Universidade de Aveiro
Ano 2023

Tiago
Almeida Cruz

Assinatura Biométrica da Emoção: Estatística Multivariada

Biometric Signature of Emotion: Multivariate Statistics

Universidade de Aveiro
Ano 2023

Tiago Almeida Cruz

Assinatura Biométrica da Emoção: Estatística Multivariada

Biometric Signature of Emotion: Multivariate Statistics

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 Eletrónica e Informática de Aveiro (IEETA), da Universidade de Aveiro.

Dedico este trabalho à minha família pelo incansável apoio e a todos os meus amigos que me ajudaram ao longo do meu percurso académico.

o júri

presidente

Prof. Doutor Carlos Davide da Rocha Azevedo
Investigador Auxiliar em Regime Laboral, Universidade de Aveiro

vogais

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

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

agradecimentos

Em primeiro lugar, quero agradecer à melhor orientadora que podia ter tido, a Doutora Susana Brás, por todo o apoio e ajuda incansável, por estar sempre disponível para me ajudar, por todo o conhecimento transmitido e pela liberdade, sempre orientada, no desenvolvimento desta dissertação, promovendo as minhas ideias e raciocínios, tornando-a um desafio enriquecedor.
Quero agradecer também aos meus amigos com quem partilhei todos estes anos memoráveis.
Finalmente, agradeço à minha família, sobretudo aos meus pais e ao meu irmão, que são as pessoas que sempre me apoiaram de forma incondicional e estiveram sempre presentes para me apoiar e depositar confiança.

palavras-chave

Emoção, Sinais fisiológicos, Análise Fatorial, Análise de Clusters, Análise de Componentes Principais, Classificação emocional, SVM, Modelo Discreto.

resumo

O reconhecimento emocional desempenha um papel fundamental na vida do ser humano, nomeadamente, na compreensão do seu comportamento, nas suas tomadas de decisão e bem-estar geral. Com os avanços tecnológicos, têm sido investigados e desenvolvidos diversos métodos de reconhecimento emocional, sobretudo com base nos sinais fisiológicos. No entanto, continuam a existir desafios no desenvolvimento de modelos de reconhecimento emocional robustos e fiáveis. A subjetividade inerente às experiências emocionais, as diferenças entre indivíduos e as influências contextuais são aspetos que devem ser tidos em conta aquando da construção dos ditos modelos. Nesse sentido, a consideração e avaliação, a priori, das características e estados emocionais dos participantes, através de questionários devidamente implementados para esse efeito, num processo de reconhecimento emocional e o estabelecimento de grupos de indivíduos por contexto com base em semelhanças, quer em termo de respostas fisiológicas quer em termos de respostas aos questionários referidos, pode conferir melhores resultados aos sistemas de reconhecimento emocional.

Neste trabalho foi efetuado o agrupamento dos participantes, de forma a perceber se isso conferia melhores resultados ao modelo de reconhecimento emocional, comparativamente com os resultados obtidos segundo a consideração de todos os indivíduos como um todo. Para tal foram consideradas, para o estabelecimento dos grupos, a análise fatorial e a análise de clusters, sobre as respostas a dois tipos de questionários baseados na *State-Trait for Cognitive and Somatic Anxiety* (*STICSA*) e na *Visual Analog Scale (VAS*) e os sinais fisiológicos. Foram construídos modelos de classificação para a combinação de cada conjunto de participantes com cada conjunto de sinais fisiológicos. Verificou-se que a consideração do eletrocardiograma, da atividade eletrodérmica e do eletromiograma em conjunto permitia obter geralmente melhores resultados de classificação emocional face à sua consideração em separado. Além disso, averiguou-se que a construção de modelos emocionais por cada grupo gerava modelos mais robustos. O que levou a concluir que a separação dos indivíduos por grupos baseados no seu estado emocional e a adaptação dos modelos aos mesmos permite a obtenção de uma melhor descrição fisiológica do processo emocional.

keywords

Emotion, Physiological Signals, Factor Analysis, Cluster Analysis, Principal Component Analysis, Emotional Classification, SVM, Discrete Model.

abstract

Emotional recognition plays a fundamental role in the life of human beings, namely in understanding their behavior, decision-making and general well-being. With technological advances, several methods of emotional recognition have been investigated and developed, mainly based on physiological signals. However, challenges remain in developing robust and reliable emotional recognition models. The subjectivity inherent to emotional experiences, the differences between individuals and the contextual influences are aspects that should be considered when building these models. In this sense, the consideration and evaluation, a priori, of the characteristics and emotional states of the participants, through questionnaires properly implemented for that purpose, in an emotional recognition process, and the establishment of groups of individuals per context based on similarities, both in terms of physiological responses and in terms of answers to the mentioned questionnaires, may give better results to emotional recognition systems.

In this study, the grouping of participants was carried out to understand if this gave better results to the emotional recognition model, compared to the results obtained by considering all individuals as a whole. To this end, factor analysis and cluster analysis were used to establish the groups, on the answers to two types of questionnaires based on the State-Trait for Cognitive and Somatic Anxiety (STICSA) and the Visual Analog Scale (VAS) and on the physiological signals. Classification models were built for the combination of each set of participants with each set of physiological signals. It was found that the consideration of the electrocardiogram, electrodermal activity and the electromyogram together allowed for generally better emotional classification results than considering them separately. In addition, it was found that the construction of emotional models by each group generated more robust models. This led to the conclusion that the separation of individuals by groups based on their emotional state and the adaptation of the models to them allows a better physiological description of the emotional process to be obtained.

Contents

List of Figures

List of Tables

List of Acronyms

APA	American Psychological Association
BVP	Blood Volume Pulse
ECG	Electrocardiogram
EDA	Electrodermal Activity
EEG	Electroencephalogram
EMG	Electromyogram
EMG MF	Electromyogram Frontalis Muscle
EMG TR	Electromyogram Trapezius Muscle
EMOTE	Emotion Multimodal Database
EOG	Electro-oculography
FN	False Negative
FP	False Positive
GSR	Galvanic Skin Response
HRV	Heart Rate Variability
KMO	Kaiser-Meyer-Olkin
NaN	Not a Number
PAD	Pleasure-Arousal-Dominance
PCA	Principal Component Analysis
RBF	radial basis function
RFECV	Recursive Elimination with Cross-Validation
RM ANOVA	Repeated Measures ANOVA
RSP	Respiration
STICSA	State-Trait for Cognitive and Somatic Anxiety
SVM	Support Vector Machine
TAS-20	Toronto Alexithymia Scale
TN	True Negative
TP	True Positive
VAS	Visual Analog Scale

1 Introduction

1.1 Context

Emotion plays a fundamental role in human life. It allows each human being to protect himself from dangers, from possible pathogens present in decaying food to poisonous animals, for example, and to react to them. It allows us to enjoy a certain experience, as well as acting as a communicative tool that often allows each person to identify the emotional state of another in a given situation and help in decision-making. In this sense, any emotion plays an essential role in human life [1]. The negative character that is sometimes associated to certain emotions ends up being a consequence of the misinterpretation that each individual makes and because it is a not so pleasant emotion to experience [1]. Over time, several definitions of emotion have emerged, almost a different definition per researcher both in the areas of psychology and affective computing, and there is no consensual one [2]. However, emotions are seen as responses to a given stimulus, and can manifest themselves through speech, facial expressions, changes in physiological signals in the human body, such as changes in heart rate, blood pressure, breathing, among others, and/or behavioral changes, which leads us to approach and interact or withdraw from a given experience [3]–[6].

With the studies and evolutions around the concept of emotion, some emotion models have emerged, two of which have become more widely adopted: the dimensional model and the discrete or categorical model, which allow characterizing a given emotion [7]–[9]. In the dimensional model, emotions are characterized based on their intensity in relation to a given dimension (valence, arousal, domain) [7], [8]. Valence represents how pleasurable emotion is by assuming a polarity, i.e., ranging from negative to positive, e.g., being happy is more positive/pleasurable than being disappointed [8]. Arousal represents how active the emotion is, e.g., excited represents a higher state of activation than serenity or tiredness [8]. Dominance consists of the emotion's ability to influence the environment or be influenced by it, assuming a restrictive character [8]. In the discrete or categorical model, there is characterization according to a limited list of emotions [7]–[9].

Emotion recognition is a core area and can be established essentially through recognition of facial expressions, speech analysis and analysis of physiological signals, or a combination of these methodologies. Facial and discursive emotion recognition can be performed through observation and listening, respectively. In peer-to-peer communication, emotion recognition by observing face and posture, as well as the sound of the voice, allows for fast and effective emotion recognition. In automatic detection facial features have a large description in the literature, often constituting a ground truth for emotion studies. However, these features are easily manipulated, making emotional recognition either through observation or affective computation ineffective [3], [4]. Emotional recognition through physiological signals such as Electroencephalogram (EEG), Electrocardiogram (ECG), Electromyogram (EMG), Electrodermal Activity (EDA), among others, is more reliable and robust given the high difficulty of manipulating them due to their association with the autonomic nervous system, their involuntary nature, and the use of robust equipment to acquire them [5], [6]. The major disadvantage of using these signals is related to the intrusiveness of placing electrodes, however, with the development of new wearable technologies, it is becoming increasingly comfortable and efficient to collect this information. The combination of several emotion recognition methodologies may constitute an even more robust system for the identification of a given emotion.

The induction of a certain emotion in a subject can be done in several ways, such as through memories, images, music, videos, among others. However, the use of videos can confer the

activation of several senses simultaneously and can be more effective in this process. Obviously, the theme of the videos is essential in order to trigger the desired emotion. For example, the use of horror videos will be associated with the triggering of fear in the analyzed individuals. However, it is important to note that each subject reacts differently to the same stimulus, in this case, to a particular video. What may cause greater intensity of fear in one individual may cause a less exacerbated reaction in another. In this sense, given the differences between individuals, even in terms of changes in physiological signals when experiencing the same emotion, it is important to establish a profile associated to each subject. The analysis and consideration of the perception of individuals about their own emotions, through questionnaires for example, and their a priori introduction in an emotional recognition process, can give better results to emotional recognition systems. The process of collecting, cleaning, and analyzing the obtained signals are fundamental processes in these systems. The identification of emotions, based on the extracted features, is performed through classification models, namely machine learning or deterministic.

Nowadays emotional recognition has become a key concept and is extremely important in any area of performance, life stage or situation [10]. In this sense, emotional recognition has captured great interest, and consequently the number of studies in this area has increased [10]. The ability to recognize emotions through the use of technological systems can improve people's quality of life and well-being. In the health sector, emotion recognition can be used to diagnose and monitor mental health conditions such as depression, anxiety, among others, or simply to monitor patients' emotional states and provide them a personalized care. In the education sector, it can allow an improvement and adaptation of teaching methods, in order to allow a better performance of the students [7], [11]. In the entertainment sector, it can allow the adaptation of artists to their target audience, and on the other hand improve interactions in virtual games [12], [13]. In the robotics sector, it can contribute to the improvement of the interaction between humans and robots by allowing them to respond to humans' emotions [12]. This can provide greater comfort and well-being, through developments associated to assisted living, in customer services performed by robots, among others. In the economic sector, namely, in the business area, emotion recognition may allow a better evaluation of the customers' satisfaction level, a better emotional control when executing investments, the creation of more attractive advertisements, among others [7]. Overall, the applications of emotion recognition are broad and diverse, and the field is continuing to evolve as researchers study and develop systems capable of recognizing, detecting and processing emotions [14].

1.2 Objectives

Emotional recognition, based on physiological signals and self-assessment questionnaires of the emotional state, is important to understand how human beings react physiologically to a given emotional stimulus and how they describe what they experience. Human beings are complex systems and react in different ways to the same emotional stimulus, which may make it important to consider, a priori, groups of individuals who respond in a similar way to the stimulation generated, in order to build more efficient emotional recognition models. Therefore, this work aims to interpret and describe an emotional profile through a priori characterization of participants. To accomplish this major goal, some open questions had to be analyzed and discussed:

1. Analyze the questionnaires and the physiological signals and define groups of individuals from them.
2. Building and optimizing machine learning models for each group individually and for the global case where all participants are considered together.

3. Understand if the Principal Components Analysis, by bringing changes in the dimensional space, allows obtaining better results in emotion classification.
4. Understand if the analysis of the signals separately gives better results in terms of performance of the model developed in comparison with the consideration of the signals together.
5. Understand whether the separation into groups confers improvements in terms of the results of the emotional classification model compared to the global case (without separation by groups).

1.3 Outline

This document is divided into a sequence of 6 chapters which were stipulated in order to achieve the objectives proposed in the previous section.

- **Chapter 1:** contextualizes the theme of this exploratory study and presents and describes the main objectives defined.
- **Chapter 2:** describes the state of the art associated with this work.
- **Chapter 3:** establishes a detailed characterization of the database considered.
- **Chapter 4:** describes the base methodology for the definition of the participant groups and for the construction of the emotion models.
- **Chapter 5:** presents the results of the definition of the groups, the exploratory data analysis and the classification of the emotional models developed, as well as their discussion.
- **Chapter 6:** summarizes the results and conclusions generated from the analysis carried out.

2 Background

2.1 Emotion

The word emotion comes from the Latin *movere*, which means movement outwards in order to communicate, interact and react to stimuli from the environment [15].

According to the American Psychological Association (APA), emotion is defined as "a complex reaction pattern involving experiential, behavioral and physiological elements" [16]. The definition presented is merely exemplificative, with a subjective character and not consensual. In fact, there is no consensus, in terms of the scientific community, about the definition of emotion and how it should be interpreted [15]. Due to the associated subjectivity, it is as if there were as many definitions, or attempts at them, as there are researchers [2].

The only aspects on which there seems to be some consensus are the consideration of the existence of a relationship between emotions and the nervous system and the consideration of three factors in the definition of emotion. These are the occurrence of physiological alterations, behavioral alterations, or reactions, which is related to the adoption of a certain reactive behavior before a certain emotion, and the experiential factor, which is associated to the description of each individual in relation to what he is experiencing before a certain emotion, presenting a subjective character [17].

It is essential to mention that emotion is different from feeling. Emotions cannot be rationalized, they are merely felt and expressed unconsciously. Feelings on the other hand are experienced consciously [18]. Feelings are personal, and therefore subjective, perceptions of emotions. They are the conscious form of emotions [19]. For example, when a university student manages to graduate, the emotion of joy usually manifests itself. Through awareness of both this fact and the emotion itself, feelings of gratitude and satisfaction, among others, emerge.

It is important to clarify that no emotion is harmful and that when we talk about positive, negative, or neutral character, it is not related to an emotion being better or worse, good, or bad. All emotions are essential [1]. For example, being afraid is not a problem except in complex and exaggerated cases such as phobias [1]. Fear arises as a form of self-protection, it makes us rationally fear something that we know can hurt us [1]. Another example is disgust, which, if we are not faced with excessive cases such as germaphobia, is another essential form of self-protection [1]. Feeling disgust makes us avoid ingesting or coming into contact with substances that are potentially harmful to health. Disgust is physically manifested by closing our mouths and half-closing our nostrils, precisely as a protective reaction in which we filter our breathing more carefully and effectively, preventing the entry of these potentially harmful substances [1].

2.1.1 Emotional models: Discrete model and Dimensional model

The discrete model and the dimensional model, as previously mentioned, stand out for their greater acceptance, although there is no unanimously accepted emotional model. The discrete model establishes a qualitative description of emotions, which may not be possible in cases of simultaneous experience of several emotions. The dimensional model, on the other hand, establishes a quantitative description, which may allow combating the difficulties associated with the discrete model when it comes to the recognition of different emotions [7], [8].

Within the dimensional model, the Pleasure-arousal-dominance (PAD) stands out [8]. In this model emotions are represented based on three parameters: valence, arousal, and dominance [8]. Since valence and arousal are sufficient to obtain a representation of most emotions, a 2D model is used

[8]. This model allows, based on the quantification of the valence and arousal parameters, carried out from the own recognition of the felt emotion, to categorize it [8]. As can be seen in figure 2.1, the vertical and horizontal axes are represented, respectively, by the arousal and valence parameters. In the first quadrant we can find an emotional state characteristic of happiness, in the second quadrant of anger, in the third of sadness and in the fourth of pleasure/relaxation [8], [9].

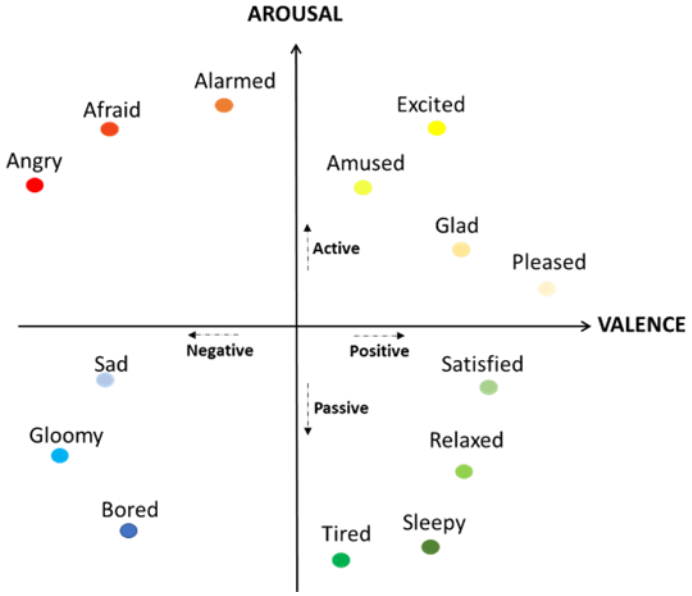

Figure 2.1: Two-dimensional model of valence and arousal, based on [20].

