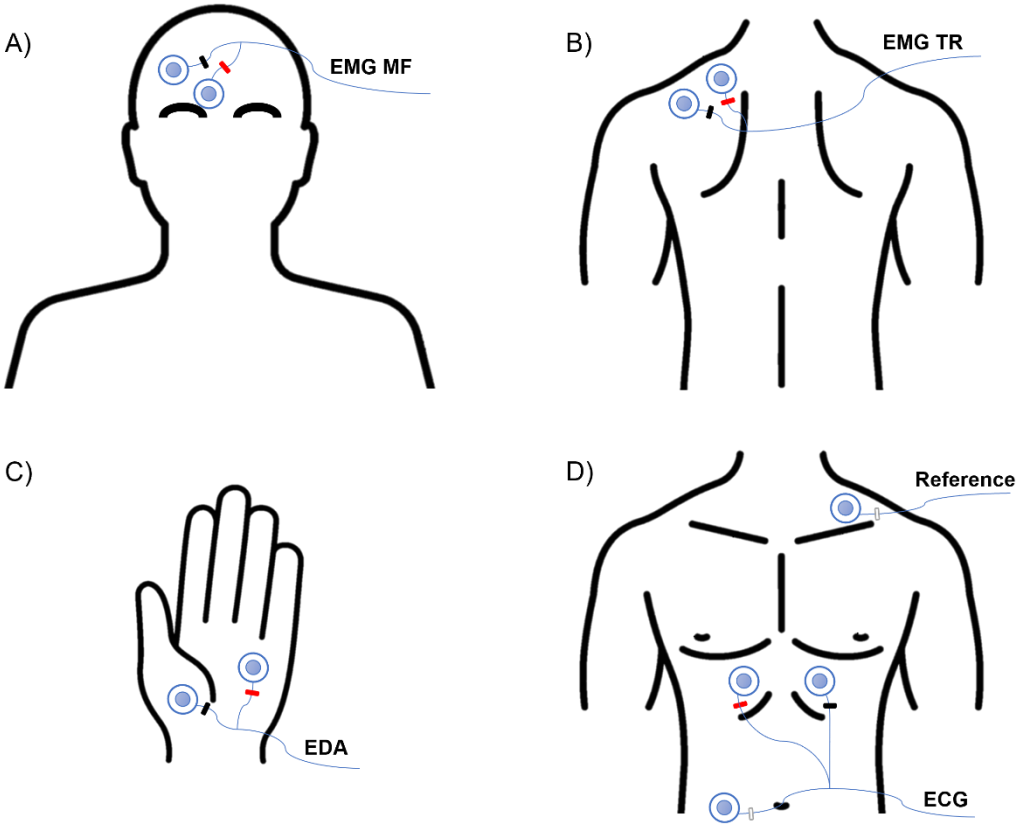

Figure 9 *- Electrode placement scheme for each sensor: A) EMG MF, B) EMG TR, C) EDA, and D) ECG and Reference.*

All experiments were conducted at the Instituto de *Engenharia Electrónica e Informática de Aveiro (IEETA,* http://www.ieeta.pt*) at University of Aveiro (UA,* https://www.ua.pt*)*, in a room prepared specifically for data collection. The surfaces in the room (tables and shelves) were all cleared out to minimize (as much as possible) external interference in the participant's experience. As presented in Figure 10, the equipment was placed in a corner of the room, the existing window was covered to minimize the visual external stimuli and the headphones were used to reduce external auditory stimuli. Also, a second computer was needed to control what was seen by the participant in his/her monitor, in addition to the computer that was connected to the biosignalsplux hub and where was installed the platform of visualization and acquisition of the signals, Opensignals (not shown in Figure 10). The second computer was also used to participants answer the questionnaires, through the monitor to avoid sharing of the equipment.

Figure 10 *- Experimental setup.*

3.5 Experimental Procedure

Before the participant arrival, it is necessary to ensure that all equipment is ready for use:

- Turn on the computers and the monitor;
- Open Opensignals on computer I;
- Open questionnaires and the movie excerpts on computer II;
- Place electrodes on sensors;
- Connect the headphones to computer II.

As soon as the participant arrives at the collection room, but before data collection, it is explained how the experiment will be conducted, he/she is prepared for the acquisition of physiological signals and some needed questionnaires are answered. The procedure is the following:

- Thank their participation and reinforce that they can give up from the study at any time;
- Inform how the experiment will be conducted and how their data will be used;
- Sign the informed consent;
- Place the sensors on the participant, according to the instructions of Figure 9, except for the EDA sensor that is only placed after answering the initial questionnaires;
- Lead the participant to sit on the chair near the monitor, and ask him/her to answer the initial questionnaires: STICSA-Trait, TAS-20, STICSA-State, and VAS-Pre;
- Make sure that the participant understands all questions and if there is any doubt, clarify it;
- When the participant finishes answering, place the EDA sensor and the headphones;
- Turn on the biosignals hub;
- Update the sensor on the Opensignals, define 1000 Hz for the sampling rate and start recording;
- Check if all sensors are well placed and if the acquisition track is ok;
- Open the movie excerpts' sequence and adjust the sound volume to 50 (on computer scale);
- Ask the participant to relax and to be as attentive as possible to the videos;

While the participant is viewing the excerpts, it is necessary to use the trigger to record the moments when there are set transitions. The five moments are:

- Beginning of baseline video;
- End of baseline video and Beginning of the Emotional Set I;
- End of Emotional Set I and Beginning of the Emotional Set II;

- End of Emotional Set II and Beginning of the Emotional Set III;
- End of Emotional Set III;

As soon as the stimulus ends, the participant answers the last VAS-Pos questionnaire, and the equipment is removed. The steps are:

- Stop the acquisition of the physiological signals in Opensignals;
- Turn off the biosignalsplux hub;
- Remove the headphones and the EDA sensor;
- Close movie excerpts;
- Open the VAS-Pos questionnaire;
- Ask the participant to answer the questionnaire and explain that the assessment of emotional state must be done using memory and to think about the most intense moment of each set;
- When the participant finishes, ask to get up and remove the remaining electrodes carefully. If necessary, clean the skin of the participant;
- Ask the participant if he/she knew any of the films and if so, point out which ones;
- Thank again for the participation, and if it is the first session, schedule the second session.

After the participant leaving, it is necessary to reorganize all the equipment and proceed to the disinfection of the entire room, following the order:

- Open the window and the door to air the room;
- Save the acquired physiological signals;
- Close Opensignals;
- Turn off the computers and the monitor;
- Disinfect biosignalsplux hub, sensors, monitor, mouse, keyboard, and headphones;
- Disinfect surfaces and chairs;
- Store biosignalsplux hub, sensors, and headphones.

3.6 Emotion MultimOdal daTabasE (EMOTE)

The EMOTE is the acronymous for Emotion MultimOdal daTabasE where “emote” stands for “give emotion to, in a stage or movie role”. This dataset is composed of files with simultaneous physiological signals (EMG MF, EMG TR, EDA, and ECG, 1000 Hz of sampling rate) collected from 30 emotionally stimulated participants in two sessions using movie excerpts. Each file contains an experimental setting (one participant in a given session) stored in a *.txt file. As the stimulation is continuous, each database file, also includes the moments of an emotional transition, marked with the trigger. Furthermore, information such as date and time of collection, sampling rate, number of channels used, the correspondent sensor, and color sleeve, are identified at the beginning of each file.

At a sociodemographic level, the 30 recruited participants are aged between 18 and 25 years (21.45±1.35), consisting of 70% female and 30% male (Figure 11). Regarding their professional status, 90% are students, 7% are student-workers, and 3% have another status (Figure 11). When asked about habits: 77% of participants drink coffee and 23% do not drink, whereas 63% practice physical exercise and 37% do not practice. However, should be noted that on the hour before of acquisition, participants did not drink coffee or practice exercise.

According to the participants' self-assessment, regarding their emotional state, while watching the videos, it can be concluded about the effectiveness of emotion stimulating and the variability of the

emotional experience of participants. Table 1 shows the mean and standard deviation of participants' responses to the VAS-Pre questionnaire for session 1.

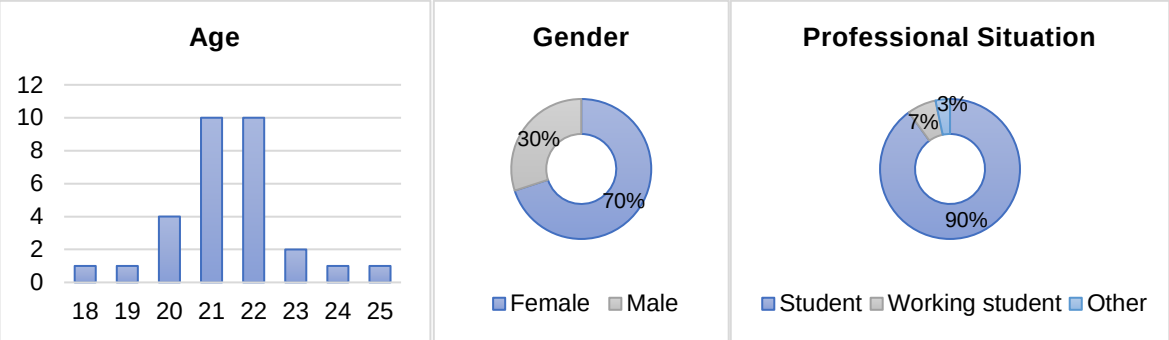

Figure 11 *- Results of the sociodemographic questionnaire regarding age, gender, and professional situation (total number of participantes = 30).*

Movie type (Emotion)	Mean ± STD					
	Anxiety [0,100]	Happy [0,100]	Fear [0,100]	Stress [0,100]	Arousal [-5,5]	Valence [-5,5]
Horror (Fear)	66.55 ± 19.17	13.79 ± 14.24	56.21 ± 23.84	58.28 ± 22.60	2.28 ± 2.48	-1.79 ± 1.94
Comedy (Happiness)	10.34 ± 13.77	73.10 ± 14.41	3.10 ± 8.35	6.55 ± 12.11	2.93 ± 1.36	3.72 ± 1.08
Documentary (Neutral)	11.38 ± 16.55	48.28 ± 22.90	3.10 ± 7.00	7.24 ± 11.11	0.69 ± 2.07	1.76 ± 1.75

Table 1 *- Results of the VAS-Pre questionnaire regarding the participants' emotional state after emotion stimulation in session 1 (total number of participantes = 30).*

In the case of horror movies, the average of responses for fear rating was 56.21 ± 23.84 (on a 0-100 scale), being above the middle of the scale, and thus considered that fear was stimulated. However, as can be seen, there was a high variability of answers that can be justified by the fact that some participants reported knowing some of the excerpts, and for that reason, their emotional experience may have been different. It should also be noted that participants reported high levels of anxiety and stress, which can be explained by the fact that horror movies may induce that kind of emotion. As for the valence and arousal dimensions, the classification is in accordance with the expected: positive arousal (2.28 ± 2.48) and negative valence (-1.79 ± 1.94).

For comedy movies, the average of responses for happiness is 73.10 ± 14.41 (on a 0-100 scale), thus indicating that this emotion was properly stimulated. In this case, the variability of responses is smaller, which is indicative that most participants had more similar intensities of emotional response (in comparison to horror movies). Although they also reported knowing some of the showed excerpts, they were still effective. Regarding valence and arousal, as expected, participants reported positive arousal (2.93 ± 1.36) and positive valence (3.72 ± 1.08).

Regarding documentaries, participants reported an average of 48.28 ± 22.90 for happiness. Despite the average value being less than 50, the standard deviation is quite high, indicating that the participants' emotional experiences were diversified and, therefore, for some of them, the emotional state neared a happy state. Regarding the assessment through dimensions, it was found that for arousal, participants reported values close to 0 (0.69 ± 2.07), which was the objective, and for valence, reported relatively higher intensities than expected (1.76 ± 1.75), which agrees with the answers through discrete emotions. Thus, can be concluded that the induction of neutral state was the least effective, with a great diversity of responses.

4. Methodology and Implementation

The execution of the experimental protocol allowed the creation of a diverse and extensive database now to be analysed. The following step was to design an emotional model that encompasses methods and techniques of accomplishing the final three goals of this work: understand which signals and features best describe emotions, construct and optimize an emotion identification system, and understand the impact of each model (discrete and dimensional) on emotion classification.

The main steps that constitute the emotional model are: reading and data organization, feature extraction, feature selection, and classification (Figure 12). The model takes the physiological data that comes from the database (Section 3.6), composed by the EMG of medial frontalis (EMG MF), EMG of trapezius (EMG TR), EDA, and ECG, from session 1; and returns a final emotion identification of the three emotional states: fear, happiness and neutral.

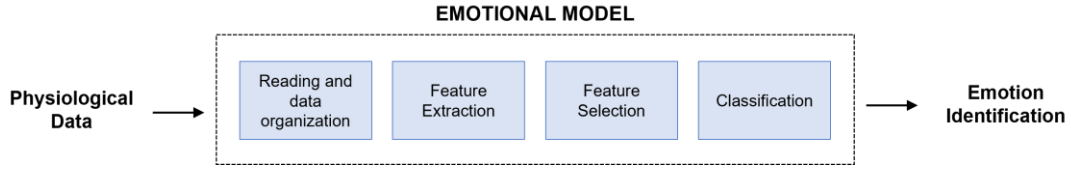

Figure 12 *- Sequence of main steps of the proposed emotional model.*

Each one of these steps has a specific purpose, and the combination of all has a great impact on the prediction of emotions by the optimized emotional model. Therefore, in the following sections, each step's purpose is detailed as well as the methods and implementations that were selected. For the implementation, Python (version 3.8.5) was chosen as the programming language being composed by variety of packages and functions that allowed the creation of the whole emotional model using only this language.

4.1 Reading and data organization

The first main step of the emotional model aims to organize the physiological data. The files created by the signal acquisition system contain different types of data, so it was important to design a data reading software. To automatically read all files, the function was designed to deal with possible irregularities, and their organization was planned so that in the rest of the work its accessibility was easy and fluid.

Regarding irregularities, the function was designed to (1) detect which sensors were used, and if any do not have values, report the occurrence, (2) confirm the placement of the sensors, and if there is any modification, reorganize the data and (3) detect the triggers, and if there is more than what is supposed (a result of a reflection or anticipation of the transition), eliminate the additional one based on the time of the transitions. By incorporating these tests into the function, it was ensured that the data used were correct, belonging to the right signal and to the right induced emotion.

With the data properly organized, the physiological signals were divided by participant and condition. Knowing the moment when there is a transition, through the trigger, it is possible to access the signals' data and divide them by condition (Baseline, F, H and N). That is, from the first trigger to the second, the signals always correspond to the baseline, from the second trigger to the third, corresponds to the first emotional state, and the division continues until the end. It should be noted that the order of the emotional states varied from participant to participant, therefore, this function

also had to incorporate the sequence of each participant to match each condition to the right emotion and guaranty the correct association to the set of signals.

The data was organized following the dictionary format, in which each participant is associated with the four conditions, and each condition is associated with the four physiological signals. The reading of all files was performed, and the resulting dictionary was saved so that it is only necessary set a given participant, condition, and physiological signal, to access the data.

4.2 Feature Extraction

The feature extraction encompasses the preprocessing of signals and the extraction of feature data that can be used in the classification step (Figure 13). Through the preprocessing, time series of features with reduced dimensionality are obtained, when compared to signals (smaller number of temporal observations). Then, their statistical characteristics are inspected to appropriately describe them through time and to reduce the computational cost in the emotional classification. Therefore, the final feature data are statistical metrics from the feature time series but with an even reduced dimensionality (a vector of features per set of signals/series).

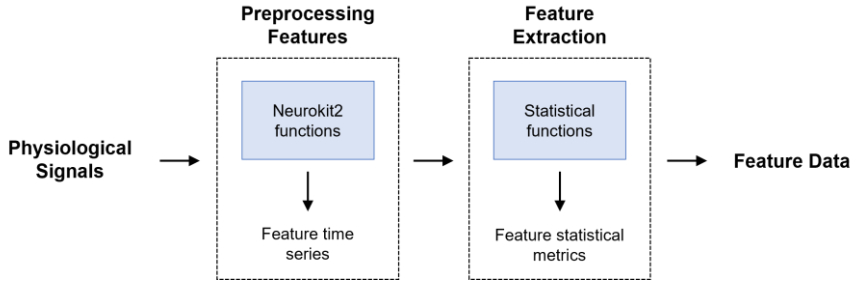

Figure 13 *- Methodology of feature extraction step.*

4.2.1 Preprocessing

The preprocessing of physiological time series was implemented using the NeuroKit 2, a Python toolbox with advanced biosignals' processing [52]. It contains specific functions to extract specific features of given physiological signals, so the preprocessing of each signal was performed independently. Following the logic of the reading function (described in Section 4.1), the preprocessing step was repeated for each signal, condition, and participant.

Electromyogram (EMG)

The preprocessing of EMG MF and EMG TR was performed using the same functions: nk.emg_clean, nk.emg_amplitude, and nk.emg_activation.

The nk.emg_clean function cleans the EMG signal using a fourth-order 100 Hz highpass Butterworth filter and returns a vector of the filtered signal that is used by the other two functions.

The nk.emg_amplitude takes the filtered signal and computes the amplitude by calculating the linear envelope of the signal. As the return, the function gives a vector containing the envelope amplitude.

The nk.emg_activation detects onsets and offsets of the signal based on the amplitude threshold. Takes as parameters the cleaned signal, the sampling rate, and a method or threshold. In this case, the default method was not suitable, so it was chosen the method 'biosppy', in which the threshold corresponds to 1.2 times the mean of the absolute smoothed full-wave-rectified signal. As the return, it is obtained a dictionary with all the onsets and offsets of the activation moments.

With the information of activation's onsets and offsets was calculated the duration, the maximum peak amplitude, the average of amplitudes, and the area of the EMG activations. To complete the EMG feature data, it was also selected the envelope amplitudes of all activations.

Electrodermal activity (EDA)

In the preprocessing of the EDA signal, it was used the nk.eda_process and nk.eda_sympathetic functions.

The nk.eda_process is a general function able to generate different types of data taking the raw signal, the sampling rate, and a method to define the processing pipeline, that in this case was the 'neurokit'. The function returns a dataframe of the same length as the signal, containing several columns with different features, and a dictionary, containing information of the Skin Conductance Response (SCR) peaks. From the dataframe, the selected features were the tonic and phasic components of the EDA and the amplitude of the SCR signal containing the tonic component. From the dictionary were selected the amplitude, rise time, and recovery time features.

Regarding the nk.eda_sympathetic, the function takes the raw signal, the sampling rate, and the method 'posada' to compute the EDA indexes of the sympathetic nervous system normalized, obtained by dividing it by the total power, and not normalized. Since these two features are an index, they are not time series as the previous ones.

Electrocardiogram (ECG)

The preprocessing of the ECG signals takes the functions: nk.ecg_clean, nk.ecg_peaks, nk.ecg_delineate, nk.ecg_rate, and nk.hrv.

The nk.ecg_clean cleans the signal and prepares it for R-peak detection. It takes the raw signal, the sampling rate, and the method that defines the processing pipeline used was the 'neurokit', based on the Butterworth filter. The function returns the clean ECG that is then used by the nk.ecg_peaks and the nk.ecg_delineate.

The nk.ecg_peaks uses the clean signal and the sampling rate, and finds the R-peaks of the ECG signal using the default method 'neurokit'. The result of the function is a dictionary that contains the samples at which R-peaks occur.

The nk.ecg_delineate delineates the QRS complex using the cleaned signal, the sampling rate, and the dictionary from the nk.ecg_peaks. The function returns a dictionary containing all the peaks, and with the T wave offsets and peaks, it was calculated the half duration of the T wave.

Lastly, the nk.ecg_rate takes the R-peaks and the sampling rate to calculate the instantaneous heart rate as a vector, and the nk.hrv takes the R-peaks and the sampling rate to compute the HRV indices in the time, frequency, and nonlinear domain. The function returns a dataframe with HRV features from all domains, represented by only one value and not time series as the previous functions.

4.2.2 Feature Extraction

The features associated with a given time series obtained in the preprocessing step (for a given data file) were then extracted. This allowed obtaining several features to represent a certain aspect of the physiological time series by a single number (e.g. mean, standard deviation, etc).

The statistical metrics extracted from a given time series included the mean, the standard deviation, the variance, the skewness, and the kurtosis. All these metrics were calculated using functions from the NumPy module [53], in the case of the mean (np.nanmean), standard deviation (np.nanstd) and variance (np.nanvar), and from the statistical functions of the SciPy module (scipy.stats) [54], in the case of skewness (stats.skew) and kurtosis (stats.kurtosis).

The final feature data obtained in this step was saved in a dataframe and organized following the scheme of Figure 14, where each participant was divided into conditions and each condition divided into extracted features, now represented by single values and not time series.

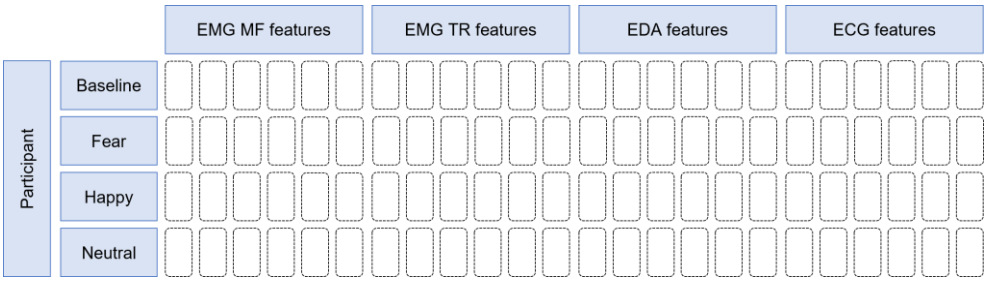

Figure 14 *– Representation of the organization of feature data in a dataframe.*

Stationarity Analysis

The process of extracting statistical metrics from a time series assume that the the time series are a realization of a stationary stochastic process. This hypothesis was investigated through the Augmented Dickey-Fuller (ADF). The ADF tests if an unit root (that causes non-stationarity in trend) is present. The null hypothesis is that there is a unit root, which implies that the time series is a realization of a non-stationary process, against the alternative hypothesis stating that there is no unit root. Thus, it is a favorable decision if the resulting p-value is lower than the significance value so that the null hypothesis of non-stationarity is rejected [55].

The code available on [56] was used to implement the ADF test, and the package used was the Statsmodel [57], which provides functions for the estimation. The function to perform the test is called statsmodels.tsa.stattools.adfuller, and it returns the p-value of the test. The analysis of the results and the conclusions about the stationarity considered a 5% significance level.

4.3 Feature Selection

The feature selection step (Figure 15) is composed of a univariate and bivariate statistical analysis that allows the reduction of the dimensional space of the variables by removing redundant data. The univariate analysis intends to detect which features can distinguish at least one of the four conditions, and, consequently, understand which physiological signals are more related to each condition. The bivariate analysis intends to find patterns and relationships between pairs of features. Additionally, at the beginning of this step, a normality test is performed to evaluate the distribution of each feature.

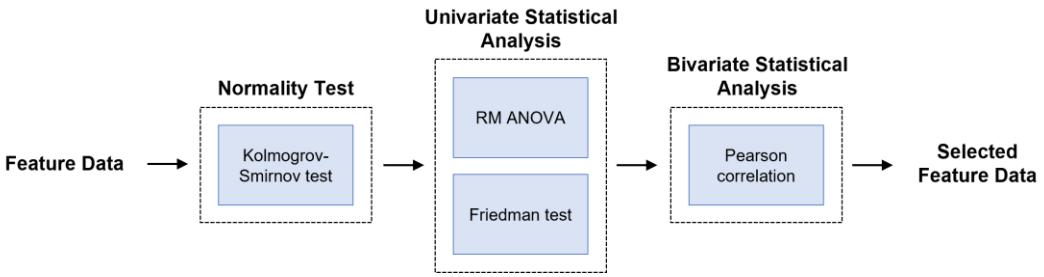

Figure 15 *- Methodology of feature selection step.*

4.3.1 Normality Test

The Kolmogorov-Smirnov (KS) test is based on the maximum vertical difference between the empirical and hypothetical cumulative distribution. Considering H(x) the hypothetical distribution, and F(x) the distribution estimated for the data samples, the null hypothesis states that the two distributions are identical H(x) = F(x) for all x, which means that data follow a specific distribution, whereas the alternative hypothesis states that H(x) ≠ F(x) for at least one value of x, meaning that data does not follow a specific distribution [58], [59].

The KS test was implement with the function stats.kstest from the SciPy module [54], which takes the samples and the type of distribution to test, which in this case was the normal distribution ('norm'), and performs the test, returning the p-value. If the p-value is higher than the significance value of 5% then the null hypothesis is not rejected, and one can consider that data comes from a normal distribution.

In addition to the KS test, the normality was investigated through the plots generated by the sns.ecdfplot function from the Seaborn module [60] that computes the cumulative distribution of a given data. The final plot shows the cumulative distribution calculated by the function for the data and the normal distribution with parameters from the data. Through this, it was possible to visualize the vertical difference between the two distributions and to inspect the normality assumption.

Both analyses were performed to set of samples for each feature in the four conditions. The results varied from feature to feature and also from condition to condition, having features where data followed a normal distribution for one condition and not for another.

4.3.2 Univariate Statistical Analysis

Each feature has values associated with the four conditions of each participant (Figure 14). Considering conditions as groups, each one is represented by all participants, meaning this that groups are not independent and there is a relationship between them. Thus, to execute the univariate analysis, the most indicated type of ANOVA was the Repeated Measures ANOVA (RM ANOVA).

The RM ANOVA is similar to one-way ANOVA, however, the groups are related and not independent. The goal of this analysis is to detect significant differences between the means of the groups, that is if there are one or more groups that are significantly different from the rest [61]. It requires a continuous dependent variable and a categorical independent variable, which in this case were the feature samples and the four conditions (groups), respectively.

The assumptions for the RM ANOVA are that the dependent variable should be continuous and normally distributed (which was inspected by KS test), that the samples between subjects are independent, and that variances are homogeneous among groups (meaning that variances must be equal between groups - the assumption of sphericity) [62], [63]. For the variables that were not normally distributed (rejecting the KS null hypothesis) a non-parametric test was used, described further below.

The null hypothesis states that the expected value of each group are equal, and the alternative hypothesis states that at least two groups are statistically different, which means that one group's expected value is different from that of another group. Thus, if the p-value is less than the significance level of 5% then the null hypothesis is rejected and at least the expected value of one group is statistically different from that of the remaining groups. This indicates that at least one condition stands out from the rest and that the feature might be able to descriminate emotional states.

The F-statistic for the RM ANOVA is:

$$F = \frac{MS_{conditions}}{MS_{error}}$$

with $MS_{conditions}$ being the mean sum of squares for conditions and the MS_{error} being the mean sum of squares for error [63].

The advantage of using the RM ANOVA instead of the convencional ANOVA is that the groups are represented by the same participants so there is no subject variability, which decreases the MS_{error} and increases the F-statistic.

This analysis was performed with the function pingouin.rm_anova from the Pingouin module [64]. The function takes the data and it was defined the dependent variable, that are the features, the within factor, that are the conditions, and the subject, that are the participants. Besides performing the RM ANOVA, the function also computes the Mauchly's test to evaluate the sphericity and determine if the p-value must be corrected or not. Thus, the function returns the p-value of the RM ANOVA, the F-statistic of RM ANOVA, the p-value of the Mauchly's test, and the Greenhouse-Geisser corrected p-value of RM ANOVA. Every time that the p-value of the sphericity test is higher than the significant level of 5%, the p-value of RM ANOVA considered to the analysis was the corrected one.

Regarding the normality assumption, for cases where the null hypothesis of normality was rejected, the Friedman test was applied. The Friedman test, unlike ANOVA, is a non-parametric test that can be applied to cases where normality is not assured. It is the alternative to the RM ANOVA, so it is very similar and tests exhibit similar null and alternative hypothesis [65].

The implementation of the Friedman test used the function pingouin.friedman, also from the Pingouin module [64], and takes the same parameters as the pingouin.rm_anova, but does not compute the statistical correction. It returns the F-value and p-value of the Friedman test and when the p-value is lower than the significant level of 5% then the null hypothesis is rejected, meaning that at least one group is statistically different from the remaining ones.