In the discrete model, as the name implies, emotions are defined as classes according to models, in which Ekman's six basic emotions model and Plutchik's emotional wheel model stand out [7], [9]. Ekman's model, proposed in the 1970s, is based on six basic emotions, these are: surprise, sadness, happy, fear, disgust and anger, as shown in figure 2.2, on the left side [7], [9]. In the 1990s, Ekman himself added other emotions to the six he had previously defined. This new set, seen as a secondary class, includes the emotions: Shame, sensory pleasure, satisfaction, relief, pride in achievement, guilt, excitement, embarrassment, contentment, contempt, amusement [7], [8]. Plutchik's emotional wheel model includes eight basic emotions (Grief, anger, anticipation, sadness, surprise, fear, trust and joy) considered as a base, and the emotions associated with them, with a positive or negative character in terms of intensity, as shown in figure 2.2, on the right side. For example, fear according to this model constitutes a basic emotion, which would be considered as the base, occupying the central circular crown as can be seen in figure 2.2 on the right side, while terror would be a more intense version of fear representing a negative character and apprehension presents a lower intensity exhibiting a positive character [7], [9]. Thus, in this model the more intense emotions are arranged in the central region and the less intense ones in the extremities.

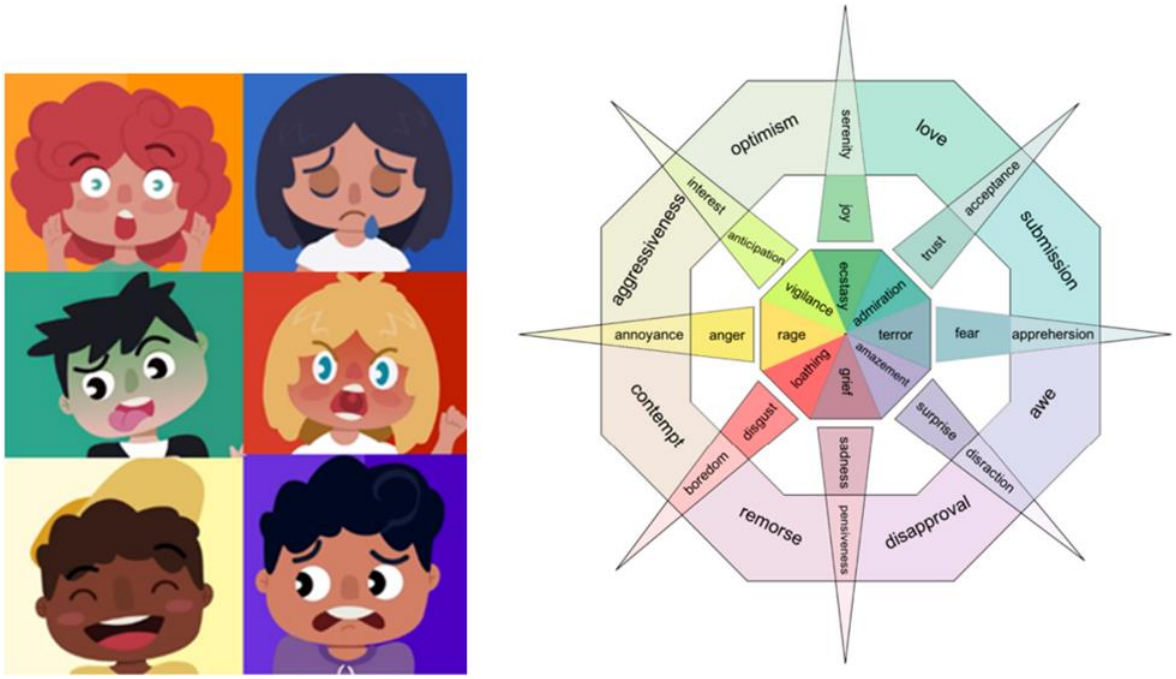

Figure 2.2: Discrete models of emotion. Ekman's 6 basic emotions model illustrated through cartoons (on the left), based on [21], Plutchik's emotional wheel model (on the right), taken from [20].

2.2 Emotion Recognition

Emotion can be recognized in various ways. It can be recognized, for example, through facial expression, speech, gestures, or physiological signals. Each of the emotion recognition methods has advantages and disadvantages, and in the present work the focus is on emotion recognition based on physiological signals [20], [22].

The facial recognition of emotion presents limitations such as the need for a good image quality and a frontal centering of the face and not moving it in front of the acquisition device, for the detection of the facial expression, it can also generate problems related to personal privacy, being of difficult acceptance by the target audience and is easily susceptible to manipulation, in the sense that one can fake facial expressions relatively easily and fool the system. However, facial expression can contain rich information in terms of the experience of a particular emotion, being non-intrusive [3], [20], [22].

Emotion recognition through speech has some associated difficulties such as the need for large databases due to the huge amount of differences in individual speech signals, the quality of the signals is highly dependent on background and ambient noise, and it is easy to manipulate since one can fake emotions through speech with relative ease. However, it can present itself very advantageous in non-intrusive and remote analysis, for example, in police or emergency situations via call centers, in which the only interpersonal contact is established via speech [20], [22].

The consideration of physiological signals as a basis for emotional recognition has limitations such as the data acquisition system, as is the case, for example, of the electrodes, whose placement may become intrusive, which may increase the difficulty to collect the data [20]. However, it has advantages such as greater robustness, that is, the physiological signals are spontaneous by themselves, preventing manipulation of the system [20]. Furthermore, with the introduction of wearable technologies, data collection has become simpler allowing mobility to the individual during data collection [20], [22].

Despite the various and different definitions of emotion, there is an inevitable aspect that is the relationship between emotion and our body, more specifically with our nervous system [12]. This means that when faced with a particular emotion, a reaction is triggered in our body in physiological, behavioral, communicative terms, among others [12].

In this sense, emotional recognition is based on the detection and identification of the triggered reaction and the recognition of patterns of relationship between a certain change and the emotion witnessed. Emotion recognition through audio and facial recognition of emotions are topics widely covered in the existing bibliography. However, these types of recognition are subject to a significant level of inefficiency due to the disadvantages presented above. To overcome the problems associated with them, over the last few years, new methodologies for emotional recognition have emerged, namely, the recognition of emotions through physiological signals (Electroencephalogram (EEG), Electrocardiogram (ECG), Electromyography (EMG), Electrodermal activity (EDA)) collected through sensors placed on the human body as exemplified in figure 2.3 [20]. Changing features in these signals have been proven to be related to emotional changes [20]. Multimodal analysis, either by considering information from different physiological signals or by combining physiological signals with behavioral information: facial expressions, audio analysis, among others, can confer greater robustness to an emotional recognition system [12].

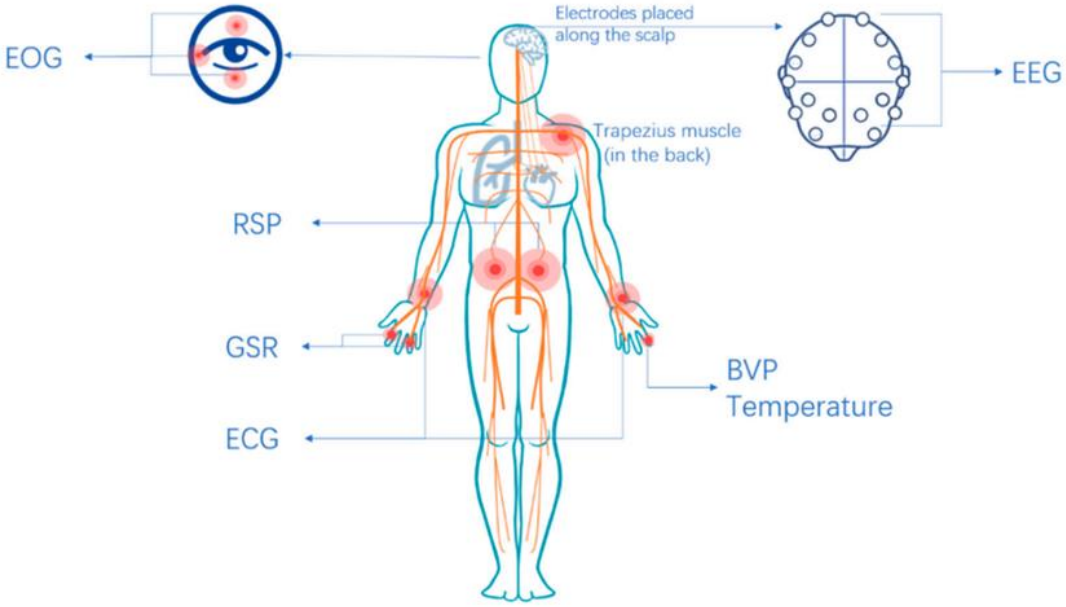

Figure 2.3: Positioning of sensors to collect physiological signals, Electrocardiogram (ECG), Blood Volume Pulse (BVP), Temperature, Galvanic Skin Response (GSR), Respiration (RSP), Electro-oculography (EOG) and Electroencephalogram (EEG). Taken from [20].

Difficulties may arise, of generalization of a certain emotional recognition system, arising from the fact that all human beings present differences among themselves. That is, the same situation may unleash diverse emotions in distinct subjects.

3 Database

In this work, it was considered a multimodal emotion database of physiological signals using movies as stimulus, called Emotion Multimodal Database (EMOTE), created within the scope of a master's thesis, at University of Aveiro, entitled of "Discrete to dimensional physiological emotion classification" [23]. The data that make up this database were collected using an experimental protocol that was approved by the Ethics and Deontology Committee of the University of Aveiro and by the delegate for the data protection affairs, with evaluation approval 12-CED/2020 [23].

3.1 Inclusion criteria

To be included in the experiment the participants needed to have an age between 18-35 years old, no psychological or psychiatric diagnosed disorder, no physical condition that could impact the experiment (e.g. cardiac arrhythmia) and no medication that could impact the experiment or that may be indicative of a psychological or psychiatric disorder (e.g. anti-depressives). Based on this inclusion criteria participants were selected or not to integrate in this experiment [23].

3.2 Setup

The physiological signals considered at the time of data collection were ECG, EDA and EMG. These signals were gathered by the 4-channel biosignalplux from the Biosignalplux Research Kit, utilizing five sensors [23]. More specifically one ECG sensor, one EDA sensor, two EMG sensors (one for EMG MF and the other for EMG TR), and an extra lead ground cable to act as a reference to the other sensors, allowing the acquisition of four signals [23]. The EDA and EMG sensors have two electrodes: one positive (red) and one negative (black). On the other hand, the ECG sensor has an additional electrode corresponding to a ground truth electrode (grey) [23]. Electrodes placement, which follows the scheme represented in figure 3.1, was equal for all participants to ensure that it was not a variable that can cause variations on the acquired data. For more details on the positioning of the electrodes and on the signal acquisition process please consult this reference [23].

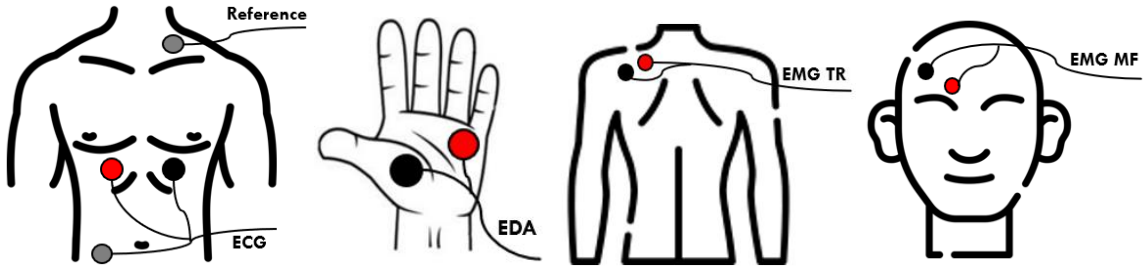

Figure 3.1: Positioning of the electrodes for the acquisition of the ECG, EDA, EMG TR and EMG MF signals, respectively, based on [23].

Each participant watched a movie compilation characterized by three sequential short movie excerpts with high intensity. There were used comic, terror and neutral movie types, to induce three different emotions: happiness, fear and neutral. There were acquired participants physiological response to those movies. That acquisition was made in two sessions, with a week of separation [23].

In the first session, there were considered a sequence of three sets, each composed of movie excerpts of the same movie type, ordered ascendingly by emotion intensity, and had a total of approximately 10 minutes. The first, second and third sets were composed of movie excerpts from

horror, comedy, and documentary movies, to cause fear, happiness, and neutral state, respectively [23].

In the second session, there were considered a sequence of three sets, following the idea of the first session. Although, in each set there were considered 5 minutes of repeated excerpts, in other words, excerpts used in the first session, followed by 10 minutes of new excerpts [23].

It should be noted that because of the variability of physiological response between participants, there was, at began of each experimental session, considered a five-minute neutral documentary, that consists in a baseline that triggered no emotion to the participant [23].

3.3 Questionnaires

Firstly, each participant was asked to answer a sociodemographic questionnaire, to acquire personal information like age, nationality, gender, education, professional situation, health issues and restrictions and other questions about exercise practices and personal habits [23]. This questionnaire allowed the selection of the participants, according to the inclusion criteria described in section 3.1 [23].

Three more types of questionnaires were used in addition to the sociodemographic questionnaire to assess the participants' characteristics and emotional states [23]. They are Toronto Alexithymia Scale (TAS-20), State-Trait Inventory for Cognitive and Somatic Anxiety (STICSA), and Visual Analog Scale (VAS) [23].

The TAS-20 is self-administered questionnaire, composed by 20 items, that measures difficulty in identifying emotions, describing emotions and externally oriented thinking, which are the alexithymia dimensions. There are five choices for each item, each one based on the degree of agreement with it (1 – Strongly Disagree to 5 – Strongly Agree) [24], [25]. The total alexithymia score is the sum of all item's response (0-51 No alexithymia, 52-60 Possible alexithymia, 61-100 Alexithymia present). This instrument is used to detect participants that have Alexithymia, once that can compromise the self-assessment of their emotional state during sessions [24], [25].

The STICSA is a self-report questionnaire consisting of 42 – items, used to assess cognitive and somatic dimensions of trait and state associated with anxiety [26]. It is composed of affirmations about symptoms related with anxiety [23], [26], [27]. Participants should indicate how they feel according to items rating from 1 (Never) to 4 (A lot). For items related to anxiety trait, the participant should answer how usually, in general, the affirmation is true to him. In contrast, for items related to anxiety state, he should answer in terms of how he feels at a specific moment [23], [26], [27]. Higher scores indicate higher levels of anxiety [23], [26], [27].

Participants answered the STICSA-Trait questionnaire only at the beginning of the first session, and the STICSA-State questionnaire at the beginning of both sessions [23]. This because it is important to know the participants state, in other words, how they feel at the start of each session, once that it can have influence and explain their answers in VAS-Pre and Vas-pos questionnaires and their physiological responses to the different movie excerpts [23].

The VAS-Pre and VAS-Pos are, respectively, used to the participant assess their own emotional state before and after, watching the movie excerpts, in first and second sections [23]. Each questionnaire is composed of two parts [23]. One, it's constituted by a scale from 0 to 100% related to anxiety, happiness, fear and stress [23]. The other part is about the valence and arousal dimensions, where the scale goes from -5 to 5 [23].

3.4 Dataset description

Emotion Multimodal Database (EMOTE) is composed of 29 participants, considering that one was removed due to duplicate information, that were emotionally stimulated using horror, comedy and documentary movie excerpts [23]. For each participant there are a sociodemographic, TAS-20, STICSA-Trait, STICSA-State, VAS pre and VAS pos questionnaires, and 2 files with simultaneous physiological signals, more specifically ECG, EDA, EMG MF, EMG TR sampled at 1000 Hz sampling rate [23]. These files correspond to different sessions, separated in time by one week [23].

The 29 selected individuals are aged between 18 and 25 years (21.45±1.38), with 20 (69%) of them being female and 9 (31%) of them being male, as can be seen from the distributions in figure 3.2. Around 90% of the participants are students, approximately 7% are student-workers, and about 3% have another professional status, as shown in figure 3.2. When it comes to health 14% have a psychological or psychiatric problem, 69% have vision problems and 38% take any medication. Regarding their habits, around 76% admitted that usually drink coffee or consume products with caffeine and 24% do not. In terms of exercise, 34% don't practice regularly and 66% do. It is important to note that individuals did not drink coffee or exercise one hour before the signal collection. 14 participants don't have alexithymia, 12 have the possibility of having it and 3 present this difficulty or incapacity to express their emotions.

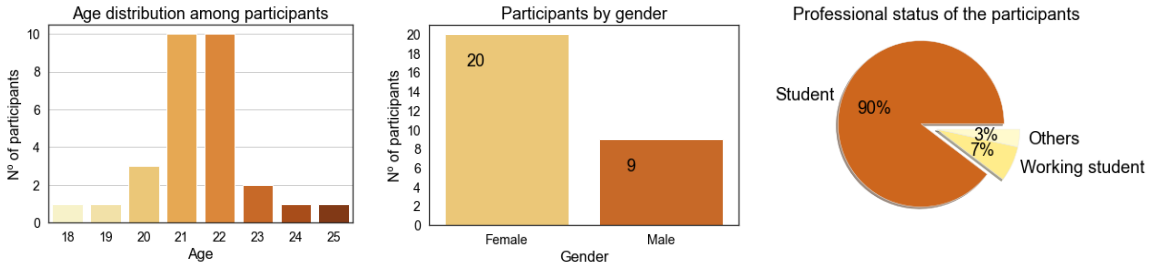

Figure 3.2: Statistical description about age, gender and professional situation on the sociodemographic questionnaire.

4 Methodology

4.1 Grouping of participants

Emotional recognition models aim to identify the emotional state of the individual. In this sense, they should be built in order to obtain a good performance. Emotion manifests itself through bodily and physiological responses given its relationship with the nervous system. Interpersonal differences mean that the experience of the same emotion may be seen differently among individuals, both in terms of physiological signals and in terms of their own description of the emotion experienced.

A priori definition of groups composed of individuals with similar responses, in terms of physiological signals and specific questionnaires about emotional recognition (STICSA-State, STICSA-Trait, VAS), can reveal itself as an essential step in the establishment of emotional recognition systems with better results, given their targeting to the subjects' profile.