To complement the analysis of RM ANOVA and Friedman, violin plots were generated using the function sns.violinplot from the Seaborn module [60], where it was possible to identify the distribution of feature samples for each condition. The plot is a combination of boxplot and kernel density estimation and when a feature is able to describe emotional states, then in the plot at least one of the conditions is misaligned from the others.

The last univariate analysis was a post hoc test to produce multiple t-tests between all possible combinations of the groups with Bonferroni correction. With this test, in addition to knowing if there is statistical evidence, it was possible to determine which groups are statistically different and, consequently, which features describe each condition.

This post hoc test was implemented via the pingouin.pairwise_ttest from the Pingouin module [64]. The function takes the data and was defined the dependent variable, that is the feature, the within, that are the emotions, the subject, that are the participants, and the post hoc test to perform the correction of the p-value, that was the 'bonf' (Bonferroni correction). If the p-value of each combination of groups is lower than the original significant level of 5 % then those two groups are statistically different.

4.3.3 Bivariate Statistical Analysis

Through the bivariate statistical analysis, it is possible to determine relationships between pairs of features. The selected approach used the Pearson's correlation to detect pairs of highly correlated features and, therefore, redundant. For the classifier, only one feature of the pair is needed since it contains most of the useful information, so the other one can be disregarded from the analysis [66].

The Pearson correlation coefficient is a measure of the linear relationship between two continuous variables (X and Y), based on the method of covariance:

$$\rho = \frac{cov\ (X,Y)}{\sqrt{var(X) \cdot var(Y)}}$$

The coefficient is such that $-1 \leq \rho \leq 1$, whereas ρ = 0 represents no correlation between variables, $\rho = -1$ negative correlation, and $\rho = 1$ positive correlation.

The Pearson correlation assumes that the two variables are continuous and paired, meaning that for each case (participant) there are two observations (one for each variable); the cases should be independent, and theoretically, variables should have a linear relationship, follow a bivariate normal distribution and have the same finite variance – homoscedasticity [61], [67].

It was used the function pandas.corr from the Pandas module [68] to implement the analysis and it was defined the 'pearson' method to compute de Pearson correlation. The function returns a correlation matrix that was graphically represented using the sns.heatmap function from the Seaborn module [60], to complement the analysis and easily identify the highly correlated pair of features.

For feature selection, the Pearson correlation coefficient was used to detect highly correlated feature pairs and with the p-value of the univariate statistical analysis (RM ANOVA or Friedman), the feature with the lowest p-value, the one with the highest statistical significance between conditions, was selected for the final set of selected features. The correlation coefficient threshold considered to detect the highly correlated pairs was 0.8, allowing to reduce about half of the features.

4.4 Classification

As presented in Figure 16, the classification step encompasses the classifier, data standardization, feature optimization, hyperparameters optimization, and, lastly, building up the optimized emotional model. All these processes aim to find the optimized model that best describes the different emotional states. Data standardization was used to reduce the effect of the features value's range, feature optimization to select the best combination of features, and hyperparameters optimization to select the combination of hyperparameters that increases the performance of the final optimized emotional model.

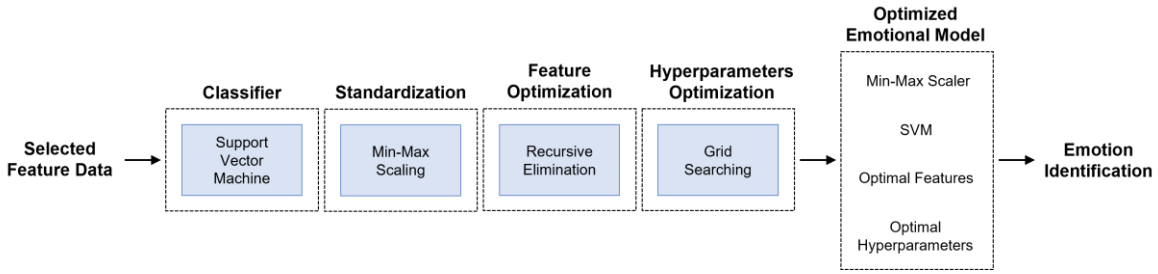

Figure 16 *- Methodology of the classification step.*

4.4.1 Classifier

Support vector machine (SVM) is a supervised-learning model based on the construction of one or more hyperplanes in a high-dimensional space (Figure 17.A), which can be used for regression or classification tasks, and even for outliers' detection. The goal is to find the hyperplane in a N-dimensional space that best distinguishes the data points (Figure 17.B).

Support vectors are data points that define the position and orientation of the hyperplane (Figure 17.B). They are the data points considered by the SVM to determine the hyperplane, meaning that if one of them is moved then the margin of the hyperplane also changes, which won't happen if the other data points (that are not support vectors) move.

The optimal hyperplane is the one that maximizes the margin, i.e. the distance between the data points of different classes (Figure 17.B), and the greater the margin, the greater is the model's ability to predict future data points.

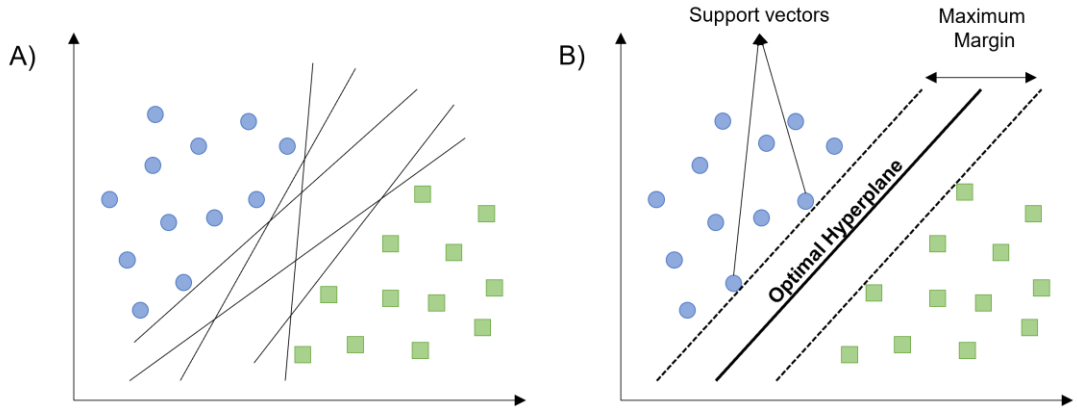

Figure 17 *- A) Representation of data points and possible hyperplanes. B) Representation of data points and the optimal hyperplane that maximize the margin. Near the margin are also represented the support vectors.*

The original SVM algorithm is linear, and the task is to find a linear classification hyperplane $w^T x + b$ that satisfies:

$$y_i(w^T x_i + b) \geq 1, i = 1,2, \dots, N$$

, where x_i corresponds to the feature vector and y_i to the actual label [69], [70].

The above equation is strictly used for linearly separable data. To extend the SVM algorithm to the case of nonlinearly separable data, the loss function hinge loss is used, and the problem is then described by:

$$y_i(w^T x_i + b) \geq 1 - \max\big(0,1 - y_i(w^T x_i - b)\big), i = 1,2, \dots, N$$

Considering $\xi_i = \max\big(0,1 - y_i(w^T x_i - b)\big)$, that is the allowed degree of deviation from the ideal linear separability condition, the final goal of optimization is to minimize the cost function:

$$\frac{1}{2} w^T w + C \sum_{i=1}^{N} \xi_i$$

, where C is a penalty factor to control the degree of penalty to misclassification of samples [69], [70].

To transform SVM more global, a nonlinear SVM was later proposed by applying the kernel trick to maximum margin hyperplanes. The algorithm fits the maximum margin hyperplane in a high dimension transformed feature space, where data can be separable.

The SVM classifier used in this work was the sklearn.svm.SVC(), from the module Scikit-learn [71], and its implementation is based on [72]. It is a function that computes the SVM algorithm, that may be used for binary or multiclass classification and allows the user to define some entry hyperparameters that highly influence the results.

The multiclass classification is handled according to a one-vs-one scheme. One-vs-one refers to the use of $k(k-1)/2$ binary classifiers and when there are new samples, the outputs of all classifiers are aggregated by the majority voting approach [69].

The primal problem solved by the function is:

$$\min \quad \frac{1}{2} w^T + C \sum_{i=1}^{N} \xi_i$$

$$subject\ to \quad y_i(w^T \phi(x_i) + b) \geq 1 - \xi_i \qquad \xi_i \geq 0, i = 1, \dots, n$$

, and the dual problem is:

$$\min \quad \frac{1}{2} \alpha^T Q \alpha - e^T \alpha$$

$$subject\ to \quad y^T \alpha = 0 \qquad 0 \leq \alpha_i \leq C, i = 1, \dots, n$$

, where e represents the vector of all ones, α_i represents the dual coefficients upper-bonded by C, and Q is an n by n matrix $Q_{ij} = y_i y_j K(x_i, x_j)$, where $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$ represents the kernel function [73].

Finally, when the optimization problem is solved, the decision function is modified to:

$$\sum_{i=1}^{L} y_i \alpha_i K(x_i, x) + b$$

, where L is the number of support vectors, and the predicted label corresponds to the sign of the decision function [69].

4.4.2 Standardization

The SVM model is not scaler invariant so that the data must be scaled. Besides that, it is already a good practice to scale data since the features with different scales are seen with different importance by machine learning algorithms. Usually, features with high scales tend to have a more decisive role when training.

When choosing the best scaler, it is crucial to consider the type of data, in particular, its distribution and the presence of outliers. Therefore, the choice of the method, besides having the intention of improving the performance of the model, must also admit if the outliers are (or not) to be consider.

Among the different methods of standardization, some are highly sensitive to outliers and others are more robust. From the most sensitive, the most used ones are the StandardScaler and the MinMaxScaler. StandardScaler assumes that data is normally distributed and scales each feature so that the mean is zero and the standard deviation is one; MinMaxScaler assumes that data is not normally distributed and transforms each feature individually to the range between zero and one, or between -1 and 1 when there are negative values. If data contains outliers scaling using mean and

standard deviation is not the best option that is why RobustScaler is based on percentiles decreasing the effect of marginal outliers [74].

Among all scalers, the selected one was the MinMaxScaler because it does not assume the normality of data and has a more suitable response to the presence of outliers. Testing the final feature data, with the Kolmogorov-Smirnov test, was verified that the features do not follow a normal distribution, as would be expected since each feature has values of the four conditions and therefore there must be variability. Regarding the outliers, the violin plots detected some outliers, however, the number of features in which they are present is quite small. In addition, the outlier itself may not be a true outlier, but a value that, despite being apart from the rest, represents an intense emotional response, depending on the feature. That is why the MinMaxScaler was considered to be more suitable.

The function to implement the standardization was the sklearn.preprocessing.MinMaxScaler from the Scikit-learn [71].

4.4.3 Feature Optimization

There are several methods for feature selection, and they can be divided into three main categories: filter-based, where a metric is specified and features are filtered according to its value, wrapper-based, where the selection is based on searching, and embedded, that uses algorithms incorporating selection methods [75].

The previous feature selection was based on metrics such as the correlation between features and the ANOVA p-value, so it belongs to the filter-based methods. However, to find the features that optimize the model's performance was chosen a wrapper-based method, the backward elimination, so that is possible to make the selection based on the SVM model and to know which features have the greatest influence on the recognition and distinction of emotions.

The function used was the sklearn.RFECV(), Recursive Elimination with Cross-Validation, from the Scikit-learn module [71]. This technique aims to find the optimal number of features and even the ones that have the highest ranking, the optimal ones. It begins by building a model with all data and computing the importance score for each feature. The feature with the lowest score is the less important one, so it is removed, and the process is repeated all over again. Finally, the subset of features that has the best score is selected.

In this case, the function receives as an estimator, the SVM classifier; as cross-validation technique, the StratifiedKFold, that is specially used for a binary or multiclass label; and as scoring, the accuracy. Since the score used is the accuracy, when the process ends, it was possible to analyze the evolution of the accuracy through the subsets of features.

4.4.4 Hyperparameters Optimization

The search for the optimal hyperparameters was the final step to achieve the optimal emotional model. The hyperparameters to optimize were the kernel function, and the parameters: C, gamma, and degree.

Starting with the kernel function hyperparameter, three stand out: linear (kernel='linear'), polynomial (kernel='poly'), and radial basis function (RBF) (kernel='rbf') [76].

The other three hyperparameters that have a critical influence on the SVM's performance are the hyperparameter C, which is the penalty error that represents the misclassification, and 'tells' the SVM how much error is bearable, the hyperparameter gamma, which defines how much influence a single training sample has, and the hyperparameter degree that specifies the degree of the polynomial

kernel. Regarding C, a low value makes the decision surface smooth, the margin is big and can lead to underfitting, while a high C aims to correctly classify all training samples, have a small margin, and, thus, can lead to overfitting [73]. Regarding gamma, it is only related to RBF kernel function, and when it has a high value, only the nearby samples to the decision boundary are considered, however when its value is low, samples further away from the hyperplane are also considered.

Given the great influence of these hyperparameters on the model and to overcome the random or intuitive choice of them, the best method of finding the best combination of hyperparameters is to use a grid search methodology. To implement this method was used the sklearn.model_selection.GridSearchCV() function, from the Scikit-learn module [71].

The GridSearchCV function takes a dictionary of hyperparameters including different kernel functions, C's and gamma's; and an estimator, that was the sklearn.svm.SVC() without any hyperparameter. First, by fitting the data, a loop with cross-validation is run to find the best combination of hyperparameters, then with this combination, a new fit to the data is made, without cross-validation, to build the final model using the best hyperparameters. This process is exhaustively repeated, and the final model is the sklearn.svm.SVC() with the hyper tuned parameters that maximize the score of prediction.

The way that data was split into training and testing sets highly influenced the results of the grid-searching. To overcome this problem, the grid searching of the best parameters was also repeatedly performed using different splits to guaranty that the selection of the optimal hyperparameters was not biased to the choice of the sets.

5. Results and Discussion

5.1 Exploratory Data Analysis

The feature extraction step identified a total of 147 features being 26 from the EMG MF, 26 from the EMG TR, 33 from the EDA, and 62 from the ECG. Comparing the number of features extracted from each signal (and corresponding time series), the ECG signal stand out, from which many features were extracted.