The grouping of participants based on physiological signals and questionnaires may be carried out in several ways and through different techniques. In this work we initially tested the hypothesis of forming groups of participants based on the scores of the factors generated from the factor analysis. This is because each factor is associated with each feature, and the scores associated with the factors may thus allow for the effective formation of clusters. Assuming the difficulty of validating this methodology through existing literature, cluster analysis was also assumed as an alternative hypothesis for the formation of the said groups.

4.1.1 Data normalization

When variables present high difference in scale, the normalization of values allows a true comparison between them. Also, the building of Machine Learning models based on normalized variables allows a better parameter adjustment. There are some data normalization methods, among which the Min-max method and the z-score method stand out [28]–[30]. The Min-max method places the values associated with the variables in a range between 0 and 1 [28]–[30]. The z-score method places the values centered on a zero mean and with a standard deviation equal to unity [28]–[30]. In case the standard deviation is very small, or the distribution does not follow a Gaussian, the Min-max method should be used instead of the z-score method [28]. In the case of many outliers, the z-score method may be more adequate, however, in this situation it is best to test and evaluate the result and then decide which methodology is better for the data and models considered [28].

Min-max normalization method [29], [31]: Transforms the values to a range between 0 and 1. The maximum value of the variable is max and the minimum value is min. The equation that represents it is:

$$Z_i = \frac{X_i - \min(X_i)}{\max(X_i) - \min(X_i)}. \tag{4.1}$$

Standardization method (z-score) [29], [31]: Assuming that the variable X_i has a normal distribution with mean μ_i and variance $Var(X_i)$ (standard deviation σ_i). The equation that represents it is:

$$Z_i = \frac{X_i - \mu_i}{\sqrt{Var(X_i)}}. \tag{4.2}$$

4.1.2 Factor analysis

Factor analysis is used in order to simplify and describe the variables as a function of so-called factors. Factors are common to the original variables. They are neither observable nor measurable [29], [31].

Through the analysis of loadings (weights) or correlations, we can perform a grouping of the original variables according to each factor considered [29], [31].

Factor analysis defines p original variables as a linear combination of m factors, also called latent factors, where the number of factors is always less than the number of original variables ($p > m$) [29], [31]:

$$X_1 - \mu_1 = l_{11} \times F_1 + l_{12} \times F_2 + \ldots + l_{1m} \times F_m + \varepsilon_1 \quad (4.3)$$

...

$$X_p - \mu_p = l_{p1} \times F_1 + l_{p2} \times F_2 + \ldots + l_{pm} \times F_m + \varepsilon_p . \quad (4.4)$$

Where $L = \begin{bmatrix} l_{11} & \cdots & l_{1m} \\ \vdots & \ddots & \vdots \\ l_{p1} & \cdots & l_{pm} \end{bmatrix}$ corresponds to a matrix of the factor weights, where l_{pm} corresponds to the weight of the m-th factor on the p-th original variable, $F_1, \ldots, F_m$ to the factors or latent variables, $\varepsilon_1, \ldots, \varepsilon_p$ represent the errors or specific factors, and $\mu_1, \ldots, \mu_p$ correspond to the averages of the corresponding original variables that allows in the equation to generate centralized data.

Figure 4.1 illustrates this transformation of the original variables into the factors in terms of tabular indices.

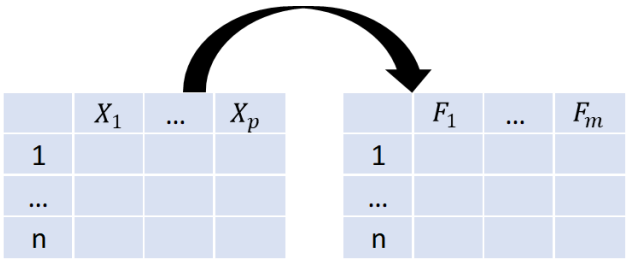

Figure 4.1: Transformation in terms of tabular indices resulting from the transformation performed by factor analysis.

Figure 4.2 shows the relationship between the original variables and the factors, which is established through the weights of the factors against each of the original variables.

	F_1	...	F_m
X_1	l_{11}	...	l_{1m}
...	...	...	...
X_p	l_{p1}	...	l_{pm}

Figure 4.2: Matrix representing the relationship between the original variables and the latent factors.

This model leads to the covariance matrix being equal to the matrix of factor weights (L),

$$Cov(X_i, F_k) = l_{ik} , i = 1,2, \ldots p \; and \; k = 1,2, \ldots m . \quad (4.5)$$

Given that,

$$Corr(X_i, F_k) = \frac{\left(Cov(X_i, F_k)\right)}{\sqrt{Var(X_i) \times Var(F_k)}}\ , \tag{4.6}$$

we can conclude that the greater the covariance, and therefore, according to equation 4.6, the greater is the weight of the factor on a given variable, the greater the correlation between said factor and the said original variable [29], [31].

To obtain the factors there are some methods, among which we highlight the principal component method and the maximum likelihood method [29], [31].

Prior to implementing the factor analysis and analyzing the associated results, a check should be made on the appropriateness of the factor analysis model to the data. To be adequate, it is necessary that elements outside the diagonal of the data covariance matrix, and therefore of the correlation matrix, are not approximately null, that is that the variables are correlated among themselves [31]. In other words, if the covariance matrix of the data, or the correlation matrix, is significantly different from an identity matrix, then factor analysis will be appropriate for the data [31]. To verify the same one can resort to Bartlett's test of sphericity which is based on the following hypothesis test [29], [31]: $H_0: \sum = I \; Vs \; H_1: \sum \neq I$, with $\sum$ being the correlation matrix or covariance matrix and I the identity matrix [32]. This test allows by means of the test statistic to analyze the null hypothesis that the variables are not correlated among themselves in the population considered [32]. Obtaining a high value test statistic, or a p-value lower than the significance level, usually given by 1%, 5% or 10%, leads to the rejection of the null hypothesis and the assumption of the alternative hypothesis H1, which would mean an adequacy of the factor analysis model to the data considered before this test [32].

In order to assess whether the samples are adequate for the procedure of a factor analysis, the KMO (Kaiser-Meyer-Olkin) measure can be used [31], [32]. The analysis of the KMO values allows us to discuss whether we should remove some variables, to maximize the efficiency of the model. Table 4.1 allows us to analyze the degree of adequacy of the factor analysis against the KMO value obtained for each variable [33]. The original variables with an associated KMO value less than or equal to 0.5 are generally disregarded from the study [29].

KMO value	FA model
> 0.9	Excellent
Between 0.8 and 0.9	Good
Between 0.7 and 0.8	Medium
Between 0.6 and 0.7	Mediocre
Between 0.5 and 0.6	Poor but acceptable
≤ 0.5	Unacceptable

Table 4.1: Measure of sampling adequacy, KMO [29].

Another issue to be taken into account is that if it is not possible to understand what the factors might represent or how many factors should be used, the factor analysis can also be considered inadequate [31].

There are parameters to take into consideration and to be optimized such as the number of factors, the rotation of the factors, the method for estimating the factor scores, and to verify if the factor analysis is adequate to the data and if all the variables involved are relevant or not in this analysis [29], [31].

An increase in the number of factors has an influence on the weights estimated by the maximum likelihood method, although the same does not occur in the principal component method [29], [31]. The number of factors to be considered can be obtained in several ways. One of them is through the Kaiser rule, in which one chooses to retain the number of factors with explained variance greater than 1, that is, the factors whose associated eigenvalues are greater than the unit [29], [31]. Another way, this one more visual and subjective, is the choice of the number of factors through the elbow rule. Through the construction of a Scree plot, it is possible to perform this analysis, in which the factors are considered until reaching one in which the difference of the variability explained between it and the next factor is little accentuated [31].

In parallel to the Scree plot, an analysis can be approached, where databases and correlation matrices are randomly generated with the same characteristics as the one used, in order to make a comparison of eigenvalues. The factors are maintained if their eigenvalues are greater than the eigenvalues of the randomly generated databases [31].

In terms of factor rotation, this generates a change in terms of variance explained by each of the factors. There is no rule on which to base a choice of a better rotation. What varies across this rotation is the interpretation of the factors. There are some orthogonal rotation methods, such as the varimax method, which allows obtaining a strong association and, therefore, a high weight between each original variable and a determined factor, and a reduced weight (weak association) with the rest of the factors considered [31]. This type of rotation allows the establishment of groups based on the relationship between each of the factors and a set composed of some of the variables [29], [31]. The quartimax rotation in turn is assumed as a method in which arises a factor that is strongly associated with all the variables, that is, it allows one factor to be obtained that can be seen as being general, and other more specific factors with which each original variable would be more associated with one of them and less with the others [29], [31].

In terms of estimating factor scores, the Thompson method and the Bartlett method arise [29], [31]. Thompson's method is based on the use of the method of least squares and requires that the original variables follow a multivariate normal distribution, and therefore that each variable follows a univariate normal distribution [29], [31]. Bartlett's method, on its turn, is based on the weighted least squares method and does not present requirements on the data [29], [31].

4.1.3 Cluster Analysis

In terms of clustering algorithms, there are essentially two classes of methods, hierarchical and non-hierarchical [29].

4.1.3.1 Non-hierarchical methods

Non-hierarchical methods are techniques that allow dividing the data into n clusters, whereby the number of clusters is defined and therefore known upfront [34]. These clusters can be generated from a random division of the elements that compose the said initial clusters, or selecting initially, in a random way, a set of points that will be the nuclei of the initial clusters [29], [34]. These methods can be used on high dimensionality data sets without strongly influencing their computational efficiency, constituting an advantage of non-hierarchical methods over hierarchical ones [29], [34]. In terms of non-hierarchical methods in practice, the use of the K-means method stands out.

K-means method:

In this method k initial points are defined which constitute centroids of k clusters to be formed [29]. Then, each individual is assigned to the cluster that presents the centroid closest to it [29], [34].

After each addition of a new element to a cluster, the centroid of each cluster is updated [29], [34]. This process is executed until there is no more change in the centroids of each cluster between successive iterations [29], [34].

4.1.3.2 Hierarchical methods

Hierarchical methods consist of techniques that allow the establishment of groupings of data in a hierarchical manner, based on the degree of similarity or dissimilarity between them. This set of methods is subdivided into agglomerative methods and divisive methods [29], [34]. Agglomerative methods are the most widely used. In this sense, in this paper only the agglomerative methods are considered.

Agglomerative methods:

The agglomerative algorithm starts with a number of groups equal to the number of individuals (or variables), that is, with each of the groups containing one individual (or variable) [29], [34]. Subsequently, a matrix of distances between the groups is generated, in this case, the distances between individuals two by two [29], [34]. From the generated matrix, a new group is formed consisting of the two groups that were close to each other [29], [34]. Then, the groups are grouped recursively, taking into account the use of a similarity measure and a clustering criterion, until the point is reached where there is a single grouping, composed of all individuals (or variables) [29], [34]. It is thus important to consider the proximity measure, which allows measuring the similarity or dissimilarity between two individuals or two variables, and the clustering criterion [29], [34].

In terms of proximity measures, when we want to group individuals, we usually use distance measures, such as Euclidean distance, Minkowski distance, Canberra distance, among others, according to which the smaller the distance measure, the greater the proximity between the individuals involved. In turn, when it is intended to group variables, measures of association are usually used, such as Pearson's correlation, the cosine measure, measures based on Jaccard, Sokal, Rand, Belson, χ^2, among others [29], [34]. The higher the value of the association measure, the greater the proximity of the variables. In terms of cluster aggregation criteria, which are used when each cluster holds more than one element, the methods, Single linkage, Complete linkage, Average linkage, Centroids and Ward's method can be considered mainly [29], [34]. In this work it was decided to choose as a measure of proximity the Euclidean distance and as a criterion for aggregation of clusters the Complete linkage method.

Complete linkage method:

The complete linkage method considers the distance between two groups to be the distance between the two elements, from each group, that are furthest apart [29].

$$d_{A,B} = max\{d_{ij}\}, \tag{4.7}$$

with $i \in A$ and $j \in B$.

This method favors minimization of the intra-group distance, leading tendentiously to find compact groups [29].

Internal validation measure (Average silhouette width):

To assess the quality of clustering results, the average silhouette width was considered. The average silhouette width evaluates the internal consistency of the clusters formed by measuring the cohesion of individuals in their cluster, that is, how similar an individual is to his own cluster in comparison to other clusters. In other words, it assesses the average distance of individuals to their own clusters relative to the average distance to neighboring clusters [29]. The values of the average silhouette width vary between -1 and 1 [29]. The higher its value the better, since it means that

individuals are better connected to their own clusters than to neighboring clusters, and that therefore the groups formed are well defined and well separated [29].

4.2 Statistical analysis

The grouping of participants is considered on the various questionnaires (STICSA-State, STICSA-Trait, VAS-Pre), and even on each physiological signal (ECG, EDA, EMG). Initially, the intention is to group according to each of these questionnaires and validate them using the groupings made based on the baseline of each physiological signal. In other words, the groups formed from the questionnaires would be compared with those formed from the baseline of the physiological signals, through the participants coinciding between groups. If the level of coincidence of participants between groups formed from a given questionnaire and each of the physiological signals is high, the groups of participants formed from this questionnaire will be considered for future analyses. Alternatively, if an adequate level of coincidences between the referred groups is not verified, the questionnaire which allows the definition of groupings with a more consistent internal structure, based on the average silhouette width, will be defined as ground truth. And the groups formed from the remaining questionnaires and from each physiological signal will be compared with that ground truth. The groups that present the greatest difference between them, according to the values of the questionnaire assumed as ground truth, will be selected. This difference will be analyzed and defined according to statistical tests, namely One Way ANOVA or Kruskal-Wallis test. In this case, it will be necessary to implement a normality test on the features of the questionnaire defined as ground truth in order to choose between using the Kruskal-Wallis test or the One Way ANOVA.

4.2.1 Normality Test

A normality test was performed to evaluate the distribution of each feature and to choose the most appropriate statistical test for each case. In this work the Shapiro-Wilk test was considered, given the small sample size.

The Shapiro-Wilk test is a statistical test used to assess whether a given dataset follows a normal distribution [35]–[37]. This test is based on the correlation between the observed data and the expected values of a normal distribution with the same mean and variance [35]–[37]. It calculates the Shapiro-Wilk test statistic (W statistic), which measures the deviation of the sample distribution from a normal distribution [35]–[37]. Higher values of the Shapiro-Wilk test statistic generally suggest a better fit to normality [35]–[37]. This test also gives the p-value that is used to determine the significance of the deviation from normality [36].

The null hypothesis states that the data come from a normal distribution whereas the alternative hypothesis states that data are not taken from a normal distributed population [37].

If the p-value is equal to or less than the significance level of 5% then the null hypothesis is rejected, and we can consider that data don't comes from a normal distribution. In other hand, if p-value is higher than the significance level of 5%, then the hypothesis of normality is accepted [37].

The Shapiro-Wilk's test was implemented with the R function *shapiro.test(),* which takes the feature data and performs the test, returning Shapiro-Wilk statistic, and the p-value [38].

4.2.2 Univariate statistical analysis

4.2.2.1 One-way ANOVA

One-way ANOVA is a statistical method whose goal is to find if there are any statistically significant differences between the means of the groups, that is whether there is one or more groups that are significantly different from the others [39].

The assumptions for One-way ANOVA are that the dependent variable need to be continuous and normally distributed, which was checked by Shapiro Wilk test, that independent variable should consist of two or more categorical, independent groups, typically of three or more categorical and independent groups, independence of observations, that is, no participants being in more than one group, and that the groups variances are homogeneous, which means that the variances should be approximately equal between groups [39].

For checking the homogeneity of variances, the Levene's test was used. In cases where the assumption of homogeneity of variances fails, Welch's ANOVA is used instead of the One-way ANOVA [39].

One-way ANOVA tests the null hypothesis that the groups mean are equal [40]. It calculates the F-statistic, which is the ratio of the between-group variance to the within-group variance [41], [42]. If the F-statistic is greater than the critical value at a given significance level, then the null hypothesis is rejected, and it is concluded that there is a significant difference between the means of the groups.

The F-statistic for the One-way ANOVA is:

$$F = \frac{MS_{\mathrm{b}}}{MS_{\mathrm{w}}} = \frac{MS_{\mathrm{b}}}{MS_{error}}. \tag{4.8}$$

Where MS_{b} is the mean sum of squares for between groups and MS_{w} or MS_{error} is the mean sum of squares for within groups [41], [42].

It also calculates the p-value associated with the F-statistic [42]. If it returns a p-value less or equal to the significance level of 5%, the null hypothesis is rejected and the alternative hypothesis assumed [42]. Which means that at least one group differs significantly from another group. In other words, it means that the mean of at least one group is statistically different from that of another group.

In that case where the null hypothesis is rejected, post-hoc tests or additional analysis can be used to identify which specific groups have significantly different means between themselves [42]

This analysis was performed in R, using the function *aov* [43]. In this function the dependent variable, that is, the values of anxiety, happiness, fear or stress, the independent variable or within factor, which in this case are groups 1, 2 or 3, and the dataframe with the data under analysis are introduced as arguments. Based on the *aov* function and using the summary function, we obtain the F-statistic (*F value*), the p-value for the F-statistic (*Pr(>F)*), the degrees of freedom (*Df*), Sum the squares (*Sum sq*) and mean squares (*Mean sq*) [43].

4.2.2.2 Kruskal-Wallis test

The Kruskal-Wallis test is a non-parametric statistical test whose goal is to determine if there are any statistically significant differences between the medians of two or more independent groups [44]. The assumptions for Kruskal-Wallis test are that the dependent variable needs to be

continuous, that independent variable should consist of two or more categorical, independent groups, independence of observations. It can be used when variables are not normally distributed. Kruskal-Wallis test tests the null hypothesis that there are no significant differences between the medians of the groups being compared. In other words, the null hypothesis states that the medians of all groups are equal [44].

The Kruskal-Wallis test calculates the H statistic, which is a measure of the degree of difference between the ranked data across all groups [45].