Table 2 summarizes the features extracted from each physiological signal. In the rest of the document, the denomination used for each feature is the combination of what is in brackets on the left side of the table, regarding the signal, with the features that are designated on the right. The features on the right side are a combination of the feature time series with the statistical metric that represents it, except for the HRV features.

Signal	Feature
EMG MF (EMG_MF_)	Activations_N; Duration_Mean; Duration_Std; Duration_Var; Duration_Skew; Duration_Kurt; MaxPeakAct_Mean; MaxPeakAct_Std; MaxPeakAct_Var; MaxPeakAct_Skew; MaxPeakAct_Kurt; MeanPeaksAct_Mean; MeanPeaksAct_Std; MeanPeaksAct_Var; MeanPeaksAct_Skew; MeanPeaksAct_Kurt; all_Amplitude_Mean; all_Amplitude_Std; all_Amplitude_Var; all_Amplitude_Skew; all_Amplitude_Kurt; Area_Mean; Area_Std; Area_Var; Area_Skew; Area_Kurt
EMG TR (EMG_TR_)	Activations_N; Duration_Mean; Duration_Std; Duration_Var; Duration_Skew; Duration_Kurt; MaxPeakAct_Mean; MaxPeakAct_Std; MaxPeakAct_Var; MaxPeakAct_Skew; MaxPeakAct_Kurt; MeanPeaksAct_Mean; MeanPeaksAct_Std; MeanPeaksAct_Var; MeanPeaksAct_Skew; MeanPeaksAct_Kurt; all_Amplitude_Mean; all_Amplitude_Std; all_Amplitude_Var; all_Amplitude_Skew; all_Amplitude_Kurt; Area_Mean; Area_Std; Area_Var; Area_Skew; Area_Kurt
EDA (EDA_)	Symp; SympN; Tonic_Mean; Tonic_Std; Tonic_Var; Tonic_Skew; Tonic_Kurt; Phasic_Mean; Phasic_Std; Phasic_Var; Phasic_Skew; Phasic_Kurt; SCR_Height_Mean; SCR_Height_Std; SCR_Height_Var; SCR_Height_Skew; SCR_Height_Kurt; SCR_Amplitude_Mean; SCR_Amplitude_Std; SCR_Amplitude_Var; SCR_Amplitude_Skew; SCR_Amplitude_Kurt; SCR_RiseTime_Mean; SCR_RiseTime_Std; SCR_RiseTime_Var; SCR_RiseTime_Skew; SCR_RiseTime_Kurt; SCR_RecoveryTime_Mean; SCR_RecoveryTime_Std; SCR_RecoveryTime_Var; SCR_RecoveryTime_Skew; SCR_RecoveryTime_Kurt
ECG (ECG_)	Rate_Mean; Rate_Std; Rate_Var; Rate_Skew; Rate_Kurt; Tduration_Mean; Tduration_Std; Tduration_Var; Tduration_Skew; Tduration_Kurt;
ECG (HRV_)	RMSSD; MeanNN; SDNN; SDSD; CVNN; CVSD; MedianNN; MadNN; MCVNN; IQRNN; pNN50; pNN20; TINN; ULF; VLF; HTI; LF; HF; VHF; LFHF; LFn; HFn; LnHF; SD1; SD2; SD1SD2; SCSI; CVI; CSI_Modified; PIP; IALS; PSS; PAS; GI; SI; AI; PI; C1d; C1a; SD1d; SD1a; C2d; C2a; SD2d; SD2a; Cd; Ca; SDNNd; SDNNa; ApEn; SampEn

Table 2 *- Extracted features from each physiological signal.*

Due to the inter and intra variability that characterizes the acquired signals, the inspection of the stationarity assumption was used to validate the reduction of the time series into a value expressing a certain feature. Through this analysis, most of the cases were considered as stationary. Nevertheless, a small number of participants present some series that were not realizations of a stationary process (at 5% level), which may be related to a greater difficulty in eliciting emotions in these participants, and as result the trend of the emotional response through the 10 minutes is time dependent. Regarding the EDA time series, the stationarity test was not performed because these

series had the particularity of having few samples. Thus, the stationarity was verified through the cumulative distribution plots.

Physiological data, in general, are prone to outliers and artifacts that may mislead the conclusions of any statistical analysis. Therefore, it was carried out an analysis to verified the existence of NaN (Not a Number) values in the extracted features. From this analysis, two features from the ECG: 'HRV_ULF' and 'HRV_VLF' were eliminated, because they only contained NaN values. This action led to 145 features remaining.

Considering the EDA features that presented one or more NaN values: 'EDA_SCR_Height_Skew', 'EDA_SCR_Amplitude_Skew', 'EDA_SCR_Risetime_Skew', and 'EDA_SCR_Recoverytime_Skew', a detailed analysis was carried out in the two participants with this particularity (Figure 18). By the visual inspection of the signals, it was observed that both signals were partially saturated. This means that the phasic component of the signal was not detected, and the tonic component although detected, its quality and use in the study, was still questionable. Considering this, the detailed analysis was very important to understand properly the context of these events and how to proceed.

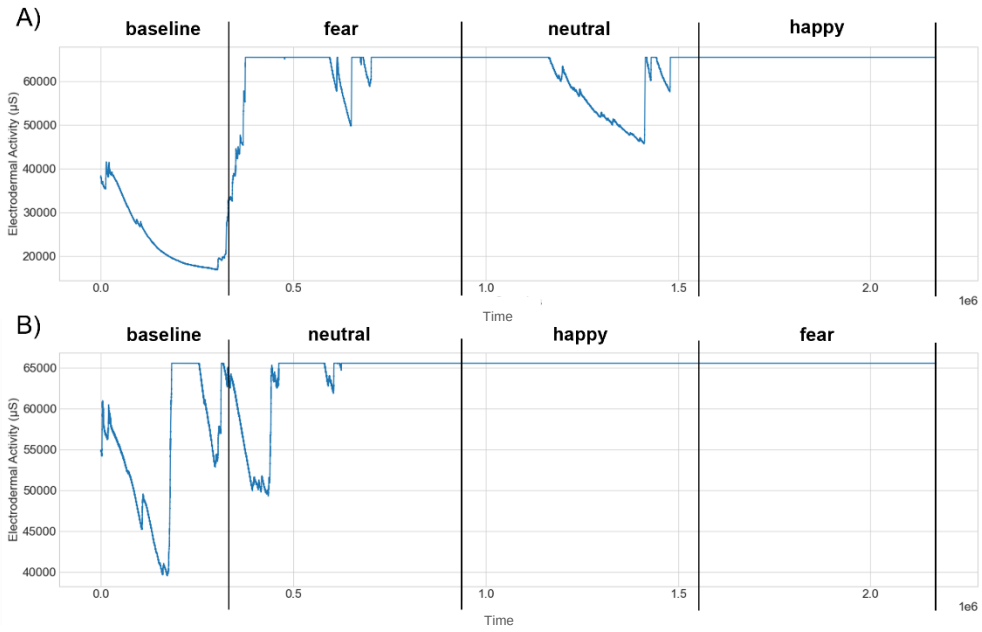

Figure 18 *- Electrodermal activity for the two participants that presented NaN values in EDA features.*

In the signal represented in Figure 18.A), corresponding to one of the participants, can be seen that the saturation is only present in the elicited emotional states and not in the baseline. Based on the participant's baseline, it is verified that there is no high electrodermal activity, only when the first emotional state is stimulated there is an increase in the activity, which indicates that this increase is mainly related to the emotional state and not to the individual physiological characteristics. It is also verified that in the transition from fear to the neutral state, there is a decrease in electrodermal activity, and the signal segment is not fully saturated, which can demonstrate that the emotional activation is lower when compared to the other emotional states.

According to the responses of the questionnaires of this participant, it was reported an arousal of 5 and 4, (on a scale of -5 to 5) for fear and happiness, respectively, and 1 for neutral, supporting that the emotional arousal is lower in this state and hence the decrease of the electrodermal activity. This demonstrates the correlation between electrodermal activity and emotional stimulation and, since

fear and happiness are classified with high arousal, the correlation of electrodermal activity with arousal is also demonstrated, as expected [77].

Regarding Figure 18.B), were it is shown the EDA signal of the second participant, a partial saturation is detected in the baseline, which may be related to the participant's physiological characteristics or due to a more active initial state. At the beginning of the neutral state, some activity variations are still detected, and the signal is partially saturated, but as soon as it stabilizes, the signal remains saturated for the rest of the session, for both fear and happy states. Thus, it was not possible to infer if the saturation derives from emotional activation, physiological conditions, or hardware problems.

According to the questionnaires' answers for this participant, it was reported an arousal of 3 for the initial state (before baseline), an arousal of 3 for fear, an arousal of 2 for happiness, and an arousal of 1 for neutral. Considering this information and the results of the STICSA-State questionnaire where the participant did not report sweating of the hands, that could justify the high electrodermal activity, the partial saturation of the signal is probably related to an active and overexcited initial state. Contrarily to participant A), the excited initial state of the participant B) seems to have a higher impact on the rest of the electrodermal response and is not as extensive since the participant reported lower levels of arousal (happiness and neutral).

The purpose of this analysis was to decide how to deal with NaN values. There were three options to consider: (1) eliminate the features that contained NaN values, (2) eliminate the participant or (3) replace the NaN values with values that fit the data. The elimination of features was discarded since the statistical analysis had not yet been performed and some of the features may be descriptive of emotions. The second option, eliminating the participants, was also discarded because both participants reported emotional stimulation, verified in questionnaires. Besides, this work relies on a multimodal approach, this information can be used associated with the rest of the physiological signals. Lastly, the replacement was considered the best approach to deal with NaN values because their number was considerably small and unlikely to have a great impact on the construction of the model. The features in question are the skewness of feature time series related to the phasic component of the signal. The phasic component was not detected by the feature extraction methodology, due to the saturation, so it creates a time series of equal values. Since skewness evaluates the symmetry and the values are all equal, the function to calculate it returns a NaN value by default, however, it may be replaced by zero, which corresponds to the case of symmetry.

Univariate Statistical Analysis

The main goal of the univariate statistical analysis was to explore if at least one of the extracted features was able to discriminate one condition. If that was the case for a relatively high number of features, then the step of feature extraction was validated. Besides that, this information can still be very useful so that in future work it will not be necessary to extract such a large number of features but focus on those that are more relevant and informative.

The results are resumed in Table 3, where from the 145 features, 86 features were able to distinguish at least one of the conditions, while 59 could not. Within the 86 features, 17 were from the EMG MF, 6 from the EMG TR, 18 from the EDA, and 45 from the ECG signal.

		EMG MF	EMG TR	EDA	ECG	Total
Null hypothesis	Rejected	17	6	18	45	86
	Not Rejected	9	20	15	15	59
Total		26	26	33	60	145

Table 3 *- Results of the univariate statistical analysis (RM ANOVA and Friedman) per physiological signal.*

The high number of features from the EDA and ECG to reject the null hypothesis agrees with the literature, which reports that these physiological signals had shown to describe significantly emotional states [78]. The same applies to EMG MF [79], where more than half of its features distinguish at least one condition, and for EMG TR, that although the number of features is smaller, seems to also contain relevant information.

However, when analysing the violin plots, it was detected that some of the featres that rejected the null hypothesis could only distinguish the Baseline from the rest of the emotional states. Moreover, with the post hoc analysis, this was confirmed, and 19 features that rejected the null hypothesis of the RM ANOVA could only distinguish the Baseline from the rest of the elicited conditions.

At this point, it was clear that the individual characteristics (recorded in the baseline) may influence the emotional response and raised the question if the previous emotional state could also have some influence. To tackle this issue, two approaches of data normalization were considered, one using the baseline and the other using the previous emotional state. Briefly, both normalizations allow to remove the individual influence of the analysis, and the second normalization additionally deals with residual influence of the previous emotional state.

The univariate statistical analysis was repeated for the two new feature data (normalized by baseline and by the previous emotional state). Then, the results with and without normalization were compared to set the approach to be considered for the rest of the study. The discussion of these results took in consideration the p-value and F-statistic of RM ANOVA and Friedman, and the p-value for each combination of the post hoc test.

The results of RM ANOVA and Friedman (Table 4) highlight that by normalizing, the number of features exhibiting statistical significance between conditions decreases. Considering this and the post hoc results mentioned above, the influence of the baseline was confirmed.

		Without normalization	**Baseline normalization**	**Previous emotion normalization**
Null hypothesis	Rejected	86	47	34
	Not Rejected	59	98	111

Table 4 *- Results of the univariate statistical analysis (RM ANOVA and Friedman, 5% significance level) for the three approaches: without normalization, baseline normalization, and previous emotion.*

As consequence, it was chosen to perform normalization, and the final emotion classification is less affected by the intersubject variability. Moreover, normalizing by the baseline is already a practice in some reviewed literature [80] and may increase the accuracy of the final classification.

Regarding performing normalization with baseline or previous emotion, the F-statistic was the considered metric to choose which is the most suitable as the higher the value, the greater the statistical evidence of the differences between conditions.

Both normalizations reduce the denominator of the F-statistic because both normalizations remove intersubject variability. However, only previous emotion normalization removes the individual variability to each emotion, i.e., intrasubject variability. Comparing the results of F-statistics (Table 5) it was found that, rejecting the null hypothesis, the number of times that the F-statistic of baseline normalization is higher than the F-statistic of previous emotion normalization (Fb > Fp) is 53, whereas the opposite (Fb < Fp) only occurs 28 times.

		F-statistic		Total
		Fb > Fp	Fb < Fp	
Reject Null Hypothesis	Baseline Normalization (b)	38	9	47
	Previous Emotion Normalization (p)	15	19	34
	Total	53	28	

Table 5 *- Results of the univariate statistical analysis (RM ANOVA and Friedman) for the two options of normalization, using the baseline (b) or the previous emotion (p).*

This result demonstrates that by removing the previous emotion influence, i. e. removing intrasubject variability, the differences between conditions become less noticeable, indicating that too much information may have been removed and that the impact of the previous emotion is not as large as the individual characteristics of the subject.