The test statistic H for the Kruskal-Wallis test is:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{{R_i}^2}{n_i} - 3(N+1). \quad (4.9)$$

Where N is the total number of participants, in this case, k the total number of groups, n_i the number of participants in group i and R_i the total of ranks of the ith group [45].

It also calculates the p-value associated with the H-statistic. If the p-value is greater than the significance level, the null hypothesis is not rejected, which means that there is no significant difference between the median values of the groups. On the other hand, if the p-value is less or equal to the significance level of 5% the null hypothesis is rejected and the alternative hypothesis assumed, which means that at least one group differs significantly, in terms of the median value, from another group. In that case, where Kruskal-Wallis test rejects the null hypothesis, post-hoc tests or additional analysis can be done to determine which specific group medians are significantly different from each other.

This analysis was performed in R, using the function *kruskal.test* [46]. The arguments introduced in this function meet those considered in the One-way ANOVA, described above. The *kruskal.test* function returns the Test statistic H (Kruskal-Wallis chi-squared), the p-value of the test and the degrees of freedom (*df*).

4.2.2.3 Pairwise Wilcoxon test

Since One-way ANOVA and Kruskal-Wallis test do not allow us to know which pairs of groups are different, a post-hoc test is needed. Thus, a Pairwise Wilcoxon test with multiple testing correction was implemented.

Pairwise Wilcoxon test performs multiple pairwise comparisons between groups [47]. It allows to determine which groups of participants are statistically different, and therefore which physiological signals or questionnaires allow us to define statistically different groups according to the variables Anxiety, happy, fear and stress reported by them at VAS-Pre questionnaire.

This test was implemented in R, using the function *pairwise.wilcox.test* [47]. This function takes as inputs the values of the dependent variable, that is, the variables anxiety, happy, fear and stress separately, the variable composed of the defined groups and the method of correction of the p-value, which was the Bonferroni correction ('*bonferroni*'). In cases where the p-value of a given combination of groups is less than or equal to the significance level of 5% then the groups composing that pair are considered to be statistically different.

4.3 Pre-processing and Feature Extraction

The pre-processing and feature extraction process are fundamental steps to obtain relevant information from the data.

Given that there was already a pre-processing associated with this database in [23], in this work the same procedure was used, which is described in figure 4.3.

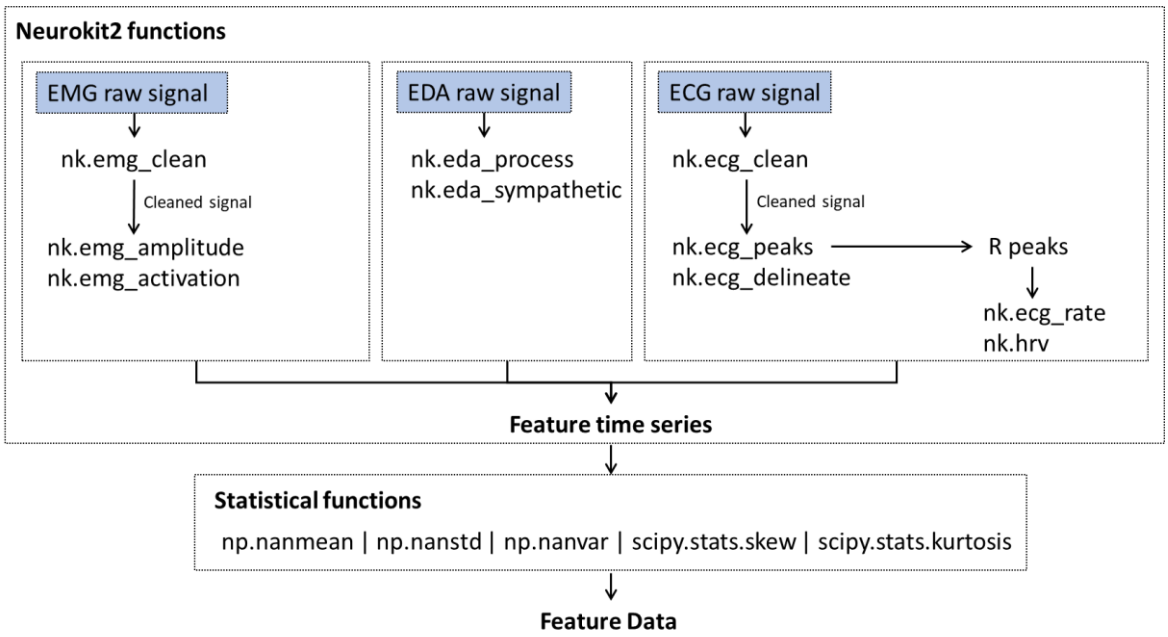

Figure 4.3: Data Preprocessing and Feature Extraction methodology used for each emotion condition and participant. Based on [23].

To preprocess the physiological signals it was used NeuroKit2, that is a Python library specifically designed for the analysis of physiological signals [23]. It provides a set of tools and functions to preprocess and analyze those signals. In this case for the preprocessing of EMG, EDA and ECG signals were used the functions *nk.emg_clean*, *nk.emg_amplitude*, *nk.emg_activation*, *nk.eda_process*, *nk.eda_sympathetic*, *nk.ecg_clean*, *nk.ecg_peaks*, *nk.ecg_delineate*, *nk.ecg_rate*, and *nk.hrv*, as shown in figure 4.3 [23]. Preprocessing each signal individually allows extracting features associated with a particular time series, except for HRV, for three features associated with EDA and one with EMG [23]. Disregarding the exceptions mentioned, for a given time series statistical metrics were extracted, namely mean, standard deviation, variance, skewness and kurtosis, based on the NumPy and SciPy module and on the statistical functions declared in figure 4.3 [23]. At the end of this process a set of features associated with each physiological signal (EMG, EDA, ECG), emotional condition (Baseline, Fear, Happy, Neutral) and participant was obtained.

4.4 Feature Selection

Initially, we intended to use principal component analysis in order to transform the initial variables into another set of independent variables consisting of linear combinations of the initial set. As principal component analysis assumes the existence of correlation between the initial variables, methodologies A and B, described in detail below, were used since they allow keeping relatively high correlation values between features and at the same time reduce the dimensional space of the features by removing redundant data. The principal components analysis also allows eliminating redundancy and obtaining principal components that consist of linear combinations of

the initial set, possibly containing "richer" information. This is why it was thought of using principal components as input to the emotional identification model.

As an alternative hypothesis, the use of methodology C and the features directly as model input, arises. Methodology C aims to remove existing redundancy and retain features to directly incorporate the classification model, without going through the principal components analysis. In addition to being less complex, it also makes it easier to later interpret the model results arising from its use.

In summary, three different methodologies (A, B and C) were considered to perform the feature selection.

4.4.1 Methodology A and B

Methodology A is based on the correlation matrix, in which the pairs of features that have the maximum correlation per line are selected, if that value was higher than 0.5. Within each selected pair, the variable that has a higher average correlation with all variables is removed. This methodology makes it possible to remove values practically equal to one, which are associated with extremely redundant data, but on the other hand it does not rule out the assumption of principal component analysis concerning the existence of correlation in the data considered.

The methodology B is a methodology used in [23], which was considered on the same database. The features selected based on that methodology were used to obtain the correlation coefficients between them. Then a correlation threshold of 0.8 was defined, allowing the detection of pairs of features with correlation coefficients equal to or greater than this value. In these pairs of highly correlated features is made the elimination of one of these variables, based on the p-value of a univariate statistical analysis (RM ANOVA or Friedman) [23]. The feature with the lowest p-value is the one that remains [23].

These methods aim at selecting a set of features to integrate the principal component analysis.

T-test: Paired two sample for means:

To compare these methodologies, the t-test for paired two samples was considered. This statistical test aims to check if the means of two paired samples are significantly different from each other or not [48].

T-test for paired two samples null hypothesis states that the means of the two samples are equal [48]. It calculates the t statistic and with the degrees of freedom, the p-value associated with that test statistic is obtained. If the p-value is less or equal to the significance level, the null hypothesis is rejected, indicating that the means of the two samples are significantly different. On the other hand, if the p-value is greater than the significance level the null hypothesis is not rejected.

4.4.2 Methodology C

Methodology C is based on the correlation matrix, in which are selected the pairs of features that have a correlation greater than 0.50. Within each selected pair, the variable that has a higher average correlation with all variables is removed.

There are two approaches addressed in this work. The approach C_1 and the approach C_2.

In approach C_1, methodology C is applied over each emotion separately and, later, a union of the selected features in each case is made, generating a final matrix.

On the other hand, the approach C2 consists in merging the feature values of the 3 emotional states (happy, fear, neutral) in a general matrix, over which the methodology C is applied.

4.5 Principal Component Analysis

The principal components analysis is based on the establishment of new uncorrelated variables, called principal components, which are the result of linear combinations of the original correlated numerical variables with the objective of explaining the data [29], [31]. Figure 4.4 illustrates this transformation of the original variables in the principal components in terms of tabular indices.

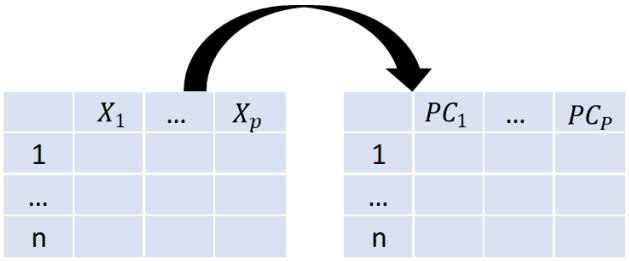

Figure 4.4: Transformation in terms of tabular indices resulting from the transformation exercised by the principal components analysis.

This method assumes the consideration and use of numerical multivariate data. Although the number of new variables, that is, of principal components is equal to the number of original variables, the transformation performed allows an attribution of relevance between variables, evaluated in the form of explained variability, in a sequential manner [29], [31]. This allows the reduction of data dimensionality, through the consideration of a number of principal components sufficiently capable of explaining a large part of the variability of the data, and an interpretation of the data that was not visible a priori through the consideration of the original variables. It is important to mention that the principal components do not have an explicit meaning [29], [31].

Principal components can be used in the establishment of a set of more "relevant" original variables, as a way of reducing data dimensionality, as previously mentioned, in the construction of graphs, including the so-called "biplots". The biplot consists of a representation of the observations, in the form of points, and the variables in the form of vectors, on the same graph, enabling data visualization and information gathering [29]–[31].

Considering the existence of p original variables $X_1, \dots, X_p$, composing the vector X ($X = [X_1 \dots X_p]'$, where $'$ means the transpose of the matrix under consideration), the principal components (PCs) are defined as [29], [31]:

$$PC_1 = a_{11}X_1 + \dots + a_{p1}X_p \quad (4.10)$$

...

$$PC_p = a_{1p}X_1 + \dots + a_{pp}X_p. \quad (4.11)$$

In a way that:

$$Var(PC_1) >= \dots >= Var(PC_p), \quad (4.12)$$

$$Cov(PC_i, PC_j) = 0, \forall\, i \neq j. \quad (4.13)$$

And where the vector $a_i = [a_{1i} \dots a_{pi}]'$, where $'$ means the transpose of the matrix, composed of the coefficients of the i-th principal component PCi, is the eigenvector with the i-th largest eigenvalue (λi), with:

$$\lambda_i = Var(PC_i) = Var(a_i'X) = a_i'Var(X)a_i. \quad (4.14)$$

The total variance of X is equal to the total variance of the principal components [29], [31]:

$$Var(PC_1) + \cdots + Var(PC_p) = \lambda_1 + \cdots + \lambda_p. \quad (4.15)$$

The proportion of variance of X explained by the i-th principal component PC_i is equal to $\frac{\lambda_i}{\sum_1^p \lambda}$ [29], [31]. There is always the question of how many components should be retained. There is no ideal or definitive answer to this question. However, there are some methodologies that can help with this question. In fact, we can even use the methodologies already discussed earlier in the factor analysis section, about how many factors should be retained. In this case, we can look precisely at the proportion of variance explained by each principal component and keep a number of PCs that allow obtaining an acceptable proportion of explained variance, that is, relatively high, in order to obtain the smallest number of principal components and at the same time bring this proportion close to 100%.

The relevance of an original variable, X_k, in the construction of a principal component, CP_i, is measured through Pearson's correlation coefficient between them [29], [31]:

$$cor(CP_i, X_k) = \frac{a_{ki}}{\sqrt{Var(X_k)}} \sqrt{\lambda_i}. \quad (4.16)$$

In the absence of knowledge about the population data, the sample is used, to estimate the principal components [31]. In graphical terms what happens is a rotation of the axes, where the direction of the axis corresponding to the first principal component minimizes the sum of the squares of the distances, in the perpendicular, of the points (data) to the axis [31]. The representation of the data transformed by this method on two selected principal components may allow discovering groups of similar data, unusual observations, among others [31]. It should be noted that the axes corresponding to the principal components are perpendicular to each other [31].

4.6 Classification

4.6.1 Support Vector Machine (SVM)

Support Vector Machine (SVM) is a machine learning algorithm that can be used for linear and non-linear classification, regression problems and outlier detection [49].

The main goal of SVM is to find the optimal hyperplane in an N-dimensional space, with N being the number of features [49], [50]. The optimal hyperplane is a hyperplane that best differentiates the data points from different classes with the maximum possible margin, as shown in figure 4.5 on the left-hand side [49], [50]. The margin is the distance between the nearest data points from different classes [49], [50]. By maximizing the margin, SVM aims to find the most robust hyperplane that generalizes well to new data, reducing the chance of misclassification [49], [50].

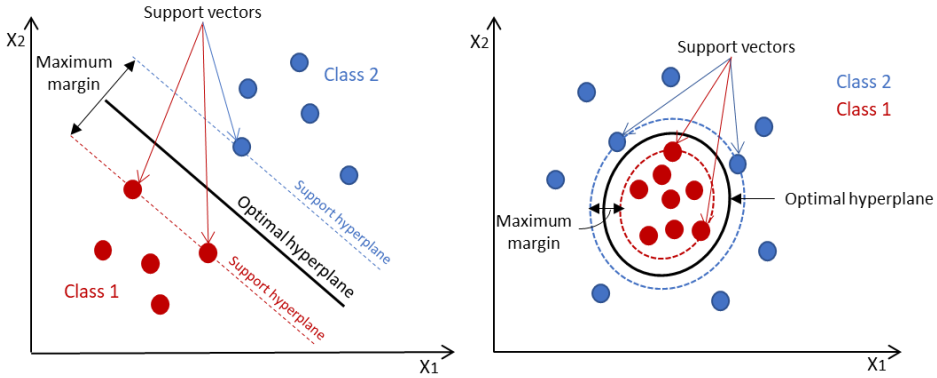

Figure 4.5: Linear (on the left) and non-linear (on the right) SVM for 2-Dimensional (2D) dataset.

The orientation and positioning of the hyperplane is determined by the support vectors [49], [50]. The support vectors are data points closer to the decision boundary (hyperplane) between the classes [49], [50]. If they are moved the margin of the hyperplane changes and if they are deleted the position of the hyperplane will change [49], [50].

The original SVM can only deal with linearly separable cases, that is when the data points can be separated by a straight line, as shown in the figure 4.5 on the left side [49], [50]. If the data is not linearly separable, that is when the data points cannot be separated by a straight line, as shown in the figure 4.5 on the right side, SVM uses a kernel function to transform the input data into a higher dimensional space where it becomes separable, and a linear decision boundary can be found [49], [50]. The kernel function can be written as: $k(x,z) = \varphi^T(x)\varphi(z)$, that performs simultaneously the mapping and dot product and where φ is the feature map that maps the input features into the higher dimensional space [49], [50]. SVM can use various kernel functions, such as linear, polynomial, radial basis function (RBF), and sigmoid [49].

There are two types of classification, the binary and the multiclass classification [49].

In its most simple type, SVM algorithm does not support multiclass classification since it is a binary classifier [49]. However, the multiclassification problem can be divided into numerous binary classification problems. In that way, there are two different approaches:

1) One-vs-Rest, where binary classifiers that distinguish between one of the labels and the rest are built [49].

2) One-vs-One, where binary classifiers that distinguish between every pair of labels are built [49].

In this work was considered the One-vs-One approach. This scheme involves training $c(c-1)/2$ binary classifiers, with c being the number of classes [49]. For example, if there are three classes (Fear, Happy, and Neutral), SVM will train three binary classifiers: Fear vs. Happy, Happy vs. Neutral, and Fear vs. Neutral. Each binary classifier learns to distinguish between the two classes given classes and outputs a decision function [49]. When a new object is tested, it follows the max-wins voting strategy, which means that every classifier gives a predicted class and the class with the greatest number of votes from the binary classifiers is chosen as the final prediction [49].

To apply this scheme in this work it was used the *sklearn.svm.SVC()* [51], from the module Scikit-learn. With this function it is possible to control and optimize some key parameters that have a big impact on the result.

4.6.2 Datasets and splits

In this work several approaches were considered, and consequently different classification models were defined.

One of the aspects that provides different approaches is the use of principal components, obtained through the principal components analysis on features selected according to methodology A or B described in section 4.4.2, for each case, or the use of features selected according to methodology C directly as input to the classifier. In this case, where features are considered directly as model input and selected according to methodology C, there are two different approaches for gathering the information associated with each emotional state, as described in section 4.4.2.

Another aspect is the consideration or not of the division of each physiological signal of the participants in five excerpts, each one with two minutes duration. In cases where the division is not verified, for each emotion there is a set of features or principal components obtained from features calculated over the ten minutes of each physiological signal considered. In the cases where the division is verified, the features mentioned above are calculated over the two minutes that compose each signal excerpt.

Based on each approach and their combination, 6 types of datasets were built.

The datasets 1, 2 and 3 follow the organizational model represented on the left side in figure 4.6. Dataset 1 makes use of the principal components obtained through principal component analysis on the selected features according to methodology A or B described in section 4.4.1. Dataset 2 makes use of the features generated from the approach C_2 described in section 4.4.2. Dataset 3 makes use of the features generated from the approach C_1 described in section 4.4.2. In this sense datasets 2 and 3 differ from dataset 1 in the matter of consideration of features instead of principal components and dataset 2 and 3 differ from each other in the features selected.

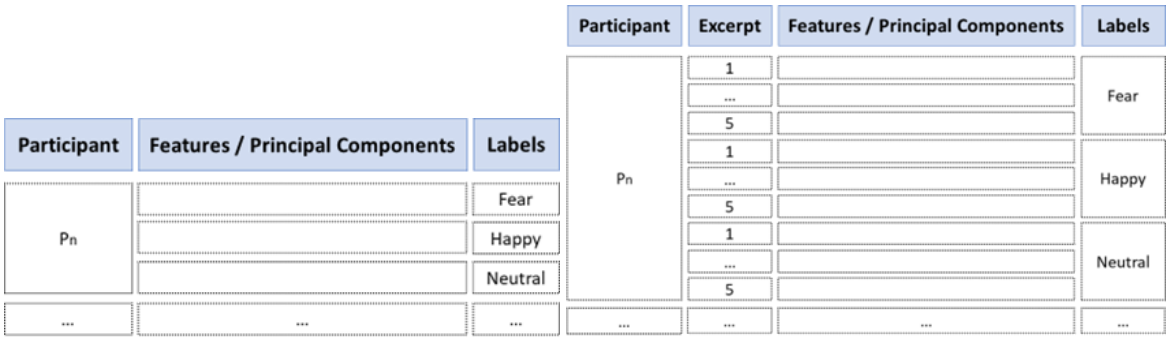

Figure 4.6: Base model of data organization in the cases of not considering the division of the complete signal into excerpts (left) and considering the division into excerpts (right).

Datasets 4, 5 and 6 follow the organizational model on the right side of the figure 4.6.

These differ from datasets 1, 2 and 3, respectively, only by considering the division of each physiological signal of each participant into 5 excerpts, each two minutes long. In the case of datasets 1, 2 and 3, in which the division does not occur, there is, for each emotion, a set of features or principal components selected or obtained from features calculated over the ten minutes of each physiological signal considered. In the case of datasets 4, 5 and 6, where the division is verified, the said features are calculated over the two minutes that make up each excerpt of the signal.

Each of these datasets contains 16 variations. For the global case, i.e. the case containing all participants (Global), and for each group defined (G1, G2 and G3), the formation of each dataset, described above, is considered for the case in which the features associated with each of the three physiological signals (EMG, EDA and ECG) are considered together and for the case in which the

features associated with those signals are considered separately. In datasets 1 and 4, in which the principal components are considered, these variations also occur, since PCA is implemented over the set of features associated with EMG, EDA and ECG separately and over the set formed by the features of these three signals together.

To split the data into the training and test sets, since there were already splits tested on this database and on similarly constructed datasets in [23], splits 2 and 3 described there were considered in the scope of this work.

On datasets 1, 2 and 3, a split was performed based on split 2 that in the case of considering the features generated on the complete signals, gave better results among those tested on the same conditions in [23]. In this split a division of the participants is made by the training (75%) and test (25%) sets, each of them composed of different participants. Each participant being represented by the three emotions (Fear, Happy and Neutral).

On datasets 4, 5 and 6 were considered a split which follows the same construction philosophy as split 3. In the training set four excerpts are used for each emotion associated with each participant, and in the test set the remaining excerpt is used in each case. The choice of excerpts for each set is random.

4.6.3 Standardization

SVMs take into consideration the distance between the hyperplane and the closest data points from different classes. Once the distance is sensitive to the scale of the features, there is a need to resort to standardization. Moreover, when the features have different scales, some of them will be seen as the dominant ones and tend to have more impact when training the machine learning model, while the others will have very little influence. Standardization ensures that all features are on the same scale so that they contribute equally, improving the performance of machine learning models.

There are several methods of standardization. The most used ones were discussed in section 4.1.1. The choice of standardization method depends on the type of data, the presence of outliers and on the machine learning algorithm used [52]. By performing normality tests on each final dataset formed, including all the variations mentioned above in section 4.6.2, it was found that the features did not follow a normal distribution. Through the construction and analysis of boxplots it was found that there were some outliers associated with a small number of features. These outliers may simply correspond to a more pronounced response generated by the imposed emotional stimulus and may even be associated with a region where the maximum stimulus peak for that participant occurred. Given this, the Min-max method was chosen and applied through the function *sklearn.preprocessing.MinMaxScaler* from Scikit-learn [51], [52].

4.6.4 Feature optimization

In the cases where the principal components are considered as input to the model, the feature optimization process consists only in using, in each case, the minimum number of principal components that present a percentage of accumulated explained variance equal or superior to 90%.

In the cases where features are considered as model input, the recursive elimination with cross-validation (RFECV) was used as feature optimization method.

Recursive Elimination with Cross-Validation (RFECV) is a feature selection technique in machine learning that aims to identify the most relevant features in a dataset that optimize the model's

performance, by recursively eliminating the least important features based on cross-validation. This is an additional feature selection method that complements the methodology C described on section 4.4.2. It allows to select the features, among that one's previously selected based on methodology C, that optimize the SVM model constructed and have more influence over the emotion recognition [53].

This method starts by fitting the SVM model on the datasets built after the feature selection based on methodology C. Then he ranks the importance of each feature based on its impact on the performance of the model and removes the least important feature and fit the model again. This process is repeated until the optimal number of features is reached. It performs a cross-validation on each subset of features to estimate the generalization error of the model. Finally, the subset of features with the best performance based on the cross-validation scores, is chosen [53].

To implement this method in this work was used the function *sklearn.RFECV()* [51]. from the module Scikit-learn. In this case, this function takes as input a machine learning estimator, the Support Vector Machine classifier, a scoring function, the accuracy, and as cross-validation splitting strategy, the StratifiedKFold, once this work involves a multiclass problem [51].

4.6.5 Hyperparameters Optimization

The performance of SVMs can change according to the hyperparameters chosen, which are parameters that need to be set before training the model. The following hyperparameters have a critical influence on the SVM's performance:

- **C:** The C hyperparameter controls the tradeoff between achieving a low training error and a low testing error. A lower value of C reduces the risk of overfitting but allows more misclassifications. A higher value of C results in the opposite [54].
- **kernel:** The kernel hyperparameter determines the shape of the decision boundary (optimal hyperplane). The kernel functions that stand out are the linear, polynomial, and radial basis functions [54].
- **gamma:** The gamma hyperparameter controls the shape of the decision boundary for non-linear kernels. A smaller value of gamma leads to a smoother decision boundary, while a larger value of gamma leads to a more complex decision boundary [54].
- **degree:** The degree hyperparameter is used in polynomial kernels to specify the degree of the polynomial function. A higher degree polynomial function can better fit complex data, but it can also lead to overfitting [54].

Due to these parameters influence over the SVM model performance, there is a need to appeal to the hyperparameter optimization, which is the process of searching for the best combination of hyperparameters for the machine learning model considered. The process of hyperparameter optimization involves a definition of the range of hyperparameters to search over, which can be defined manually or automatically by using a search algorithm. Since this work involved the construction of different models and is intended to eliminate the impact of a manual choice, even if it is correctly thought out, it was used an automatically method to define the best combination of hyperparameters for each model developed. That method was the Grid search.

Grid search technique searches over all possible combinations of hyperparameters in a predefined range and evaluates them to find the optimal one.

To apply this method in this work was used the function *sklearn.model_selection.GridSearchCV()* function, from the module Scikit-learn [51].

This function takes an estimator, that was the *sklearn.svm.SVC()*, a grid of hyperparameters that is defined as a dictionary of hyperparameters, that includes C values, kernel functions and gamma values, a cross-validation generator to split the data into training and test sets and a performance metric [51].

For each combination of hyperparameters, this method fits the model on the training data and evaluate its performance on the test data using the chosen performance metric. It evaluates all possible combinations of hyperparameters and chooses the best combination, that is the one that maximizes the performance metric. Then with that best combination this method refits the model on the entire training set using the chosen hyperparameters, building the final model.

4.6.6 Model Evaluation Metrics

The confusion matrix is a tool used in classification problems and/or evaluating the performance of classifiers. This matrix allows the perception of errors and limitations of the models built [55]. In this case of emotional recognition, the confusion matrix is a 3x3 matrix, which allows us to verify if the identification of each emotion is done correctly, that is, if the predicted emotion (Predicted Class) actually corresponds to the emotional stimulus to which each participant was subjected (True Class).

In order to obtain performance evaluations of algorithms, performance evaluators are used. In this document, accuracy, precision, recall and F1-score are considered. Accuracy is an intuitive metric and is used in the general evaluation of an algorithm's performance. However, it is a metric that can be easily misleading since it does not distinguish between true positives and true negatives. Which means that you can get a high accuracy and have a low number of true positives and a high number of true negatives, not working well for drawing important conclusions about the system. Furthermore, when the database is unbalanced, this metric usually does not represent the performance of the algorithm properly. In this sense, precision, F1-score, and recall are used to reinforce and draw reliable conclusions about the results obtained. Recall corresponds to the fraction of examples that are correctly classified as positive among the total of those that are actually positive. Precision evaluates the amount of examples correctly classified as positive over the total of those predicted as positive. The F1-score combines model precision and recall, evaluating the relationship between them [56]. The formulas used for these metrics can be found in [56].

4.7 Proposed Method

This study was planned so that, based on the review of the state of the art, a biometric signature of emotion will be established based on the questionnaires and the physiological signals collected, using statistical analysis. This is intended to be done, firstly, using Factor Analysis, or alternatively, Cluster Analysis, with the purpose of grouping participants based on the questionnaires completed before exposure to a given emotional stimulus and, secondly, grouping them based on the features selected from the baseline of the physiological signals collected. This process will result in a set of n groups (G1, ..., Gn) as well as the Global group, which is the group composed of all individuals.

In figure 4.7 the block diagram of the proposed methodology is presented. The physiological signals will be pre-processed and features will be extracted from them. These features will be subject to a selection process through methodologies A or B, discussed in section 4.4.1, and methodology C discussed in section 4.4.2.

At this stage, different analyses will be performed following the same structure, but using different combinations between groups of participants formed and sets of features associated to different physiological signals. For each combination, the procedure described below is followed.

From the features selected from methodology A or B, we will then resort to Principal Component Analysis in order to obtain the optimal principal components, i.e., those that in the smallest possible amount, allow to explain at least 90% of data variability. On the other hand, from the features selected by methodology C, RFECV will be applied, which, considering the SVM model, will allow finding the optimal features, i.e., those which optimize the performance of this model.

Considering the participants and the optimal principal components or optimal features, the appropriate splits, discussed in section 4.6.2, will be applied, dividing the data into training and test sets. In a first stage, by fitting the data, a loop (30 runs) will be run, with cross validation, in order to obtain the best combination of hyperparameters. These hyperparameters are called optimal hyperparameters and consist of parameters of the SVM model that in the set of 30 runs will allow obtaining maximum accuracy. Subsequently, based on these optimal hyperparameters, a new fit to the data will be performed, without cross-validation, allowing the construction of the optimized model.

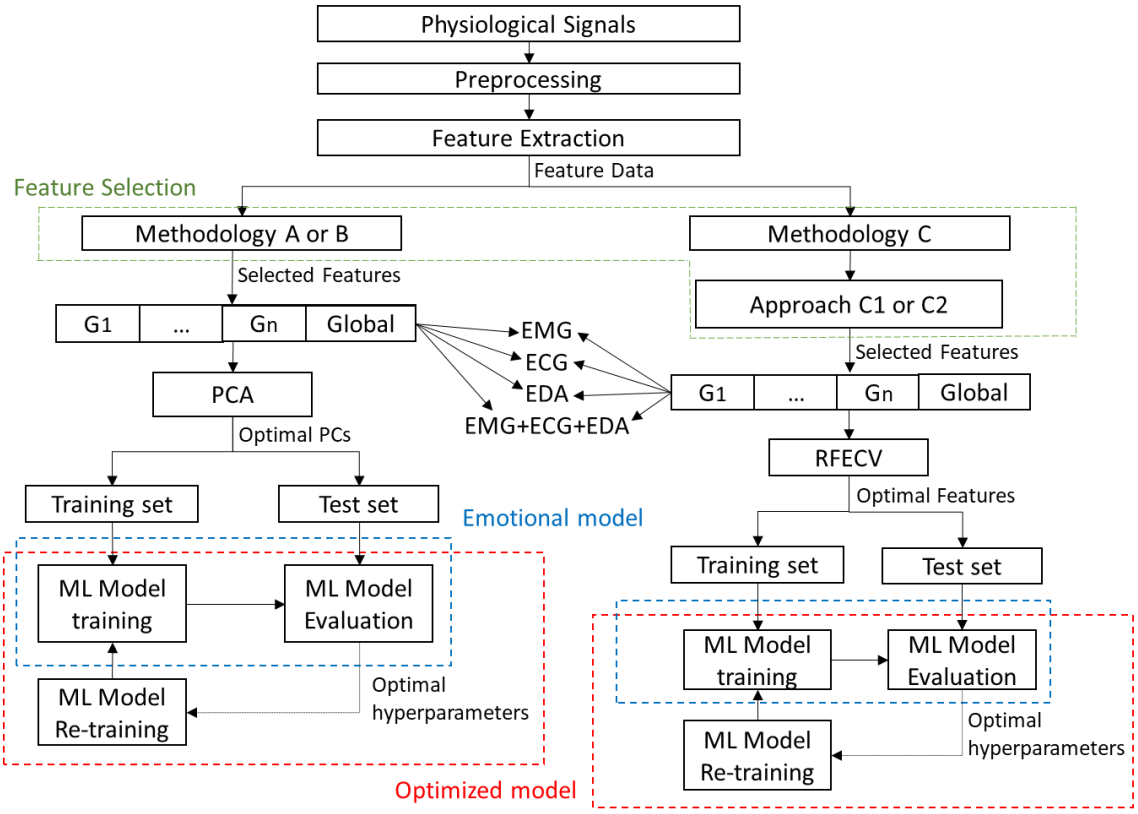

Figure 4.7: Block diagram of the proposed method.

5 Results and Discussion

5.1 Results of the defined groups

The emotional profile definition implies the study and description of the questionnaire's answers. Therefore, a factor analysis was performed on the trait and state questionnaires and on the percentages of anxiety, fear, stress and happy reported by each participant in the VAS-Pre questionnaire, with the purpose of verifying the possibility of establishing groups based on the scores associated with the factors generated in each case.

5.1.1 Trait and state questionnaires

In order to evaluate the adequacy of the factor analysis model to the data, the Bartlett sphericity test was used. In relation to these questionnaires, for session 1, it was performed the Bartlett sphericity test (p-value < 0.0001 (State), p-value < 0.0001 (Trait), significance level = 0.05) which showed the adequacy of the analysis, for both questionnaires. However, the Kaiser Meyer-Olkin measure of sample adequacy, KMO = 0.44 (State) and KMO = 0.44 (Trait), indicates that the analysis may not fit the data of these questionnaires. This is because it means that there are several variables with reduced KMO (<0.5) which should be disregarded from the study. In order not to remove most of the variables, it was initially decided to proceed with an analysis over all of them.

Scree plots were created for each case, and using the elbow rule, 6 factors were considered in the case of STICSA-State and 7 factors in the STICSA-Trait questionnaire.

A factor analysis was implemented over the data using the principal components method. The scores, for each factor, were obtained. It was verified that it was possible to associate the participants through scores, positive or negative and more or less high between the positive ones.

However, it was not possible to evaluate the groupings based on the scores and therefore to verify if there was any meaning behind them. Thus, a cluster analysis was performed, using the complete linkage method and the Euclidean distance or the K-means method, to try to validate the clusters established on the basis of the factor analysis.

It was possible to verify that some individuals belonging to the same group formed from the cluster analysis, present, similarities in terms of factor scores.

It is important to point out that the best results, in relation to the STICSA-Trait questionnaire and STICSA-State questionnaire, were obtained for the complete linkage method, since it allowed to obtain groups with a larger average silhouette width, as we can see in table 5.1. Using STICSA-State questionnaire the average silhouette widths per group were better than those obtained with the STICSA-Trait questionnaire. This means that the groups formed based on STICSA-State questionnaire have better structure than those formed based on STICSA-Trait questionnaire.

Method	Nº of Groups	STICSA - Trait (Session 1)				STICSA - State (Session 1)			
		Mean Silhouette	Average silhouette width			Mean Silhouette	Average silhouette width		
			Group 1	Group 2	Group 3		Group 1	Group 2	Group 3
Complete Linkage	3	0.10	0.05	0.05	0.17	0.28	0.22	0.11	0.33
K-means	3	0.07	0.03	0.02	0.16	0.128	0.06	0.03	0.22

Table 5.1: Silhouette results from the STICSA – Trait and STICSA – State questionnaires, for session 1, using complete linkage and K-means methods and considering 3 groups per method. Note: The composition of each group (1, 2, and 3) varies from method to method.

The same analysis was carried out on the data from session 2, with similar values but with differences in terms of the composition of the groups formed in terms of the participants in them.

5.1.2 VAS-Pre questionnaire

Additionally, an analysis was carried out on the visual analogue scale questionnaires (VAS-Pre questionnaire), in which the variables anxiety, happy, fear and stress were considered. It was performed the Bartlett's sphericity test (p-value < 0.0001 (Session 1), p-value < 0.0001 (Session 2), significance level = 0.05), which showed the adequacy of the analysis for both sessions. The Kaiser-Meyer-Olkin measure of sample adequacy, KMO = 0.69, at session 1, and KMO = 0.62 at session 2, reinforced that the analysis may fit the data of these questionnaires, respectively.

A factorial analysis was carried out on the data of these questionnaires, for both sessions. Using the Scree plot and the elbow rule, the use of two factors was considered. The principal components method was considered.