On this basis, the baseline normalization was performed and the influence of the individual characteristics and the initial emotional state of the participant was removed from the analysis. Note that the following results are from the analysis of baseline normalized data and the conditions fear, happiness, and neutral, are from now on normalized by baseline.

Table 6 shows that among the 47 features that rejected the null hypothesis: 2 belong to EMG MF, 0 to EMG TR, 12 to EDA, and 33 ECG. Once more, it appears that EDA and ECG are the physiological signals in which a higher number of features can discriminate at least one emotion from the others. For both EMG MF and EMG TR, the number of features has decreased when compared to the results presented in Table 3, meaning that these signals were mainly discriminating the baseline.

		EMG MF	EMG TR	EDA	ECG	Total
Null hypothesis	Rejected	2	0	12	33	47
	Not Rejected	24	26	21	27	98
	Total	26	26	33	60	145

Table 6 *- Results of the univariate statistical analysis (RM ANOVA and Friedman, 5% significance level) for the feature data normalized by the baseline, discriminating the physiological signals.*

The results of the post hoc test (Table 7) show that there is a higher number of features that can discriminate fear from happiness, and that some signals can discriminate better some combinations than others.

The ECG features mainly distinguish fear from the neutral state and happiness from the neutral state, which suggests that, in general, it is the neutral state that stands out from the rest. Still, some ECG features can distinguish fear from happiness, opposite emotions in the valence axis, pointing that from the ECG, it is possible to discriminate emotions through the valence dimension.

The EDA features mainly distinguish fear from happiness and, as in the ECG, it may distinguish emotions through the valence dimension. However, it is known that EDA is related to arousal [77], and that the level of arousal of each of the emotions (fear and happiness), despite being in the same quadrant, has different intensities. Thus, a more interesting conclusion can be drawn namely the

EDA features that distinguish fear from happiness may be able to discriminate the level of arousal even in the same arousal quadrant.

The features of both EMGs, when used separately, do not seem to distinguish emotions. However, as this work focuses on a multimodal approach, they may still have a relevant role when used together with the other physiological signals.

	EMG MF	EMG TR	EDA	ECG	Total
Fear - Happiness	1	0	9	8	18
Fear - Neutral	0	0	2	12	14
Happiness - Neutral	1	0	1	13	15

Table 7 *- Results of the post hoc test for the feature data normalized by the baseline, indicating the number of features from the physiological signals that can discriminate the combinations of emotions.*

Bivariate Statistical Analysis

The univariate analysis allowed the characterization of each feature independently. In the classification step, the features are used together, so their selection cannot be restricted to this analysis. The bivariate analysis allowed to reduce further the number of features by eliminating the redundant ones. In this work, a correlation analysis was carried out.

As represented in Figure 19, the heat map of the correlation matrix between all pairs of variables, shows along with the diagonal, blocks with a darker blue color evidencing high correlation between features of the same signal. The fact that they have a high correlation suggests that the variations of the features are similar for the entire signal and, therefore, represent the similar information. Furthermore, this analysis supports the use of a multimodal approach where the use of different signals allows obtaining different information (not correlated) that when used together may increase the effectiveness of the emotion classification.

The same analysis was made for each emotion independently. Through the heat maps in Figure 19, it was verified that for fear there is a greater correlation between features associated with different physiological signals. This indicates that when experiencing fear, the whole body reacts with greater intensity and, therefore, all physiological signals accuse such physiological changes. For happiness and neutral, the correlation between features of different physiological signals is lower, which may indicate that the body's physiological response is less pronounced across signals for these emotions.

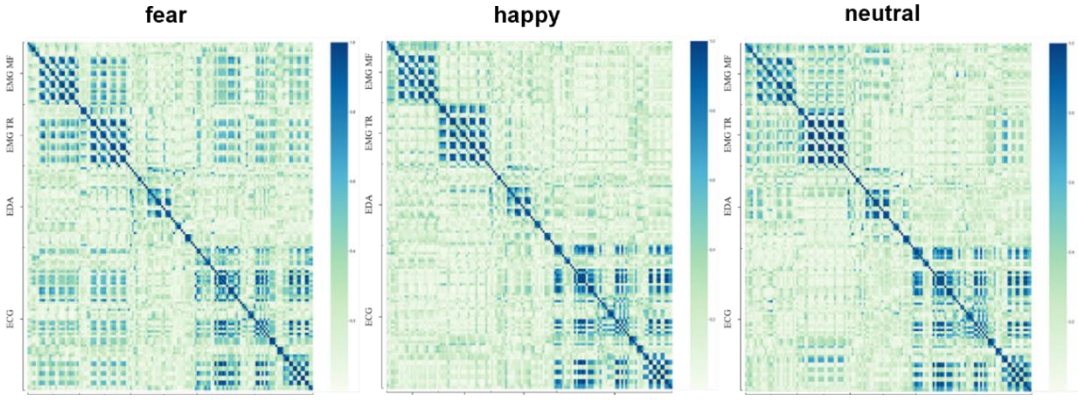

Figure 19 *- Heat maps for the three emotions representing the correlation between pairs of features.*

Regarding feature selection, following the selection criteria described in Section 4.3.3, 65 features remained, 13 from the EMG MF, 11 from the EMG TR, 18 from the EDA, and 23 from the ECG (Table 8). The number of features reduced to about half, however, there was a greater reduction in ECG features. Although more features were extracted from the ECG than from the other signals, when performing feature selection, the number of ECG features decreased, indicating that many of them are highly correlated, especially between the features extracted from the HRV.

Signal	Feature
EMG MF (EMG_MF_)	Activations_N; Duration_Mean; Duration_Var; Duration_Skew; Duration_Kurt; MaxPeakAct_Var; MaxPeakAct_Skew; MaxPeakAct_Kurt; MeanPeaksAct_Mean; all_Amplitude_Skew; all_Amplitude_Kurt; Area_Skew; Area_Kurt
EMG TR (EMG_TR_)	Activations_N; Duration_Std; Duration_Skew; Duration_Kurt; MaxPeakAct_Mean; MaxPeakAct_Skew; MeanPeaksAct_Kurt; all_Amplitude_Skew; all_Amplitude_Kurt; Area_Skew; Area_Kurt
EDA (EDA_)	Symp; SympN; Tonic_Mean; Tonic_Std; Tonic_Skew; Tonic_Kurt; Phasic_Mean; Phasic_Skew; Phasic_Kurt; SCR_Height_Kurt; SCR_Amplitude_Var; SCR_Amplitude_Kurt; SCR_RiseTime_Mean; SCR_RiseTime_Skew; SCR_RiseTime_Kurt; SCR_RecoveryTime_Mean; SCR_RecoveryTime_Skew; SCR_RecoveryTime_Kurt
ECG (ECG_)	Rate_Skew; Tduration_Mean; Tduration_Var; Tduration_Skew; Tduration_Kurt;
ECG (HRV_)	MedianNN; IQRNN; pNN50; pNN20; TINN; LF; VHF; HFn; LnHF; SD1SD2; CSI_Modified; PIP; PAS; GI; PI; Ca; ApEn; SampEn

Table 8 *- Remaining features after feature selection.*

5.2 Classification

Using the proposed emotional model, different emotional classifications were elaborated taking into account different data, different types of splitting data, and the two models of describing emotions (discrete and dimensional).

Two different approaches were considered to organize the data, as represented in Figure 20. In the Dataset1, each emotion is described by the feature through its value, and in the Dataset2, each emotion is described by a time series of the feature. The main difference between them is the window where the feature is calculated. As previously verified, most of the variables rejected the non stationarity, allowing the division the physiological signals into smaller windows and extract features for the partitions. For Dataset2, the 10 minutes windows were divided into 2 minutes, and each emotion has 5 feature values (one for each excerpt/partition).

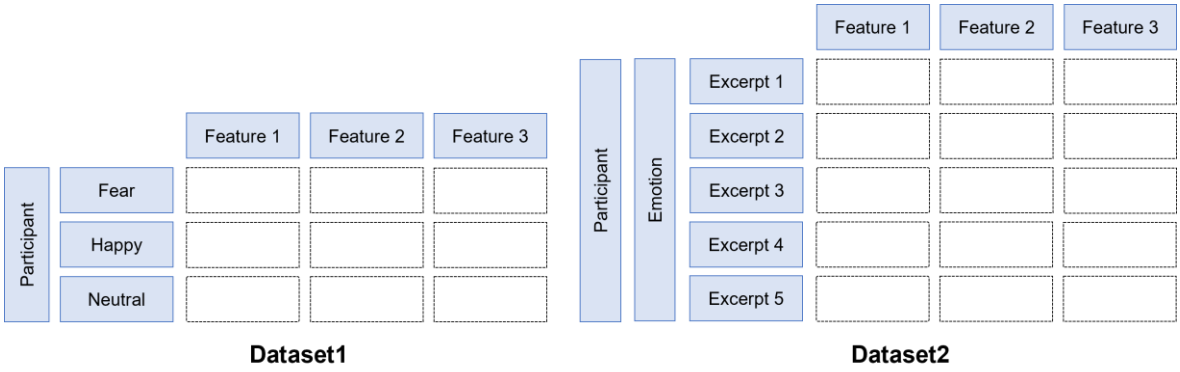

Figure 20 *- Organization of data for Dataset1 and Dataset2.*

To divide data into training and testing sets, it was not used any predefined splitting function because they randomly divide data, which does not allow controlling the source of the data. Three splitting data functions were created, denominated Split1, Split2, and Split3, and described as follows.

The **Split1** function divides data so that training and testing sets have all 30 participants. However, in the training set each participant is represented by two (of the three) emotions, and in the testing set by only one (the remaining emotion). For example, in training, if the participant is represented by fear and neutral, then in the testing is represented by happiness. It should be noted that the emotions for both sets are randomly chosen but guaranteeing that the final dataset is balanced.

In the **Split2** function, the training set is composed of 70% of the participants, and the testing set of 30% of the participants. The participants are not shared between sets, and in each set, the participants are represented by all emotions. This division allows to test the model in a subject-independent approach since the data to be tested are from new participants.

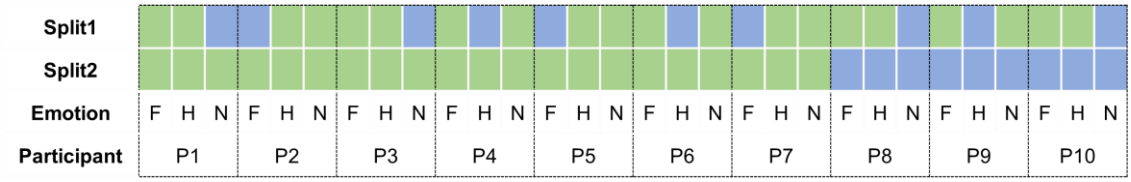

Figure 21 *- Scheme of splitting into training (green) and testing (blue) sets for Split1 and Split2 (schematic representation with 10 participants but applied to the 30 participants).*

The **Split3** function is exclusively for Dataset2. Since each emotion is represented by more than one value in each feature, it is possible to train the model with all participants' emotions. Thus, the training set is composed of all participants, and each emotion is represented by 4 (out of 5) excerpts of each feature. The testing set is also composed of all participants, but each emotion is represented by the remaining excerpt. Note that the excerpts are not repeated between sets and are randomly chosen.

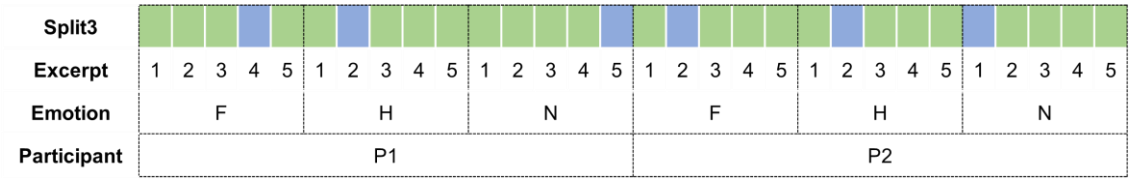

Figure 22 *- Scheme of splitting into training (green) and testing (blue) sets for Split3 (schematic representation with 2 participants but applied to the 30 participants).*

Emotions may be described by discrete or dimensional spaces; in the scope of this work both were explored. In the first approach, the discrete model was studied to identify the elicited emotions: fear, happiness, and neutral. Notwithstanding, the dimensional description can better describe the physiological response to emotions so, this approach was also studied for the dimensions valence and arousal.

The following subsections are divided into the discrete and dimensional emotional models and present the results of feature and hyperparameters optimization, and the prediction results for each dataset and each splitting function. The features belonging to each optimal subset are not presented along with the analysis of the optimal number, but can be accessed online[3].

The metrics used to evaluate the performance of the classification were the accuracy and F1-score. The accuracy measures the total correctness of the classifier based on the number of correct predictions over total instances; the F1-score represents the harmonic mean between recall and precision [81].To avoid bias to the training set, the evaluation metrics are a global average of 30 iterations of cross-validation.

[3] https://github.com/carolinafalves/Repository_CarolinaAlves.git

5.2.1 Discrete Emotional Model

Dataset1

From the backward elimination, the evolution of accuracy with the number of features is represented in Figure 23. The accuracy varies between 0.34 and 0.50, with a distinct peak in the accuracy of 0.49, for the optimal subset of 19 features. Between 16 and 22, the model performance is the highest, followed by a decrease between 25 and 30 features. Both for less than 15 to 5 and greater than 30 to 53, the accuracy varies between 0.40 and 0.44, stabilizing from 53 at approximately 0.42.

For Dataset1, it was clear that the ideal number of features varies between 16 and 22, being 19 the optimal number. The final subset of the optimal features was composed of features from all of the acquired physiological signals. As expected, the use of information from different signals optimized the emotional classification. Although all signals are represented, the ECG signal has the largest number of features, followed by the EDA features, which is in agreement with what was seen in Section 5.1. Contrary to the results presented in Table 7, where the features of both EMG seemed not to distinguish emotions and for that reason could be considered that these signals can not be used in emotion recognition, in the optimal feature set, some features were selected. This shows that the combination of these signals with the other ones improves the performance of the model and that they are relevant in emotion identification studies.