As in the previous case, cluster analysis was also applied to the data from this questionnaire. It was immediately verified, that there was an increase in terms of the value of the internal validation measure (average silhouette width, in this case) compared to the values obtained from the trait and state questionnaires, as can be seen through the comparison between tables 5.1 and 5.2. This led to the conclusion that the use of visual scalar aid allows easier and more truthful answers and, consequently, a better definition of the clusters. In other words, it allows the formulation of groupings with a more reasonable structure.

Method	Nº of Groups	VAS-Pre (Session 1)				VAS-Pre (Session 2)			
		Mean Silhouette	Average silhouette width			Mean Silhouette	Average silhouette width		
			Group 1	Group 2	Group 3		Group 1	Group 2	Group 3
Complete Linkage	3	0.37	0	0.36	0.45	0.32	0.20	0.28	0.37
K-means	3	0.379	0	0.36	0.41	0.32	0.20	0.30	0.38

Table 5.2: Silhouette results from the VAS-Pre questionnaire for both sessions, using complete linkage and K-means methods and 3 groups per method, considering the variables Anxiety, Happy, Fear and Stress. Note: The composition of each group (1, 2, and 3) varies from method to method.

It was verified through the analysis of session 1 and session 2 of the VAS-Pre questionnaire, that there is a grouping composed of a large majority of participants who, in general, remain in the same grouping. When considering three clusters, one group is composed of a large majority of participants and the remaining two groups vary in terms of participants that compose them from session to session. Some participants vary from session to session group. This leads to the conclusion that we should consider the groupings as groups with their own characteristics and not as a set of participants who must remain fixed in that group regardless of the day they answer the questionnaires. That's because between days, factors such as anxiety, happy, fear and stress will vary, due to several conditions. In that way, it's normal that participants vary from group, between sessions. This means that people respond differently to questionnaires and stimuli on different days. This could be due to several factors such as people's reactions being different to each stimulus, people's states varying over time, some people being more emotionally stable than others, among others.

Considering the analysis using STICSA-Trait, STICSA-State and VAS-Pre questionnaires, in both sessions, 1 and 2, factor scores seem to allow the groups definition. This is validated by the groupings evaluation based on cluster analysis: individuals belonging to the same group have similar scores, as shown in figure 5.1. However, that just verifies itself for the factors with highest

percentage of variance explained. That is, the scores associated with factor 1 allow for a more correct definition of groups than the subsequent factors. In the same way, the scores of factor 2 allow defining more meaningful groups than the subsequent factors, but worse than based on those of factor 1. Furthermore, when considering the use of more than one factor to calculate the scores, the disuse of rotation or the consideration of quartimax rotation and subsequent use of the scores of factor 1 to form the groupings, proved to be a little better than the use of varimax rotation, in terms of the validation of the groups formed with cluster analysis and graphical analysis.

As can be seen in figure 5.1, when considering the formation of three clusters, we found that one of them is associated with a positive and higher score in module, as well as with higher levels of anxiety and stress, and in most cases, fear, than the remaining clusters. Another cluster generally presents reduced positive scores and/or scores close to 0, as well as intermediate percentages of anxiety, stress and fear. The remaining cluster is generally composed of participants with negative scores and/or scores close to 0, and low levels of anxiety, fear and stress.

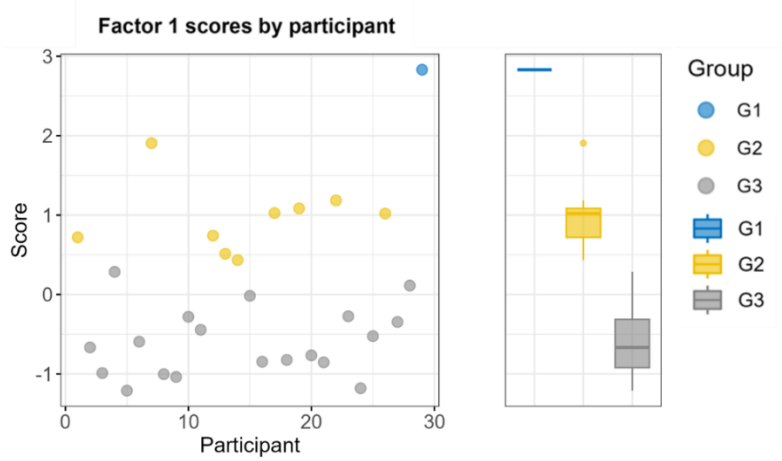

Figure 5.1: Representation of the scores associated with factor 1, generated from the factor analysis in the VAS-pre data of session 1, corresponding to each participant, belonging to the respective group defined by cluster analysis on that data (on the left) and the respective variation of scores per group (on the right).

Comparing the results obtained, it was found that the consideration of the VAS-Pre questionnaire allowed for higher average silhouette width values, and consequently more stable groups. This is associated with an improvement in the discrimination of the groups according to the factor scores obtained from the factor analysis, reinforcing the possibility of establishing groups according to factor analysis.

Factor analysis seems to allow the formation of groups with some apparent meaning, within the previously mentioned conditions in terms of factors, based on the comparison of scores between participants belonging to the same grouping defined by cluster analysis. However, there is no way to literarily validate such conclusions or to generalize them and there is no clear rule with which a separation between the groups could be precisely defined. In this sense, we proceeded to further analysis with the use of cluster analysis as a method of defining groups of participants. In addition, only one of the sessions was assumed from here on, in this case session 1 since it is the one where the answers to the questionnaires are less biased.

Cluster analysis was also implemented on the data associated with each of the physiological signals, EMG, EDA and ECG, on the baseline, as an attempt to validate the groups of participants formed from the questionnaires. However, when comparing the groupings made from the physiological signals and the questionnaires, reduced proportions of similarity were found in terms of their constitution. The majority of coincidences of participants between the groups appeared only among those made up of the majority of participants. In this sense, the validation was not verified.

Thus, a different analysis was considered to define with which groups the rest of this work would proceed. Taking into account that the VAS-Pre questionnaire allows easier and more truthful answers, as previously concluded, this questionnaire was taken as ground truth. It was used as a decision factor to decide which questionnaires (STICSA-State or STICSA-Trait) or physiological signals (EMG, EDA or ECG) would allow us to obtain the most distinct groups possible, according to the values of anxiety, fear, stress and happy reported in the VAS-Pre questionnaire.

In this sense, based on the groups formed, based on cluster analysis, from each physiological signal, at baseline, and from STICSA-State and STICSA-Trait questionnaires and on the percentages of anxiety, happy, fear and stress reported in the VAS-Pre questionnaire and associated to each of the participants, a dataframe was constructed and organized according to the scheme shown in the figure 5.2. Three groups of participants were defined for each of the cases. In this sense, each participant was assigned a number (1, 2, or 3) for each physiological signal and each questionnaire, corresponding to the group in which they were placed. The groups labeling has not a direct meaning, i.e., the number attributed may vary between groups assessment (ECG, EDA, EMG, State and Trait questionnaires).

					Group Number (1, 2 or 3)				
Participant	**Anxiety (%)**	**Happy (%)**	**Fear (%)**	**Stress (%)**	**ECG**	**EDA**	**EMG**	**State questionnaire**	**Trait questionnaire**
1									
...									
29									

Figure 5.2: Representation of the data used to build the boxplots and analyze the existence of significant differences between the groups of participants (Group 1, 2 or 3) for each signal or questionnaire.

Based on the data in figure 5.2, several boxplots were constructed and are shown below.

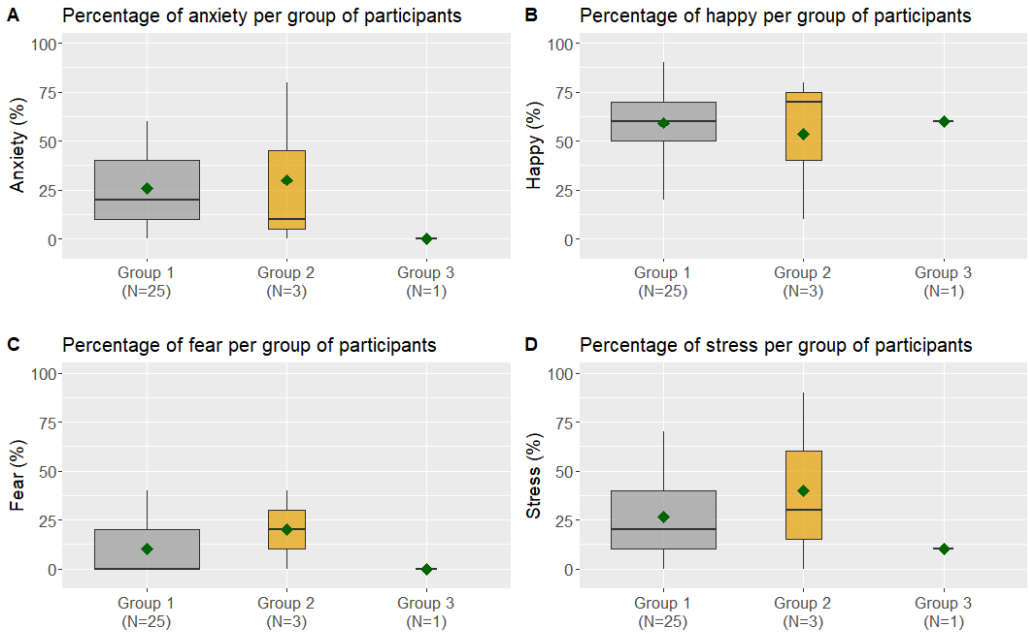

Figure 5.3: Comparison of the percentages obtained in the VAS-Pre questionnaire (Session 1), associated with the variables anxiety, happy, fear and stress, according to the groupings of participants made based on the cluster analysis on the features retained from the EMG signals (Session 1).

As shown in figure 5.3, It is not possible to visually observe a distinction of the groupings made by the EMG signal features in terms of the percentages of anxiety, fear, happy and stress reported in the same questionnaire. The results of the One Way ANOVA and the Kruskal-Wallis test on table 5.4 reinforce the absence of significant differences between the groups defined based on the selected features of the EMG signal. In this sense, the EMG signal does not allow discriminating the groups of participants according to the decision factor, not being adequate for the definition of the groupings.

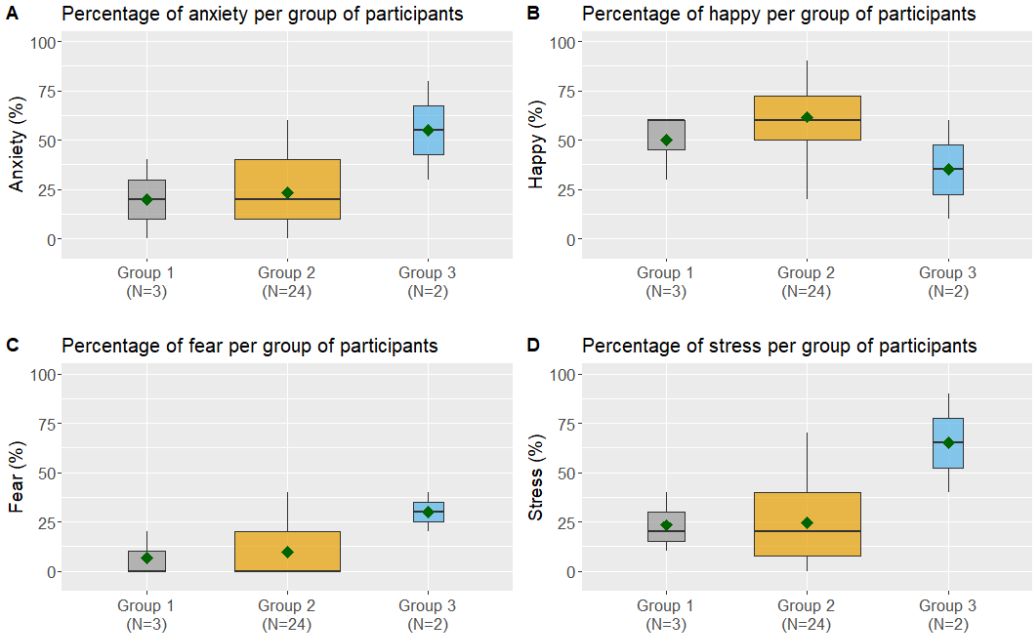

Figure 5.4: Comparison of the percentages obtained in the VAS-Pre questionnaire (Session 1), associated with the variables anxiety, happy, fear and stress, according to the groupings of participants based on the cluster analysis on the features retained from the EDA signals (Session 1).

In figure 5.4, It is possible to graphically observe a distinction of the clusters made by the features of the EDA signal, in terms of the percentages of anxiety, happy, fear and stress variables, reported in the VAS-Pre questionnaire. One group consists of the highest percentages, another of intermediate percentages, and the rest of the groups of lower percentages. The group of participants who have higher levels of anxiety, fear and stress also have lower levels of happiness, which shows consistency. However, the results of the One Way ANOVA and the Kruskal-Wallis test, shown in table 5.4, report the absence of significant differences between the groups defined based on features selected from EDA signal, in the cases of anxiety and stress. This contradicts what can be visually observed in figure 5.3 especially when Group 3 is compared to the others. However, we must take into account that we are dealing with small sample sizes, which may affect the results obtained from the statistical tests.

The better results, at least visually, associated with the EDA physiological signal, in relation to the ECG and EMG, are due to the fact that emotional stimulation correlates with electrodermal activity, since certain emotions are related to sympathetic activation, inducing changes in sweat gland activity and consequently in skin conductivity [57]. Moreover, the EDA relates directly with arousal, but the ECG does not. EDA represents only sympathetic activation. The ECG represents both sympathetic and parasympathetic activation, which may cause a lesser relation with the arousal [58]. Emotional recognition through physiological signals is established with greater association with the arousal [57], [58].

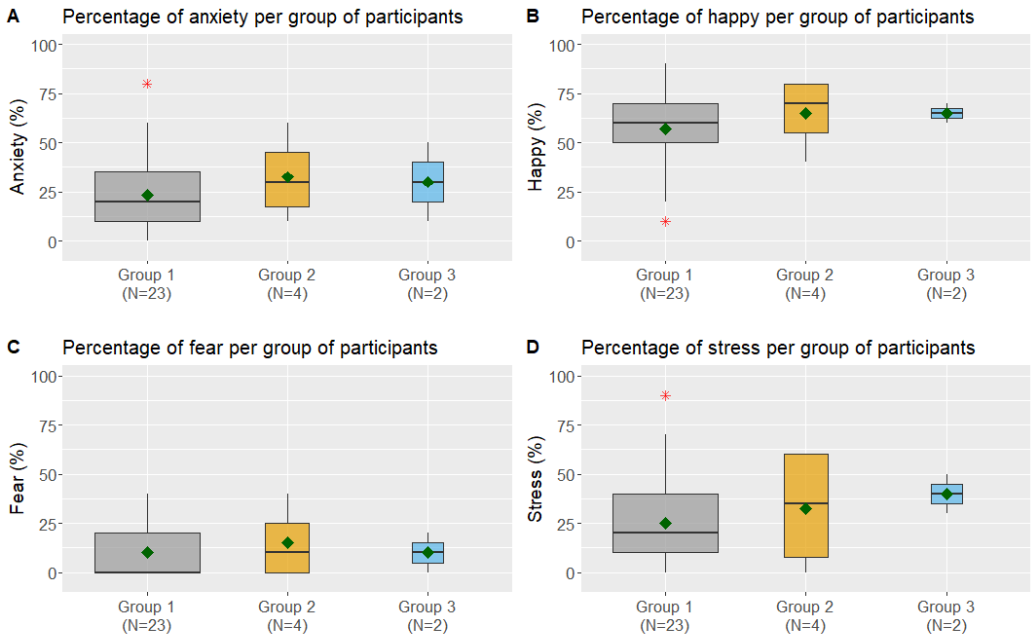

Figure 5.5: Comparison of the percentages obtained in the VAS-Pre questionnaire (Session 1), associated with the variables anxiety (A), happy (B), fear (C) and stress (D), according to the groupings of participants based on the cluster analysis on the features retained from the ECG signals (Session 1).

As can be seen in figure 5.5, It is not possible to observe a distinction of the groupings made by the features of the ECG signal, in terms of the percentages of anxiety, happiness, fear and stress, reported in the same questionnaire. The results of the One Way ANOVA and the Kruskal-Wallis test, shown in table 5.4, reinforce the absence of significant differences between the groups defined based on the selected features of the ECG signal. In this sense, the ECG signal does not allow discriminating the groups of participants according to the decision factor.

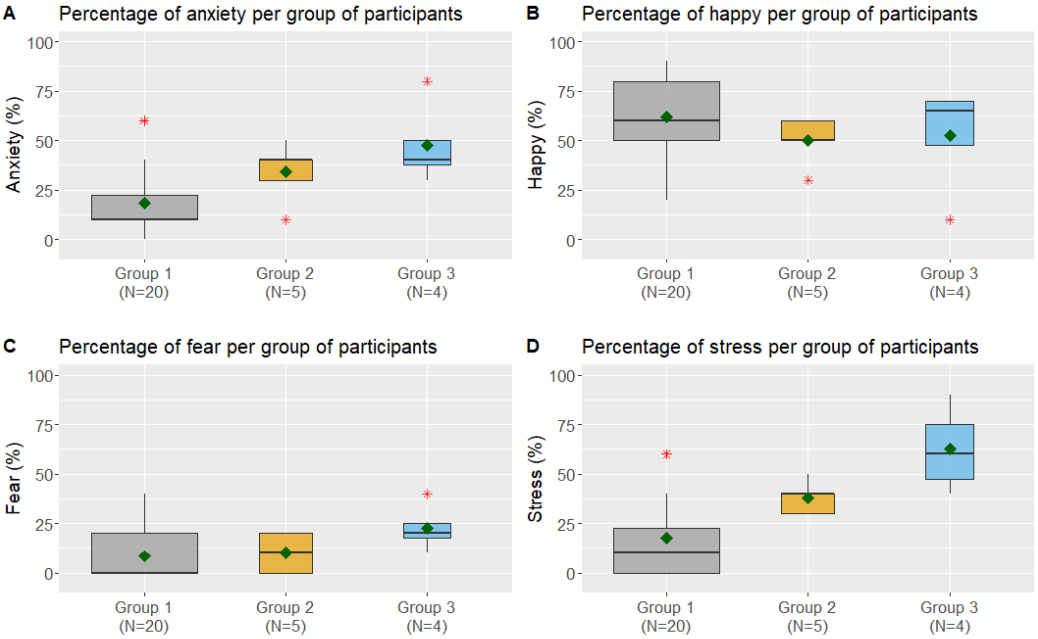

Figure 5.6: Comparison of the percentages obtained in the VAS-Pre questionnaire (Session 1), associated with the variables anxiety (A), happy (B), fear (C) and stress (D), according to the groupings of participants made based on the cluster analysis on the STICSA-State questionnaire (Session1).

As shown in the figure 5.6, It is possible to observe a distinction of the groupings made by means of the state questionnaires in terms of the percentages of anxiety, happy, fear and stress reported in the VAS-Pre questionnaire. One group is composed of the highest percentages, another of intermediate percentages, and the remaining group of lower percentages. The results of the One Way ANOVA and the Kruskal-Wallis test, presented in the table 5.4, reinforce the existence of significant differences between Group 1 and Group 3, in the case of anxiety and stress variables, defined based on state questionnaire. In the other cases that confirmation is not verified. However, we must take into account again that we are dealing with small sample sizes, which may affect the results obtained from the statistical tests. Thus, it can be seen that there is a relationship between the responses to the VAS-Pre questionnaire and the state of each participant.

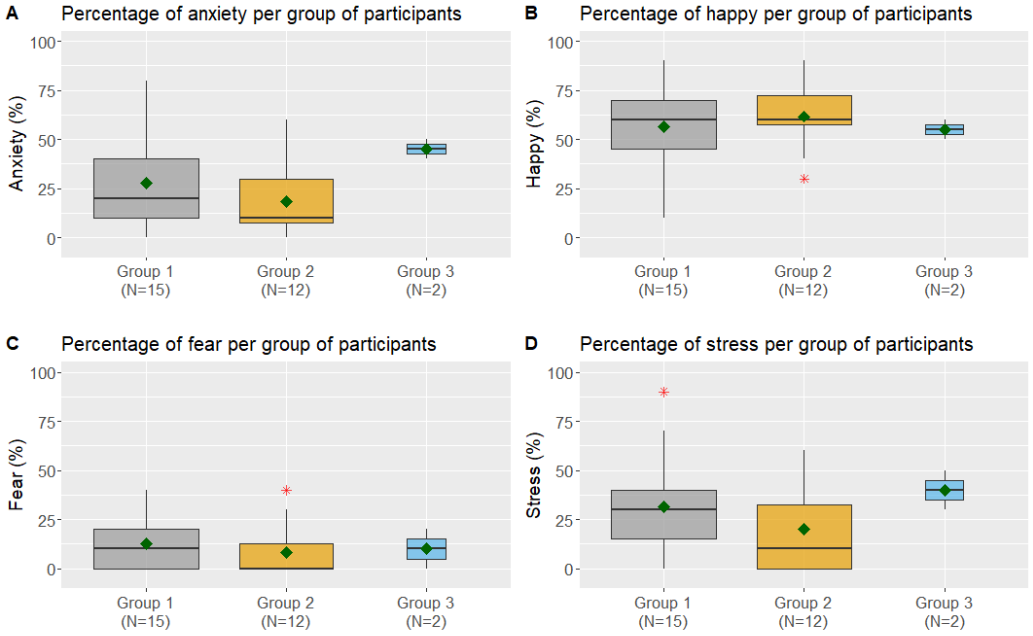

Figure 5.7: Comparison of the percentages obtained in the VAS-Pre questionnaire (Session 1), associated with the variables anxiety (A), happy (B), fear (C) and stress (D), according to the groupings of participants made based on the cluster analysis on the STICSA-Trait questionnaire (Session1).

As we can be seen from the figure 5.7, It is not possible to observe a distinction of the groupings made by means of the trait questionnaires, in terms of the percentages of Anxiety, Happy, Fear, and Stress reported in the VAS-Pre questionnaire. The results of the One Way ANOVA and the Kruskal-Wallis test, shown in table 5.4 reinforce the absence of significant differences between the groupings considered. Thus, there is no clear relationship between the responses to the VAS-Pre questionnaire and the participant's trait.

In order to obtain the conclusions it was necessary to perform normality tests for each case studied, including those not shown here (VAS-Pos), in order to find out if the values of anxiety, happy, fear and stress, followed a normal distribution or not, so as to understand which statistical test should be used in each case to understand if there really were significant differences between the defined groups.

Normality tests were conducted to set of samples for each feature in the four conditions. The results varied depending on the feature and condition, with some features having data that followed a normal distribution under one condition but not under another, as we can check on table 5.3.

Feature	Tests of Normality (Shapiro-Wilk)							
	VAS-Pre (Baseline)		VAS-Pos (Fear)		VAS-Pos (Happy)		Vas-Pos (Neutral)	
	Statistic	P-value	Statistic	P-value	Statistic	P-value	Statistic	P-value
Anxiety	0.912	0.019	0.950	0.180	0.750	0.000	0.691	0.000
Happy	0.948	0.158	0.840	0.000	0.893	0.007	0.946	0.142
Fear	0.782	0.000	0.966	0.456	0.429	0.000	0.508	0.000
Stress	0.917	0.026	0.968	0.514	0.599	0.000	0.668	0.000

Table 5.3: Results of Normality tests.

Based on the results of the normality tests, the most appropriate test was applied in each case, the Kruskal-wallis test, One Way ANOVA or Welch's ANOVA. Table 5.4 shows the results for the case in which the VAS-Pre questionnaire is used, which allow us to assess the existence of significant differences or not between the groups formed from the baseline of each physiological signal collected and the trait and state questionnaires according to the values of anxiety, fear, stress and happy reported in the VAS-Pre questionnaire.

Feature	VAS-Pre (Baseline)									
	ECG		EDA		EMG		State		Trait	
	Chi-square	P-value	Chi-square	P-value	Chi-square	P-value	Chi-square	P-value	Chi-square	P-value
Anxiety*	0.996	0.608	2.397	0.302	2.523	0.283	8.005	0.018	4.451	0.108
Fear*	0.202	0.904	3.812	0.149	1.815	0.404	4.365	0.113	1.183	0.554
Stress*	1.309	0.520	3.190	0.203	0.625	0.732	12.150	0.002	2.767	0.251
Happy**	0.376	0.690	2.163	0.135	0.113	0.894	0.956	0.398	0.236	0.792

* Values obtained using the Kruskal-wallis | ** Values obtained using One Way ANOVA | Considering 2 degrees of freedom

Table 5.4: Results of the statistical tests performed on the grouped values of the features anxiety, fear, stress and happy obtained in the VAS-Pre questionnaire, in order to verify the existence or not of significant differences between the groups formed based on the ECG, EDA, Baseline EMG signals, the state questionnaire, and the trait questionnaire.

The same analysis was carried out using the VAS-Pos fear, VAS-Pos happy and VAS-Pos neutral questionnaires and the physiological signals collected during the fear, happy and neutral stimuli, respectively, as well as the STICSA-State and STICSA-Trait questionnaires. In short, the same conclusions were obtained. The boxplots obtained and the tables with the results of the statistical tests performed in each of these cases are available online[1].

Based on the previous analysis, it was decided to continue this work with the groupings defined from the STICSA-State questionnaire, since it is the one that allows, both visually and methodologically evaluation, the formation of more groups capable of being distinguished in the face of anxiety, happy, fear and stress reported by the participants in the VAS-Pre questionnaire as well as in the VAS-Pos questionnaires.

5.2 Feature Extraction Results

The feature engineering process was based on a validated method [23], that reveals effective on emotional recognition. Since the goal of this work was the evaluation and description of emotional profiles, the feature engineering process was based on previously developed studies. In this regard, the feature extraction process considered allowed the initial extraction of 52 features from EMG, 26 features from EMG MF and 26 features from EMG TR, 33 features from EDA, and 99 features from ECG signal, giving a total of 184 features extracted.

[1] https://github.com/TiagoCruzzz/Repository_TiagoCruz.git

The set of these extracted features is shown in table 5.5. The nomenclature given to each feature corresponds to the junction of each prefix, with each feature time series and with each statistical metric considered. For example, in the case of the EMG the nomenclature EMG_MF_Duration_Var can be formed.

There are some exceptions that are arranged in blue in table 5.5, on which there is no joining of each statistical metric. In addition, in the case of ECG, there is also an exception, in this case associated with Heart rate variability (HRV), which there is no consideration or association of statistical metrics.

Physiological Signal	Prefix	Feature	Statistical Metric
EMG	EMG_MF_ EMG_TR_	Activations_N; Duration; MaxPeakAct; MeanPeaksAct; all_Amplitude; Area	_Mean _Std _Var _Skew _Kurt
EDA	EDA_	Symp; SympN; Tonic; Phasic	
	SCR_	Peaks_N; Height; Amplitude; RiseTime; RecoveryTime	
ECG	ECG_	Rate; Tduration	
	HRV_	MeanNN; SDNN; SDANN1; SDNNI1; SDANN2; SDNNI2; SDANN5; SDNNI5; RMSSD; SDSD; CVNN; CVSD; MedianNN; MadNN; MCVNN; IQRNN; Prc20NN; Prc80NN; pNN50; pNN20; MinNN; MaxNN; HTI; TINN; ULF; VLF; LF; HF; VHF; LFHF; LFn; HFn; LnHF; SD1; SD2; SD1SD2; S; CSI; CVI; CSI_Modified; PIP; IALS; PSS; PAS; GI; SI; AI; PI; C1d; C1a; SD1d; SD1a; C2d; C2a; SD2d; SD2a; Cd; Ca; SDNNd; SDNNa; DFA_alpha1; DFA_alpha2; ApEn; SampEn; ShanEn; FuzzyEn; MSEn; CMSEn; RCMSEn; CD; HFD; KFD; LZC	
	HRV_	MFDFA_alpha1; MFDFA_alpha2	_Width _Peak _Mean _Max _Delta _Asymmetry _Fluctuation _Increment

Table 5.5: Initial extracted features.

It was analyzed the presence of missing values in the initial extracted features, as they can lead to erroneous conclusions by changing important metrics and causing problems in the implementation of some algorithms. It was found that the features HRV_ULF and HRV_VLF, represented in red in table 5.5, contained just Not a Number (NaN) values, so they were eliminated. Data imputation was used to suppress the other NaN entries on data. In case of consideration of the complete signal, the average of the values of the remaining participants associated to the feature in question was used. In the case of consideration of the signal excerpts, the average of other excerpts associated to the participant and feature in question was used.

5.3 Feature Selection Results

After dealing with the NaN values, the correlation matrices between pairs of features according to each emotion were generated separately. Figure 5.8 shows the heatmaps of the correlation matrices according to each emotion. As we can see, in all emotions there are high correlations between pairs of features associated with the same physiological signal, as expected. However, in the case of fear, it is possible to verify that there are also high correlations between pairs of features associated to different physiological signals, allowing us to conclude that fear induces a general

emotional response, through all body. The differences between emotions at this level induces the selection of features to be made separately for each emotion.

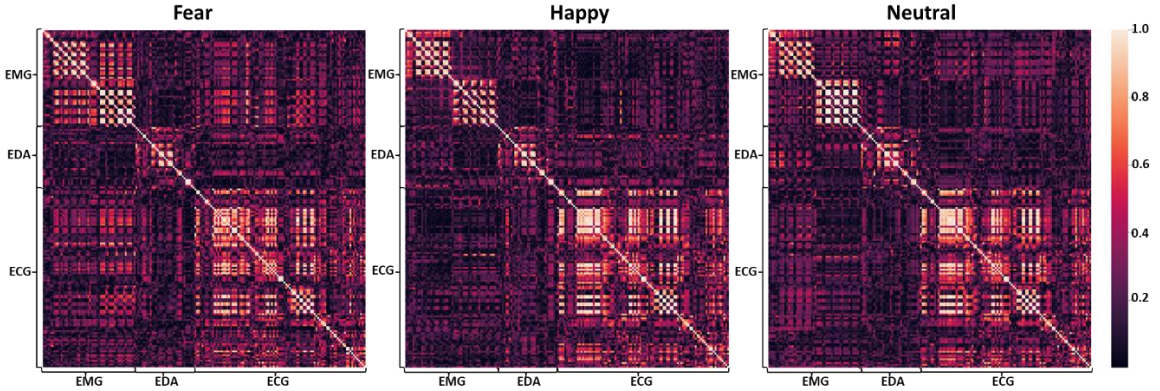

Figure 5.8: Heatmaps of the correlation matrices between pairs of features, for each emotion.

5.3.1 Methodology A and B

Each methodology allowed to obtain a set of features associated with the baseline of EMG, EDA and ECG signals. Three groups of participants were generated, using cluster analysis, for each of the physiological signals considered. Several comparisons were made between the compositions of these clusters formed, for each methodology separately, and each of the physiological signals, and the compositions of the clusters generated from each of the questionnaires: STICSA-State, STICSA-Trait and VAS-Pre.

For example, a comparison was made between the composition of the groups formed from the selected features associated with the EDA signal and the composition of the groups formed from the STICSA-Trait questionnaire. The various possible combinations between the defined groupings were considered, and for each combination the sum of the proportions of common participants was obtained, as shown in table 5.6.

	Combination (Cn)																	
	C1			C2			C3			C4			C5			C6		
EDA signal	G1	G2	G3	G1	G2	G3	G1	G2	G3	G1	G2	G3	G1	G2	G3	G1	G2	G3
Trait Questionnaire	G1	G2	G3	G1	G3	G2	G2	G1	G3	G2	G3	G1	G3	G2	G1	G3	G1	G2
Proportion	0,03	0,34	0,00	0,03	0,07	0,00	0,07	0,41	0,00	0,07	0,07	0,07	0,00	0,34	0,00	0,00	0,41	0,07
Total proportion	0,37			0,10			0,48			0,21			0,34			0,48		

Table 5.6: Combinations between the clusters formed based on the selected features of the EDA signal and the clusters generated through the trait questionnaire, and the respective total proportion of common participants between the compared clusters. The highest value of total proportion generated was selected and highlighted in green.

For each comparison performed, the largest total proportion value obtained was selected. With this, we obtained a set of total proportions associated with methodology A and another set of the same size associated with methodology B.

Based on these total proportion values, a T-test (paired) was performed to verify if the average of the total proportions based on the features selected through methodology A was significantly different from the average of the total proportions considering the features selected from methodology B, with a significance level of 5%.

	Methodology A	*Methodology B*
Mean	0,58	0,55
Variance	0,01	0,01
Observations	21,00	21,00
Pearson Correlation	0,55	
Hypothesized Mean Difference	0,00	
df	20,00	
t Stat	1,43	
P(T<=t) one-tail	0,08	
t Critical one-tail	1,72	
P(T<=t) two-tail	0,17	
t Critical two-tail	2,09	

Table 5.7: T-test: Paired Two Sample for Means

As can be seen in table 5.7, as the p-value is greater than the significance level of 0.05, the Null Hypothesis (H0) is not rejected, which means that there is no evidence to consider a significant difference between the means. Furthermore, since the critical value (t Critical) is greater than the test statistic (t Stat) it means that the test statistic is within the zone of acceptance of H0, which reinforces the conclusion already drawn based on the p-value. Given this conclusion, it was decided to continue the study with the feature selection procedure performed based on methodology A, since it is an original procedure on this database.

Methodology A was applied on the data associated with fear, happy and neutral emotion stimuli's, independently. The number of features that are rejected and not rejected by this method, in each case, are shown in table 5.8.

			EMG MF	EMG TR	EDA	ECG	Total
Nº of features	Rejected	Fear	13	16	16	55	100
	Not Rejected		13	10	17	44	84
	Rejected	Happy	14	14	17	56	101
	Not Rejected		12	12	16	43	83
	Rejected	Neutral	12	14	19	60	105
	Not Rejected		14	12	14	39	79

Table 5.8: Number of features rejected and not rejected after implementation of the method A.

The table with the nomenclature of the selected features is available online[2]. The selected features about each emotional state were later used in the principal component analyses performed.

5.3.2 Methodology C

As mentioned in section 4.4.2, there were two ways of proceeding in which methodology C was used. These were approach C1 and approach C2.

[2] https://github.com/TiagoCruzzz/Repository_TiagoCruz.git

The number of features rejected and not rejected by considering the approach C1, in each case, are shown in table 5.9.

		EMG MF	EMG TR	EDA	ECG	Total
Nº of features	Rejected	18	19	13	67	117
	Not Rejected	8	7	20	32	67
	Total	26	26	33	99	184

Table 5.9: Number of features rejected and not rejected after applying the approach C1.

The number of features selected or not selected by approach C2, in each case, are shown in table 5.10.

		EMG MF	EMG TR	EDA	ECG	Total
Nº of features	Rejected	21	21	17	79	138
	Not Rejected	5	5	16	20	46
	Total	26	26	33	99	184

Table 5.10: Number of features rejected and not rejected after applying the approach C2.

The tables that contain the nomenclature of features selected from each approach are available online[3]. The selected features were later used directly as input in the emotional model constructed.

As seen above, the analysis of the heatmaps of the correlation matrices indicates that selecting features separately may be better, and therefore approach C1 seems to be the most appropriate. However, after selecting features on each emotion separately, this approach requires merging them in order to obtain the final matrix composed of the three emotions, to feed the emotional model. Thus, certain features, which for a certain emotion were redundant and had been removed, inevitably reappear after this union. In this sense, and taking into account that both approaches, C1 and C2, have arguments for and against, the analysis was made in the model for both, in order to understand which one generated a better performance and then decide with which one to follow the analysis.