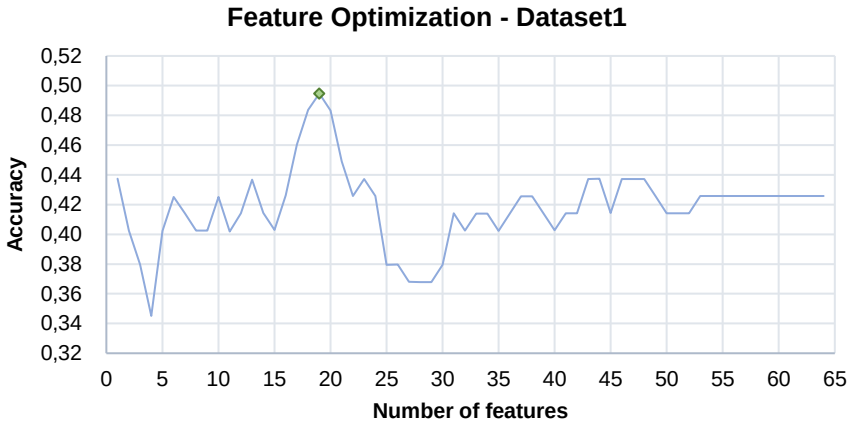

Figure 23 *- Evolution of accuracy with the number of features using backward elimination for Dataset1.*

With the best features selected, the emotional model with grid searching was applied to Dataset1 using the two splitting functions, Split1 and Split2. In addition to discovering the best combination of hyperparameters, the goal was to analyze how the model behaves in the subject-dependent and independent approach. Table 9 displays the results of this analysis for the training and testing of the classifier. Besides analyzing the final accuracy, the F1-score of each emotion was used to understand if there is any bias to the classification label and if so which one.

Comparing the classifier's training and testing in Table 9, the results for the training are considerably higher than those of the test, as expected. The accuracy difference is accentuated for Split1 where the participants are not represented by all emotions but only by two, suggesting that identifying a new emotion from an already known participant is difficult for the classifier.

According to the final accuracy of the two approaches, Split2 is superior to Split1, indicating that the classifier can extrapolate the emotional evolution of one participant to another better than discovering a new emotion from a known participant. Thus, the subject-independent approach for this dataset is

the one with the best performance. Although the performance for both splits is lower than 50%, a random choice of the classifier corresponds to 33%, which indicates that the classifier has learned something and outperforms the random results (the worst scenario).

Regarding the F1-score, in both, the identification of fear is always superior to happiness and neutral, and happiness is the one that the classifier has the greatest difficulty in identifying. Previously, in the post-hoc analysis, the Fear vs Happiness combination was the one with the highest number of features able of distinguishing them. However, when using three emotions there is a greater difficulty in distinguishing happiness. A possible explanation for this is that it was found that, for the neutral state, some of the participants report a higher level of valence in the questionnaire and a higher percentage of happiness, which can cause these two emotional states to be closer. This and the fact that happiness has high arousal like fear can difficult the discrimination of this emotion by the classifier, and for that reason, the F1-score is lower.

Dataset2

In Figure 24, the results of the feature optimization for Dataset2 are presented. The evolution of accuracy for this dataset is quite different from the previous one with a more accentuated increase in the beginning until 10 features, followed by a decrease between 10 and 20, and again an increase in the accuracy that then starts to vary in a smaller range between 0.44 and 0.48. The accuracy varies between 0.32 and 0.48 and is notable that from 20 features on, there are two regions with very similar accuracy, one for 21 and 22 features, and the other between the 45 and 47 features. However, the maximum accuracy is for the subset of 46 features with an accuracy of 0.475.

Relating to the results of Dataset1 where the highest score was between 16 and 22 features and considering that for Dataset2 there is also a peak in this zone, it appears that the ideal number of features for the emotional model round values around 20 features. Despite this finding, the function used only allows knowing the optimal set of features and not which features optimize other subsets, like the subsets of 21 or 22 features. Thus, it was considered 46 features as the optimal subset, also from all signals.

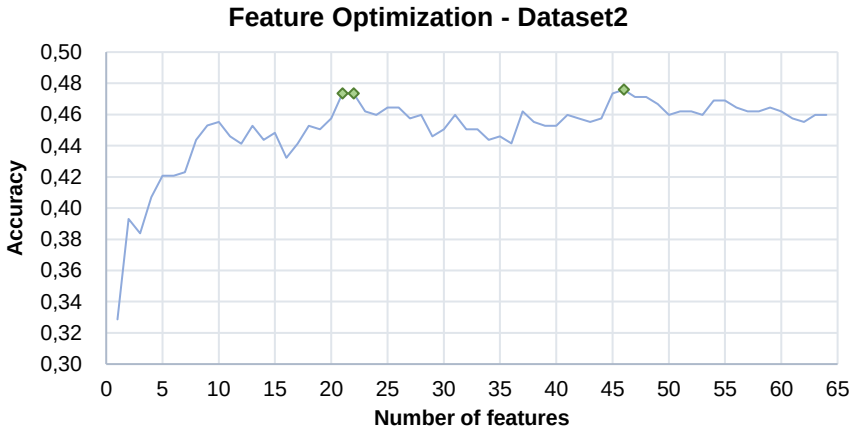

Figure 24 *- Evolution of accuracy with the number of features using backward elimination for Dataset2.*

Regarding the results of Dataset2 in Table 9, once again, it was verified that in the training test, the evaluation metrics present higher values, being the F1-score of happiness the lowest. For this dataset, the classifier knows all participants and all emotional states, consequently, the test results are superior to those of Dataset1, with a final accuracy of 60.4%.

The final accuracy of the classifier for Dataset1 is below 50% and for the Dataset2 is above. Given the difference between these datasets, that is the number of samples, these results point out that the dimension of data and amount of information may have impact on the ability of the classifier to learn and correctly identify the emotion.

Dataset	Splitting function	Accuracy ± Standard Deviation		F1-Score					
		Train	Test	Train			Test		
				F	H	N	F	H	N
Dataset1	Split1	0.810 ± 0.118	0.369 ± 0.097	0.812	0.790	0.822	0.435	0.288	0.361
	Split2	0.770 ± 0.105	0.489 ± 0.076	0.805	0.725	0.774	0.510	0.428	0.491
Dataset2	Split3	0.813 ± 0.137	0.604 ± 0.046	0.827	0.807	0.805	0.631	0.51	0.607

Table 9 *- Results of the emotional model classification for Dataset1 and Dataset2.*

Optimized Model

The optimized model is the emotional model used in the previous classification but with the optimal features and hyperparameters defined in the model itself. Instead of the model trying to find the optimal features and hyperparameters, with the backward elimination and grid searching, each time it performs a classification, in the optimized model, this is already defined.

Dataset2 showed better classification results (Table 9), so in the optimized model, this dataset was the used one. From the backward elimination to this dataset, the optimal number of features was 46, as discussed previously. Regarding the optimal hyperparameters, the grid searching was performed several times and saved all the choices of this method. As result, there was a table that contained all the combinations selected from the repeated grid searching, and to select the best combination, was analyzed the combination that appeared the most but also that had the higher accuracy. From this analysis was found that the optimal hyperparameters were the combination of the kernel function 'rbf' with C=10 and gamma=1.

From the analysis of F1-Score in Table 9, regarding the differences between emotions, the idea of applying the emotional model in a binary perspective came up. With this, it was intended to understand which emotions the classifier can identify and distinguish better, comparing only two by two. Thus, with the optimized model defined, the two types of classification were performed: multiclass and binary.

Regarding the results of the multiclass classification in Table 10, there was a slight increase in the final test accuracy from 60.4% to 66.7% and a slight decrease in the standard deviation. The final accuracy increasing, is the result of fixing the optimal hyperparameters and highlights the impact of each combination in the SVM classifier. Regarding the F1-score of each emotional state, it was verified that the metric is more similar to the rest but still it seems that there is a bias to the fear emotional state. This bias may be related to the previously identified fact that the neutral state response is closer to happiness.

Among the binary classifications (Table 10), the one with the best result is Fear vs Neutral, followed by Fear vs Happiness, and finally, Happiness vs Neutral, which is according to what was expected. Fear and neutral, in terms of valence and arousal are quite different. While fear has negative valence and positive arousal, neutral rounds 0 and 1, which makes these two emotional states very distinct, and the classifier has the better performance. Regarding Fear and Happiness, in terms of valence, they are opposites, however in terms of arousal both are positive, hence the classifier has greater difficulty in distinguishing the two emotional states. Finally, Happiness and Neutral have the lowest

accuracy, which may be, once again, related to the biasing from the Neutral to a more pleasurable state, difficulting the classification.

Summing up, when the classifier is applied to Fear-Neutral that have the two dimensions distinct, it is has a higher performance, but when applied to Fear-Happiness, which are only distinct in valence, and to Happiness-Neutral, that in this case are mainly distinct in arousal, the classifier has the lowest performance. Analyzing this, if the classifier is applied considering the description of emotions in the dimensions valence and arousal, then the performance may be higher. This was the starting point for using a dimensional-based emotional model described in the next section.

Classification	**Accuracy ± Standard Deviation**		**F1-Score**					
	Train	Test	Train			Test		
			F	H	N	F	H	N
Multiclass	0.997 ± 0.002	0.667 ± 0.044	0.999	0.997	0.996	0.712	0.635	0.656
Fear vs Happiness	1.000 ± 0.001	0.774 ± 0.047	1.000	1.000	-	0.766	0.779	-
Fear vs Neutral	0.999 ± 0.002	0.812 ± 0.047	0.999	-	0.999	0.811	-	0.812
Happiness vs Neutral	0.997 ± 0.003	0.749 ± 0.050	-	0.997	0.997	-	0.751	0.747

Table 10 *- Results of the optimized emotional model classification for multiclass and binary classification.*

5.2.2 Dimensional Emotional Model

In the previously binary analysis, the justification for the results of each of the combinations was based on the dimensional characteristics of each emotional state. Verifying that it was possible from the dimensions to support the results, indicated that the emotional description based on dimensions can improve the classification of the proposed emotional model.

This subsection is divided into the two considered dimensions, valence and arousal, and the classification of emotions through dimensions is performed independently. The dataset selected is the Dataset2, which proved to lead to better results, and the splitting function is the Split3.

In this model, the label to identify is not the discrete emotion but the corresponding quadrant for each dimension. For valence, fear is described as negative whereas happiness and neutral are positive. For arousal, fear and happiness are described as high and the neutral state as low. The attribution of the qualitative value was based on the proposed model by Russell, except for the neutral state in which the responses to the questionnaires were considered.

Valence

Figure 25 shows the evolution of accuracy with the number of features for each subset. It is noticeable that the range in which accuracy varies is much lower than the previous results, ranging only between 0.65 and 0.70. Therefore, several feature subsets provide the model with a good performance. However, the optimal number of features is 57 with an accuracy of 0.692, also from all signals.

As for the optimized discrete model, it was necessary to find out which combination of hyperparameters best fits the training data. As a result of grid searching, the best combination is composed by the kernel function ‘rbf’, C=1000, and gamma=0.1.

The valence model is a binary model to determine whether each sample belongs to the positive or negative component of the valence. However, Dataset2 for the negative component has data that

belongs to fear, and for the positive component that belongs to happiness and neutral state data, which makes the data classes unbalanced. Thus, starting from Dataset2 a new dataset (with data balancing) was created using data balancing methods. In particular, in this work, was used the imblearn.over_sampling.RandomOverSampler function from the imbalanced learn package [82], which performs random over-sampling to the minority class.

The model was applied to the dataset 'Without Data Balancing' and 'With Data Balancing', and the results are shown in Table 11. Comparing the final accuracy of the two datasets, it was found that the best results belong to the classification of the unbalanced dataset. However, when analyzing the results of the F1-score, it appears that for the minority class, which is the negative, the best score belongs to the balanced dataset, as expected, since the model has more samples to train. The F1-score of the positive class is lower for the second dataset, which may be related to the method of data balancing. Although data balancing increases the model's performance to classify the minority class, the method can be improved so that there is no decrease in performance relative to the majority class. One way to overcome this issue would be to acquire more real data and not originate by data augmentation techniques.

When comparing these results with those obtained by the discrete model for the multiclass classification, it is verified that with the dimensional approach the performance of the models is greater and, therefore, the model can more effectively determine the emotional state. Despite not detecting the specific emotion with this valence model, it is possible to determine if the participant is in a positive and negative state, which allows the emotional state characterization.

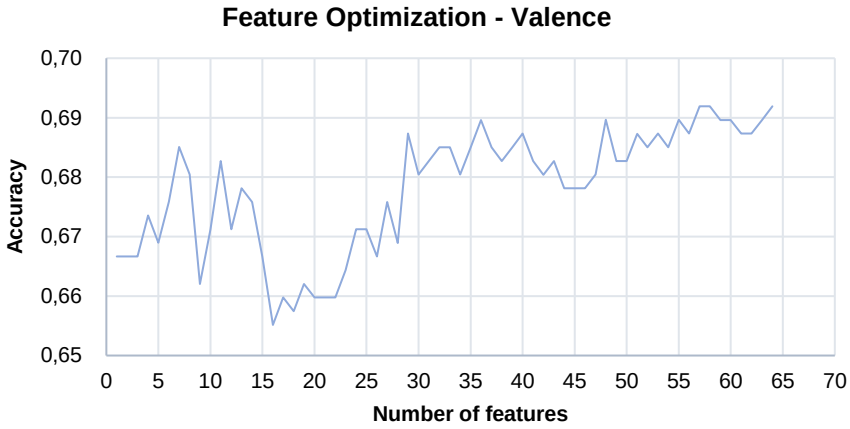

Figure 25 *- Evolution of accuracy with the number of features using backward elimination for Dataset2 valence-based.*

Classification	Accuracy ± Standard Deviation		F1-Score			
	Train	Test	Train		Test	
			Positive	Negative	Positive	Negative
Without Data Balancing	1.000 ± 0.000	0.776 ± 0.036	1.000	1.000	0.833	0.657
With Data Balancing	1.000 ± 0.001	0.737 ± 0.049	1.000	1.000	0.761	0.705

Table 11 *- Results of the valence emotional model classification.*

Arousal

Figure 26 shows the results of the feature optimization process for arousal. Contrary to the results obtained so far, the model is optimized by only the feature 'EMG_TR_MaxPeakAct_Mean' and the score is the highest until the optimal number of 8 features. From the 9 features onwards, the score varies approximately between 0.62 and 0.64, with only one peak in the subset of 26 features. Although the model determines that the best subset was composed of only one feature, considering only this feature can lead to data bias. To overcome that problem was determined that the minimal number was 10 features, and the optimal subset of features shifted to 26 features belonging to all physiological signals. Regarding the gridsearching of the hyperparameters, it was determined that the optimal combination was composed by the kernel function 'linear' and C=10.

As for the valence model in which data were unbalanced, the same was verified for the arousal model in which high arousal was related to the fear and happiness data, and the low arousal was related to the data of the neutral state. Using the same method described for the valence model, a new dataset with data balancing was obtained, in which the two classes were balanced.

With the optimized model defined, were obtained the results in Table 12, regarding the arousal classification for the datasets with and without data balancing. As in the previous valence model, it was verified that the final accuracy of the model decreases to the balanced data, that the minority class F1-score increased for the balanced dataset and that the majority class F1-score decreased for the same dataset.

When comparing the final accuracy of both dimensions models, the arousal model has a slightly lower score, yet the results are superior to those obtained for the discrete model. Again, it was highlighted that the description of emotions by dimensions is more reliable.

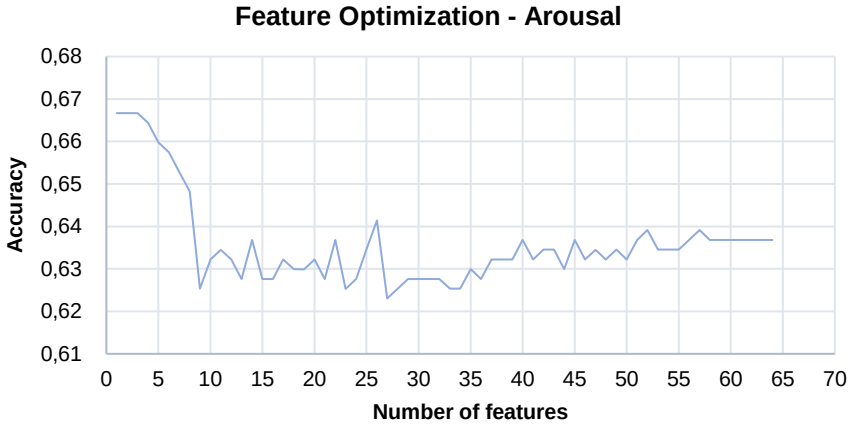

Figure 26 *- Evolution of accuracy with the number of features using backward elimination for Dataset2 arousal-based.*

Classification	Accuracy ± Standard Deviation		F1-Score			
	Train	Test	Train		Test	
			High	Low	High	Low
Without Data Balancing	0.943 ± 0.006	0.756 ± 0.042	0.958	0.908	0.821	0.613
With Data Balancing	0.958 ± 0.008	0.699 ± 0.061	0.957	0.958	0.721	0.669

Table 12 *- Results of the arousal emotional model classification.*

6. Conclusion

Emotion recognition systems are a very useful tool for people's lives, capable of increasing their well-being and health. Their variety of applications, with a particular interest in the area of education, psychology, and medicine, prove that research in this area is of great importance and that there is still much to be done.

With this work, it was intended, and it was possible to draw conclusions on the different components that constitute the emotion recognition systems, and thus, take a step forward in what is the study of emotion recognition.

Recalling the first objective of this work, it was created a complete and descriptive experimental protocol of what could be one of the procedures for collecting physiological signals using movies as a method of stimulation. Several types of analyses were considered, such as emotional transitions and response variations for the same stimulus on different days, and it was designed an optimized protocol, with which a varied and quality database was created. From the execution of the protocol in participants, it was concluded that it is adequate to its purpose and that it can be reproduced, for example, by other students or researchers in the area. Through the analysis of the participants' signals and responses, it was concluded that the movie excerpts were effective in stimulating the emotional states of fear and happiness. However, the excerpts to stimulate the neutral state were less effective, with participants reporting emotional responses closer to happiness.

In this work, the objective was not to deepen the impact of each of the excerpts used. However, as future work, an analysis can be carried out to understand which excerpts from the documentaries provoked this unwanted physiological response. For example, instead of processing the signals for each set of 10 minutes of stimulus, based on the duration of each excerpt, it is possible to divide the signals and compare the results of each with the results of happiness, and thus, understand which excerpts should (or not) be used.

In the exploratory data analysis, before concluding about the emotional description in physiological signals and respective features, the normalization of data in the three aspects proposed (without normalization and normalization by baseline or previous emotion) was studied. The results indicated that the normalization should be done since the initial state (acquired in the baseline) proved to have an impact, with features that only distinguished the baseline from the other emotions. Then, with the analysis between the two normalizations, it was found that by removing the influence of the previous emotion, i.e. the intrasubject variability, the differences between emotions were less noticeable, suggesting that a lot of information was being removed. Thus, the baseline normalization for this set of data is the most adequate, allowing to remove the impact of the initial state, i.e. the intersubject variability.

With the normalized data, it was then possible to complete the second objective of the work, to understand the emotional description in the various physiological signals and respective features. The signal that was shown to better describe emotions, by the univariate analysis, was the ECG, followed by the EDA. Regarding the two EMGs, according to the results, the extracted features did not seem to include information capable of distinguishing emotions. Afterward, in the emotional identification, the classifiers considered these features to belong to the optimal subset, allowing to conclude that the information from the EMGs complements the ECG and EDA information and, thus, they are relevant in emotion recognition studies.

As for the discrimination of pairs of emotions, evaluated by the results obtained in the post hoc test, once again, the ECG was highlighted, in which a greater number of features were able to distinguish the pairs of emotions (F-H; F-N; H-N). Among all the combinations, the ECG features mostly

discriminated fear from neutral and happiness from neutral, emotions with opposite arousal (fear and happiness have high arousal and neutral has low arousal), suggesting that the ECG can be used to distinguish emotions by dimension arousal. Still, some ECG features distinguished fear from happiness and for this reason, this signal can also be used to discriminate emotions by the valence dimension. These results demonstrated that the ECG features are relevant in the identification of emotions through the dimensional approach.

The last two objectives of this work are related to the optimization of classifiers based on machine learning techniques, and subsequent discussion between the two approaches that can be used to describe emotions (discrete and dimensional). In both approaches, the optimal features selected by the classifier are from all physiological signals, demonstrating once again that the use of more than one signal (multimodal approach) is beneficial for learning the classifier. It was also found that a larger dataset increases the performance of the classifier, and as a note for future work, it is suggested to collect more data using the proposed protocol, so that the EMOTE database is constituted by a greater quantity and variability of information, and thus increase the classifier's ability to identify emotions.

The use of an optimized model in the discrete approach caused an increase in test performance from 60.5% to 66.7%, highlighting that the definition of input parameters (hyperparameters) and features has a very important role in the construction of the classifier and that is why these steps are also crucial to proper emotional identification. Also, with this model, the binary approach was discussed from which it was reached the very important conclusion that the performance differences for each of the binary pairs (Fear-Happiness, Fear-Neural, and Happiness-Neutral) can be explained by the valence and arousal dimensions. When emotions are described by dimensions and there are notable differences in valence and arousal, which is the case of Fear-Neutral, the classifier performs better; when the differences are only noticeable in one of the dimensions, which is the case of Happiness-Neutral (for the data used) and Happiness-Fear, the classifier has a lower performance. From this, it was deduced that using the dimensional description of emotions, the classifier would present a superior performance, which was verified by the results obtained using the dimensional approach.

The two classifiers built in the dimensional approach, based on valence and arousal, achieved an accuracy of 77.6% and 75.6%, respectively. When compared with the result of the discrete multiclass classification, the results were superior, approaching even the results of the discrete binary classification, which only discriminated two emotions. With these classifiers, it is possible to identify the quadrant in the valence-arousal dimensional space and despite not identifying the specific emotion, they are able to describe the emotional state in its components.

Once the purpose of these emotion recognition systems is to be used in real-life contexts that can thrive on them, it is also important to realize the applicability of the classification. Recalling the set of diseases highlighted in this work, mental disorders, the description of a more negative or positive emotional state vs excited or calm, can be more relevant and applicable than a system that identifies the specific emotion. For example, in the case of depression, where patients may feel sad, bored, afraid, unwilling, a valence-based classifier may be more appropriate as the aim is to detect the prevalence of a negative and not the specific emotion. While an arousal-based classification can be useful, for example, in the classroom context, to detect if the students are more excited or calm, and consequently adapt the methodologies to be more appealing. These are just two of the possible applications of the dimensional-based emotional model developed in this work, but which prove that emotion recognition systems are crucial to well-being and an emotionally healthy life.

References

[1] C. for D. C. and Prevention, “Mental Health and Coping During COVID-19 | CDC.” https://www.cdc.gov/coronavirus/2019-ncov/daily-life-coping/managing-stress-anxiety.html (accessed Jun. 29, 2020).

[2] L. Shu *et al.*, “A review of emotion recognition using physiological signals,” *Sensors (Switzerland)*, vol. 18, no. 7, 2018, doi: 10.3390/s18072074.

[3] H. M. de A. P. de Carvalho, “Emoções: validação ecológica de estímulos-vídeo,” Universidade da Beira Interior, 2017.

[4] R. W. Picard, *Affective Computing*. Cambridge: MIT Press, 2000.

[5] S. D. Kreibig, “Autonomic nervous system activity in emotion: A review,” *Biol. Psychol.*, vol. 84, no. 3, pp. 394–421, 2010, doi: 10.1016/j.biopsycho.2010.03.010.

[6] G. Amaral *et al.*, “The Handbook of Psychophysiological,” *J. Petrol.*, vol. 369, no. 1, pp. 1689–1699, 2013, doi: 10.1017/CBO9781107415324.004.

[7] C. Liu, F. Liu, L. Zhang, Y. Su, and A. Murray, “Smart Wearables in Healthcare: Signal Processing, Device Development, and Clinical Applications,” *J. Healthc. Eng.*, vol. 2018, pp. 1–3, 2018, doi: 10.1155/2018/1696924.

[8] L. Shu *et al.*, “Wearable emotion recognition using heart rate data from a smart bracelet,” *Sensors (Switzerland)*, vol. 20, no. 3, 2020, doi: 10.3390/s20030718.

[9] C. Torres-Valencia, M. Álvarez-López, and Á. Orozco-Gutiérrez, “SVM-based feature selection methods for emotion recognition from multimodal data,” *J. Multimodal User Interfaces*, vol. 11, no. 1, pp. 9–23, 2017, doi: 10.1007/s12193-016-0222-y.

[10] World Health Organization, “Depression.” https://www.who.int/health-topics/depression#tab=tab_1 (accessed Jul. 06, 2020).

[11] World Health Organization, “Autism spectrum disorders.” https://www.who.int/news-room/fact-sheets/detail/autism-spectrum-disorders (accessed Jul. 06, 2020).

[12] U. Nations, “Take Action for the Sustainable Development Goals.” https://www.un.org/sustainabledevelopment/sustainable-development-goals/ (accessed Oct. 09, 2021).

[13] E. Walsh and P. E. Griffiths, “Emotion, Evolution of,” *Int. Encycl. Soc. Behav. Sci. Second Ed.*, vol. 7, pp. 467–471, 2015, doi: 10.1016/B978-0-08-097086-8.81031-4.

[14] A. Keil and V. Miskovic, “Human Emotions: A Conceptual Overview,” in *Sleep and affect: Assessment, theory, and clinical implications*, Elsevier Inc., 2015, pp. 23–44.

[15] G. J. Norman, E. Necka, and G. G. Berntson, “The Psychophysiology of Emotions,” in *Emotion Measurement*, Elsevier Ltd, 2016, pp. 84–98.

[16] S. Hamann, “Mapping discrete and dimensional emotions onto the brain: Controversies and consensus,” *Trends Cogn. Sci.*, vol. 16, no. 9, pp. 458–466, 2012, doi: 10.1016/j.tics.2012.07.006.

[17] T. Eerola and J. K. Vuoskoski, “A comparison of the discrete and dimensional models of emotion in music,” *Psychol. Music*, vol. 39, no. 1, pp. 18–49, 2011, doi: 10.1177/0305735610362821.

[18] I. B. Mauss and M. D. Robinson, “Measures of emotion: A review,” *Cogn. Emot.*, vol. 23, no. 2, pp. 209–237, 2009, doi: 10.1080/02699930802204677.

[19] A. Kolakowska, A. Landowska, M. Szwoch, W. Szwoch, and M. R. Wrobel, “Emotion recognition and its application in software engineering,” *2013 6th Int. Conf. Hum. Syst. Interact. HSI 2013*, no. June 2015, pp. 532–539, 2013, doi: 10.1109/HSI.2013.6577877.

[20] J. Cai, G. Liu, and M. Hao, “The research on emotion recognition from ECG signal,” *Proc. - 2009 Int. Conf. Inf. Technol. Comput. Sci. ITCS 2009*, vol. 1, pp. 497–500, 2009, doi: 10.1109/ITCS.2009.108.

[21] A. Dzedzickis, A. Kaklauskas, and V. Bucinskas, “Human emotion recognition: Review of sensors and methods,” *Sensors (Switzerland)*, vol. 20, no. 3, 2020, doi: 10.3390/s20030592.

[22] J. A. Domínguez-Jiménez, K. C. Campo-Landines, J. C. Martínez-Santos, E. J. Delahoz, and S. H. Contreras-Ortiz, “A machine learning model for emotion recognition from physiological signals,”

Biomed. Signal Process. Control, vol. 55, p. 101646, 2020, doi: 10.1016/j.bspc.2019.101646.

[23] C. Wei, L. lan Chen, Z. zhen Song, X. guang Lou, and D. dong Li, “EEG-based emotion recognition using simple recurrent units network and ensemble learning,” *Biomed. Signal Process. Control*, vol. 58, p. 101756, 2020, doi: 10.1016/j.bspc.2019.101756.

[24] J. Zhang, M. Chen, S. Zhao, S. Hu, Z. Shi, and Y. Cao, “ReliefF-based EEG sensor selection methods for emotion recognition,” *Sensors (Switzerland)*, vol. 16, no. 10, pp. 1–15, 2016, doi: 10.3390/s16101558.

[25] J. Chen, D. Jiang, Y. Zhang, and P. Zhang, “Emotion recognition from spatiotemporal EEG representations with hybrid convolutional recurrent neural networks via wearable multi-channel headset,” *Comput. Commun.*, vol. 154, no. 6, pp. 58–65, 2020, doi: 10.1016/j.comcom.2020.02.051.

[26] Y. Zhang, X. Ji, and S. Zhang, “An approach to EEG-based emotion recognition using combined feature extraction method,” *Neurosci. Lett.*, vol. 633, pp. 152–157, 2016, doi: 10.1016/j.neulet.2016.09.037.

[27] P. Pandey and K. R. Seeja, “Subject independent emotion recognition from EEG using VMD and deep learning,” *J. King Saud Univ. - Comput. Inf. Sci.*, no. xxxx, 2019, doi: 10.1016/j.jksuci.2019.11.003.

[28] H. W. Guo, Y. S. Huang, J. C. Chien, K. Haraikawa, and J. S. Shieh, “Short-term analysis of heart rate variability for emotion recognition via a wearable ECG device,” *ICIIBMS 2015 - Int. Conf. Intell. Informatics Biomed. Sci.*, pp. 262–265, 2016, doi: 10.1109/ICIIBMS.2015.7439542.

[29] K. N. Minhad, S. H. M. Ali, and M. B. I. Reaz, “Happy-anger emotions classifications from electrocardiogram signal for automobile driving safety and awareness,” *J. Transp. Heal.*, vol. 7, no. xxxx, pp. 75–89, 2017, doi: 10.1016/j.jth.2017.11.001.

[30] A. Goshvarpour, A. Abbasi, and A. Goshvarpour, “An accurate emotion recognition system using ECG and GSR signals and matching pursuit method,” *Biomed. J.*, vol. 40, no. 6, pp. 355–368, 2017, doi: 10.1016/j.bj.2017.11.001.

[31] H. Feng, H. M. Golshan, and M. H. Mahoor, “A wavelet-based approach to emotion classification using EDA signals,” *Expert Syst. Appl.*, vol. 112, pp. 77–86, 2018, doi: 10.1016/j.eswa.2018.06.014.

[32] V. Sharma, N. R. Prakash, and P. Kalra, “Audio-video emotional response mapping based upon Electrodermal Activity,” *Biomed. Signal Process. Control*, vol. 47, pp. 324–333, 2019, doi: 10.1016/j.bspc.2018.08.024.

[33] N. Ganapathy, Y. R. Veeranki, and R. Swaminathan, “Convolutional neural network based emotion classification using electrodermal activity signals and time-frequency features,” *Expert Syst. Appl.*, vol. 159, p. 113571, 2020, doi: 10.1016/j.eswa.2020.113571.

[34] Q. Zhang, X. Chen, Q. Zhan, T. Yang, and S. Xia, “Respiration-based emotion recognition with deep learning,” *Comput. Ind.*, vol. 92–93, pp. 84–90, 2017, doi: 10.1016/j.compind.2017.04.005.

[35] D. Kukolja, S. Popović, M. Horvat, B. Kovač, and K. Ćosić, “Comparative analysis of emotion estimation methods based on physiological measurements for real-time applications,” *Int. J. Hum. Comput. Stud.*, vol. 72, no. 10–11, pp. 717–727, 2014, doi: 10.1016/j.ijhcs.2014.05.006.

[36] S. Jerritta, M. Murugappan, K. Wan, and S. Yaacob, “Emotion recognition from facial EMG signals using higher order statistics and principal component analysis,” *J. Chinese Inst. Eng. Trans. Chinese Inst. Eng. A*, vol. 37, no. 3, pp. 385–394, 2014, doi: 10.1080/02533839.2013.799946.

[37] G. Yang and S. Yang, “Emotion recognition of electromyography based on support vector machine,” *3rd Int. Symp. Intell. Inf. Technol. Secur. Informatics, IITSI 2010*, no. 2, pp. 298–301, 2010, doi: 10.1109/IITSI.2010.122.

[38] S. Yang and G. Yang, “Emotion recognition of EMG based on improved L-M BP neural network and SVM,” *J. Softw.*, vol. 6, no. 8, pp. 1529–1536, 2011, doi: 10.4304/jsw.6.8.1529-1536.

[39] K. Gouizi, C. Maaoui, and F. Bereksi Reguig, “Negative emotion detection using EMG signal,” *Proc. - 2014 Int. Conf. Control. Decis. Inf. Technol. CoDIT 2014*, pp. 690–695, 2014, doi: 10.1109/CoDIT.2014.6996980.

[40] D. Nikolova, P. Petkova, A. Manolova, and P. Georgieva, “ECG-based Emotion Recognition : Overview of Methods and Applications,” *ANNA ’18; Adv. Neural Networks Appl. 2018*, pp. 118–122, 2018.

[41] L. B. Krithika and G. G. Lakshmi Priya, “Student Emotion Recognition System (SERS) for e-learning Improvement Based on Learner Concentration Metric,” *Procedia Comput. Sci.*, vol. 85, no. Cms, pp.

767–776, 2016, doi: 10.1016/j.procs.2016.05.264.

[42] J. T. Kraiss, M. Peter, J. T. Moskowitz, and E. T. Bohlmeijer, “The relationship between emotion regulation and well-being in patients with mental disorders : A meta-analysis,” *Compr. Psychiatry*, vol. 102, p. 152189, 2020, doi: 10.1016/j.comppsych.2020.152189.

[43] C. for D. C. and Prevention, “What is Autism Spectrum Disorder? | CDC.” https://www.cdc.gov/ncbddd/autism/facts.html (accessed Jul. 06, 2020).

[44] J. Kory and S. D’Mello, “Affect Elicitation for Affective Computing,” *Oxford Handb. Affect. Comput.*, no. September 2018, pp. 371–383, 2015, doi: 10.1093/oxfordhb/9780199942237.013.001.

[45] K. H. Kim, S. W. Bang, and S. R. Kim, “Emotion recognition system using short-term monitoring of physiological signals,” *Med. Biol. Eng. Comput.*, vol. 42, no. 3, pp. 419–427, 2004, doi: 10.1007/BF02344719.

[46] C. B. and R. M. V. Liu, “Grounded emotions,” *2017 Seventh Int. Conf. Affect. Comput. Intell. Interact.*, pp. 477–483, 2017, doi: 10.1109/ACII.2017.8273642.

[47] S. Jerritta, M. Murugappan, R. Nagarajan, and K. Wan, “Physiological signals based human emotion recognition: A review,” *Proc. - 2011 IEEE 7th Int. Colloq. Signal Process. Its Appl. CSPA 2011*, pp. 410–415, 2011, doi: 10.1109/CSPA.2011.5759912.

[48] N. Jatupaiboon, S. Pan-Ngum, and P. Israsena, “Subject-dependent and subject-independent emotion classification using unimodal and multimodal physiological signals,” *J. Med. Imaging Heal. Informatics*, vol. 5, no. 5, pp. 1020–1027, 2015, doi: 10.1166/jmihi.2015.1490.

[49] N. Praceres, J. D. A. Parker, and G. J. Taylor, “Adaptaçao portuguesa da escala de Alexitimia de Toronto de 20 Itens (TAS-20),” *Rev. Ibero-Americana Diagnóstico e Avaliação Psicológica*, vol. 9, no. 1, pp. 9–21, 2000.

[50] A. S. Mendes, “STICSA: Análise psicométrica numa amostra de estudantes universitários portugueses,” Universidade de Aveiro, 2018.

[51] S. Breast, L. Access, S. Breast, L. Access, S. B. Access, and S. Breast, “biosignalsplux User Manual.” Lisboa, pp. 1–24, [Online]. Available: http://www.plux.info.

[52] S. A. Makowski, D., Pham, T., Lau, Z. J., Brammer, J. C., Lespinasse, F., Pham, H., Schölzel, C., & Chen, “NeuroKit2: A Python toolbox for neurophysiological signal processing,” *Behav. Res. Methods*, 2021, doi: 10.3758/s13428-020-01516-y.

[53] S. J. et al Harris, C.R., Millman, K.J., van der Walt, “Array programming with NumPy,” *Nat. 585*, pp. 357–362, 2020, doi: 10.1038/s41586-020-2649-2.

[54] T. E. et al Pauli Virtanen, Ralf Gommers, “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python,” *Nat. Methods*, vol. 17, pp. 261–272, 2020.

[55] Y.-W. C. and K. S. Lai, “Lag Order and Critical Values of the Augmented Dickey-Fuller Test,” *J. Bus. Econ. Stat.*, vol. 13, no. 3, pp. 277–280, 2014, doi: 10.2307/1392187.

[56] “Augmented Dickey-Fuller Test in Python.” https://www.hackdeploy.com/augmented-dickey-fuller-test-in-python/ (accessed Apr. 21, 2021).

[57] J. Seabold, S., & Perktold, “statsmodels: Econometric and statistical modeling with python,” 2010.

[58] J. Frank J. Massey, “The Kolmogorov-Smirnov Test for Goodness of Fit,” *J. Am. Stat. Assoc.*, vol. 46, pp. 68–78, 1951, doi: 10.2307/2280095.

[59] N. Mohd and Y. Bee, “Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests,” *J. Stat. Model. Anal.*, vol. 2, no. 1, pp. 13–14, 2011.

[60] M. L. Waskom, “seaborn: statistical data visualization,” *J. Open Source Softw.*, vol. 6, p. 3021, 2021, doi: 10.21105/joss.03021.

[61] R. Schumacker and S. Tomek, *Understanding Statistics Using R*, vol. 6, no. 1. Springer, New York, NY.

[62] S. Brandt, *Data analysis: Statistical and computational methods for scientists and engineers, fourth edition*, Fourth Edi. Springer International Publishing, 2014.

[63] Laerd statistics, “Repeated Measures ANOVA.” https://statistics.laerd.com/statistical-guides/repeated-measures-anova-statistical-guide.php (accessed Mar. 11, 2021).

[64] R. Vallat, “Pingouin: statistics in Python,” *J. Open Source Softw.*, vol. 3, p. 1026, 2018, doi:

https://doi.org/10.21105/joss.01026.

[65] Laerd statistics, “Friedman Test in SPSS Statistics.” https://statistics.laerd.com/spss-tutorials/friedman-test-using-spss-statistics.php (accessed Oct. 14, 2021).

[66] S. Kumar and I. Chong, “Correlation Analysis to Identify the Effective Data in Machine Learning: Prediction of Depressive Disorder and Emotion States,” *Int. J. Environ. Res. Public Health*, p. 2907, 2018, doi: 10.3390/ijerph15122907.

[67] Laerd statistics, “Pearson Product-Moment Correlation.” https://statistics.laerd.com/statistical-guides/pearson-correlation-coefficient-statistical-guide.php (accessed Mar. 14, 2021).

[68] The pandas development, “pandas-dev/pandas: Pandas,” 2020.

[69] J. Zhang, Z. Yin, P. Chen, and S. Nichele, “Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review,” *Inf. Fusion*, vol. 59, no. January, pp. 103–126, 2020, doi: 10.1016/j.inffus.2020.01.011.

[70] A. J. SMOLA and B. SCHOLKOPF, “A tutorial on support vector regression,” *Stat. Comput.*, vol. 14, pp. 199–222, 2004.

[71] V. Pedregosa, F. and Varoquaux, G. and Gramfort, A. and Michel, P. and Thirion, B. and Grisel, O. and Blondel, M. and Prettenhofer, A. and and Weiss, R. and Dubourg, V. and Vanderplas, J. and Passos, and E. Cournapeau, D. and Brucher, M. and Perrot, M. and Duchesnay, “Scikit-learn: Machine Learning in {P}ython,” *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[72] C.-J. Chang, Chih-Chung and Lin, “LIBSVM: A library for support vector machines,” *ACM Trans. Intell. Syst. Technol.*, vol. 2, pp. 1–27, 2011.

[73] “Support Vector Machines.” https://scikit-learn.org/stable/modules/svm.html#svm-classification (accessed May 23, 2021).

[74] B. Roy, “All about Feature Scaling,” *Towards Data Science*, 2020. https://towardsdatascience.com/all-about-feature-scaling-bcc0ad75cb35.

[75] S. S. Hameed, O. O. Petinrin, A. O. Hashi, and F. Saeed, “Filter-wrapper combination and embedded feature selection for gene expression data,” *Int. J. Adv. Soft Comput. its Appl.*, vol. 10, no. 1, pp. 90–105, 2018.

[76] J. Hastie, Trevor, Tibshirani, Robert and Friedman, *The Elements of Statistical Learning*. New York, NY, USA: Springer New York Inc., 2001.

[77] D. Choudhary and M. Cescon, “EDA-sense: Dynamic Feedback Control of Sympathetic Arousal,” *IFAC-PapersOnLine*, vol. 53, no. 5, pp. 238–243, 2020, doi: 10.1016/j.ifacol.2021.04.103.

[78] A. F. Bulagang, N. G. Weng, J. Mountstephens, and J. Teo, “A review of recent approaches for emotion classification using electrocardiography and electrodermography signals,” *Informatics Med. Unlocked*, vol. 20, p. 100363, 2020, doi: 10.1016/j.imu.2020.100363.

[79] G. E. Schwartz and S.-L. Brown, “Relationships between facial electromyography and subjective experience during effective imagery,” *Biol. Psychol.*, vol. 11, pp. 49–62, 1980.

[80] S. Katsigiannis and N. Ramzan, “DREAMER: A Database for Emotion Recognition Through EEG and ECG Signals from Wireless Low-cost Off-the-Shelf Devices,” *IEEE J. Biomed. Heal. Informatics*, vol. 22, no. 1, pp. 98–107, 2018, doi: 10.1109/JBHI.2017.2688239.

[81] H. M and S. M.N, “A Review on Evaluation Metrics for Data Classification Evaluations,” *Int. J. Data Min. Knowl. Manag. Process*, vol. 5, no. 2, pp. 01–11, 2015, doi: 10.5121/ijdkp.2015.5201.

[82] C. K. A. and G. L. Fernando Nogueira, “Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning,” *J. Mach. Learn. Res.*, vol. 18, pp. 1–5, 2017, [Online]. Available: http://jmlr.org/papers/v18/16-365.html.