5.4 Principal Component Analysis Results

To accomplish the final to evaluate the relevance to design emotional profiles, an analysis between the built groups (G1, G2, G3) and the global assessment of participants has to be done. It was found that there are differences in the features that have the highest contribution, that is, a contribution higher than the threshold defined by $\frac{1}{length(f)}$, being f the total number of features, for each principal component. For example, features that contribute most to PC1 of group 1 are different from features that contribute most to PC1 of group 2, group 3, and the overall group. This, in itself, points to the importance of separating individuals into groups. Since, if this separation was not important, it is expected that the features that present a higher contribution for the same principal component would be the same for all groups. The graphs that show features contributions to PC1 are available online[3].

Considering the minimum number of principal components that explain at least 90% of the data variability in each case, when trying to make feature selection based on the contribution of the features, for each group individually, it was verified that all the features were important, some for

[3] https://github.com/TiagoCruzzz/Repository_TiagoCruz.git

a determined principal component and others for another, not having disregard of any feature through this process.

Using the first three principal components over the groups, scatterplot matrices were generated as the one represented in figure 5.9, for each emotion, making use of information of all physiological signals together and separately (figure 5.9 is an example of the results obtained, the rest of the results can be consulted online[4]). These graphs show overlapping groups, which indicates that in the new dimensions there is, for each case analyzed, an overlapping of the distribution of principal component values, not allowing to clearly distinguish the previously defined groupings. However, in general, it is also possible to verify that group 3, which in each case is composed of the majority of participants, assumes values closer to the mean, while groups 1 and 2 encompass more values close to the tails. This in turn induces the importance of separating participants into groups, since the statistical description of the groups indicates different characteristics.

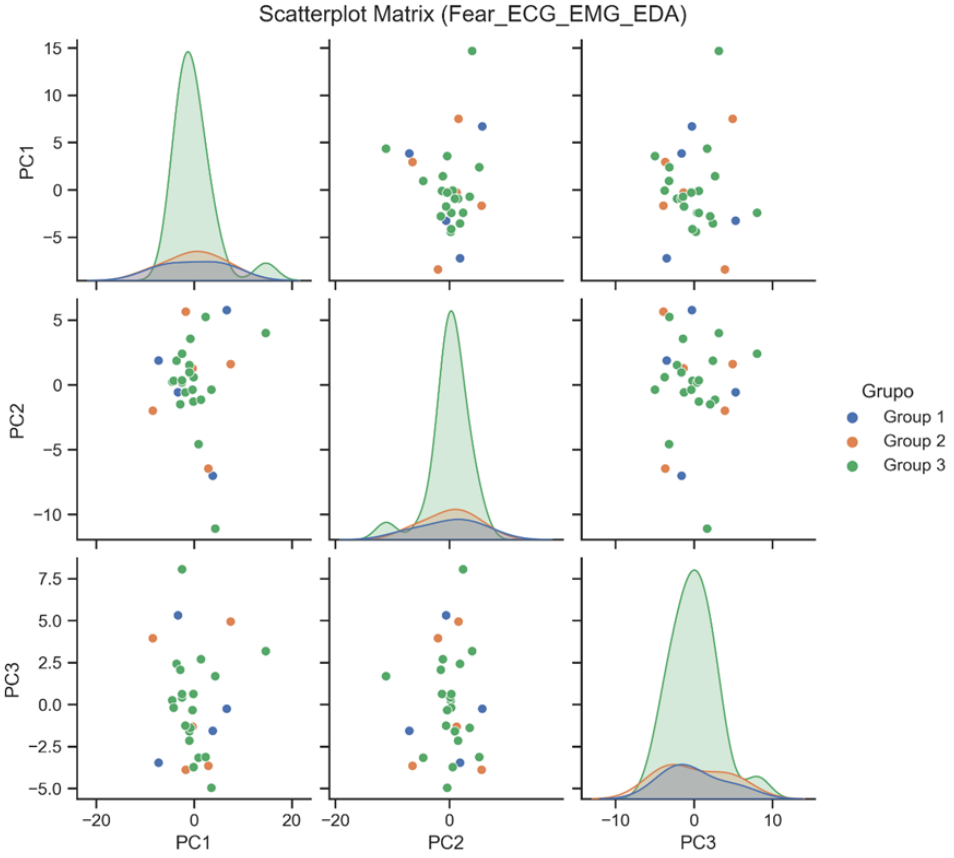

Figure 5.9: Scatterplot matrix considering the first three principal components obtained based on the principal components analysis over the features selected based on methodology A and with the consideration of all physiological signals together on the emotional state of fear.

5.5 Classification

Based on the different datasets and variations of emotional models considered, several emotional classifications were performed.

Table 5.11 displays the results in terms of accuracy and F1-score, associated to each dataset and splitting function used, for the global case and using all physiological signals together.

[4] https://github.com/TiagoCruzzz/Repository_TiagoCruz.git

Dataset	Splitting function	Accuracy ± Standard Deviation		F1-Score					
		Train	Test	Train			Test		
				N	H	F	N	H	F
Dataset 1	Split2	0.343 ± 0.050	0.327 ± 0.035	0.503	0.021	0.020	0.489	0.000	0.008
Dataset 2	Split2	0.940 ± 0.076	0.510 ± 0.073	0.936	0.936	0.949	0.455	0.525	0.519
Dataset 3	Split2	0.916 ± 0.056	0.498 ± 0.103	0.921	0.917	0.910	0.406	0.554	0.493
Dataset 4	Split3	0.999 ± 0.004	0.574 ± 0.050	0.999	0.999	1.000	0.560	0.573	0.584
Dataset 5	Split3	0.495 ± 0.017	0.451 ± 0.034	0.449	0.522	0.502	0.470	0.445	0.423
Dataset 6	Split3	0.710 ± 0.051	0.639 ± 0.036	0.758	0.684	0.689	0.669	0.620	0.628

Table 5.11: Results of training and testing of the classification model on Dataset 1, 2, 3, 4, 5 and 6, for the case where the global group and all physiological signals are considered.

The results were also obtained for the cases in which the global group and each of the physiological signals were considered separately and in which were considered the groups G1, G2 and G3 and the physiological signals separately or together. In general, taking into account the greater accuracy obtained in the test and at the same time the smaller difference between the training and test accuracy, it was verified that the use of the excerpts, and therefore of a larger number of samples, allows better results to be obtained. This indicates that the amount of information provided to the classifier has an impact on its ability to correctly identify emotions. That said, datasets 1, 2 and 3 were discarded as these are the ones where the use of the excerpts is not considered. In addition, a comparative analysis was carried out between the remaining datasets.

Table 5.12 presents the results, in summary form, in terms of accuracy for the datasets. The model was trained considering the signal windowed in 120 seconds.

		Physiological signal							
		ECG+EDA+EMG		EMG		EDA		ECG	
Dataset	Group	Accuracy ± Standard Deviation		Accuracy ± Standard Deviation		Accuracy ± Standard Deviation		Accuracy ± Standard Deviation	
		Train	Test	Train	Test	Train	Test	Train	Test
Dataset 4	Global	0.999 ± 0.004	0.574 ± 0.050	0.965 ± 0.022	0.717 ± 0.035	0.953 ± 0.035	0.768 ± 0.028	0.979 ± 0.021	0.524 ± 0.032
Dataset 5		0.495 ± 0.017	0.451 ± 0.034	0.521 ± 0.034	0.418 ± 0.037	0.449 ± 0.025	0.384 ± 0.026	0.544 ± 0.030	0.508 ± 0.027
Dataset 6		0.710 ± 0.051	0.639 ± 0.036	0.450 ± 0.021	0.394 ± 0.035	0.589 ± 0.065	0.420 ± 0.047	0.651 ± 0.029	0.635 ± 0.027
Dataset 4	G1	0.494 ± 0.195	0.331 ± 0.041	0.503 ± 0.159	0.331 ± 0.070	0.881 ± 0.081	0.453 ± 0.084	0.672 ± 0.103	0.447 ± 0.129
Dataset 5		0.980 ± 0.043	0.661 ± 0.079	0.625 ± 0.027	0.486 ± 0.073	0.840 ± 0.085	0.542 ± 0.095	0.661 ± 0.047	0.531 ± 0.083
Dataset 6		0.963 ± 0.030	0.622 ± 0.102	0.793 ± 0.099	0.489 ± 0.100	0.744 ± 0.060	0.483 ± 0.089	0.516 ± 0.025	0.378 ± 0.052
Dataset 4	G2	0.333 ± 0.000	0.333 ± 0.000	0.854 ± 0.059	0.442 ± 0.095	0.333 ± 0.000	0.333 ± 0.000	0.431 ± 0.223	0.349 ± 0.045
Dataset 5		0.901 ± 0.057	0.604 ± 0.080	0.459 ± 0.016	0.416 ± 0.099	0.627 ± 0.141	0.333 ± 0.097	0.632 ± 0.068	0.309 ± 0.075
Dataset 6		0.961 ± 0.053	0.582 ± 0.087	0.671 ± 0.101	0.440 ± 0.103	0.439 ± 0.035	0.251 ± 0.057	0.850 ± 0.185	0.447 ± 0.086
Dataset 4	G3	0.533 ± 0.311	0.365 ± 0.055	0.957 ± 0.025	0.672 ± 0.049	0.946 ± 0.019	0.696 ± 0.039	0.988 ± 0.019	0.507 ± 0.052
Dataset 5		0.719 ± 0.072	0.491 ± 0.045	0.560 ± 0.084	0.384 ± 0.036	0.431 ± 0.016	0.369 ± 0.024	0.577 ± 0.031	0.483 ± 0.033
Dataset 6		0.757 ± 0.045	0.653 ± 0.034	0.735 ± 0.130	0.397 ± 0.043	0.840 ± 0.120	0.423 ± 0.059	0.717 ± 0.288	0.686 ± 0.399

Table 5.12: Results of the classification model on Dataset 4, 5 and 6, for all groups and sets of physiological signals considered.

Model performance was evaluated in terms of the highest accuracy obtained in the test and at the same time the smallest difference between training and test accuracy (table 5.12). For a simpler interpretation of the results, the better results were marked in green.

Firstly, we can verify that, in general, dataset 6 presents better results than dataset 5, allowing us to conclude here that approach C1 gives better results than approach C2 as expected.

Secondly, it can be seen that the dataset with the best results when considering the use of the ECG signal individually almost always coincides with the dataset with the best results when considering the use of the physiological signals together (ECG+EDA+EMG), which may indicate a dominance of the ECG signal compared to the other signals. The dataset with the best results when considering

the use of the EDA signal individually almost always coincides with the dataset with the best results when considering the use of the EMG signal. Since there are more features associated to ECG than to EDA and EMG, there is an ECG domain, which makes the model selected for the ECG case to be the same as the ECG+EDA+EMG case, and that in the cases of EDA and EMG individually, the model performance is better for a different dataset.

Third, the use of PCA, and therefore the consideration of dataset 4, shows better results in certain cases. However, since PCA adds a layer of complexity that makes it less interpretable compared to those that make use only of features, and that dataset 6 allows for better results in a larger number of cases and shows greater stability of the model in general and non-existence of overfit, it was decided to proceed with the use of dataset 6, for the construction and analysis of the final optimized model.

In fourth place, assuming dataset 6, it is also found that the use of a model with information from the physiological signals together gives better results than models with information from each of the signals separately. This may suggest that the information from the three signals together is an important factor and contains information that is better able to discriminate emotions.

5.5.1 Final optimized model

The optimized model was generated, which consists of the model in which are used the optimal features obtained through Recursive Elimination with Cross-Validation (RFECV) and the optimized hyperparameters, obtained using Grid Searching. In this model, dataset 6 was used, which was the one that provided the best results. By considering this dataset and through Recursive Elimination with Cross-Validation (RFECV), the number of optimal features selected varies according to the model input characteristics (G1, G2, G3, global group and considering features from ECG, EDA or EMG). Through the implementation of Grid Searching, the combination of parameters that provided the highest accuracy was selected. It is important to note that the optimal parameters, as expected, are different depending on the model requirements (groups evaluation and signals).

The optimized model provided the results associated with emotional identification, that is, the confusion matrix, precision, recall, F1-score, and accuracy, for each of the analyzed variations, that is, for each group and set of physiological signals considered.

Figure 5.10 shows the graphs of the results obtained in terms of F1-score. The graphs of the results obtained with the optimized model in terms of Recall and Precision were also built and are available online[5].

[5] https://github.com/TiagoCruzzz/Repository_TiagoCruz.git

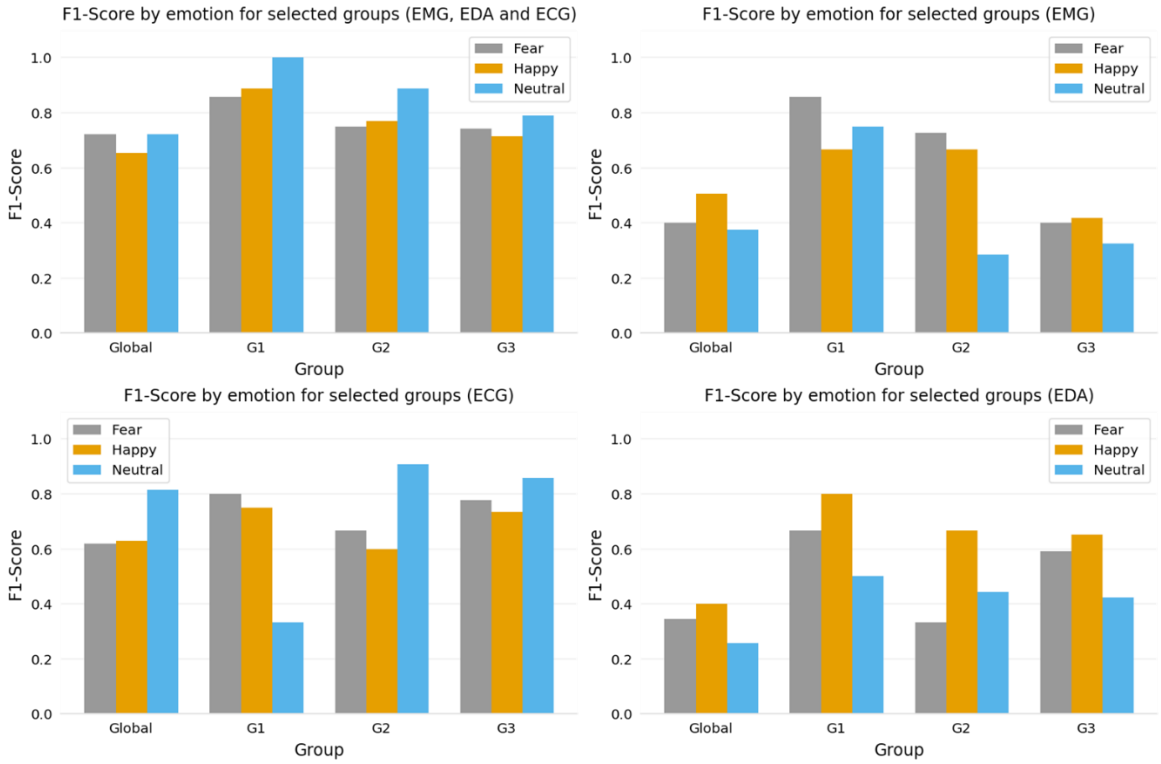

Figure 5.10: Comparative representation of the performance evaluator F1-score, obtained for each group on each emotion considered, according to the consideration of the physiological signals together (EMG, EDA and ECG) and individually (EMG), (ECG), (EDA).

The graphs in figure 5.10 reinforce the conclusion previously drawn that using physiological signals together (EMG, EDA and ECG) brings better results than separately. In addition, they show that, in general, the division by groups of participants allows better model performances to be obtained compared to the global group.

Given the heterogeneity of the population and variability in the interpretation of stimuli, building a global model, i.e., one that incorporates all participants together, will not be feasible. It would work well only near the mean, since it is the region with the largest population. In the tails, it would not work well, so one should find different models for the tails, allowing to generate adequate results to the population in study. In this way, the models are adapted to the participants and not the other way around.

6 Conclusion

With the lack of measures capable of helping in cases of mental and emotional problems and given the high prevalence of these problems, it becomes relevant the implementation and application of emotional recognition and emotion detection systems as efficient and adapted to individuals as possible. While individual emotion recognition has received significant attention, considering groups of participants is an essential aspect that can lead to improved accuracy and a deeper understanding of emotions.

Referring to the first objective of this study, mentioned above, groups of individuals were initially made based on the answers to the STICSA-State, STICSA-Trait and VAS-Pre questionnaires and based on the scores generated by factor analysis, using cluster analysis as a method to validate the groupings formed according to these scores. It was found that, considering only the scores associated with the factor with the greatest variability of the data explained, the factor analysis allowed the formation of groups with apparent meaning, although it had no basis in terms of literature that allowed generalizing this methodology as a way of establishing groupings. Assuming cluster analysis as the method to be followed to define the intended groupings, it was found that the VAS-Pre questionnaire allowed obtaining groups with a more reasonable structure, indicating that the use of a visual scale contributes to more reliable questionnaire answers. Moreover, by comparing the answers to the questionnaires between session 1 and session 2 it was found that the constitution of the groups varied, especially in the case of groups formed by fewer individuals, which is an indication that the state of the person varies over time, inducing different physiological reactions to the same stimulus or different responses between sessions. In an attempt to validate the groups formed from the questionnaires, the data obtained from each physiological signal (ECG, EDA and EMG) were also grouped, and very low proportions of participant coincidence were found between the groups formed by each methodology, which led to the need to assume the VAS-Pre questionnaire as ground truth. In this sense, the groups of participants formed by means of the remaining questionnaires and each physiological signal were statistically analyzed on the values reported for anxiety, happiness, stress and fear, by each participant in the VAS-Pre questionnaire. It was found that the STICSA-State questionnaire was the one that allowed generating groups with the greatest statistically significant difference between them. In this sense, the groups of participants defined on the basis of this questionnaire were assumed.

Several classification models were built, one for each combination between the datasets, the sets of physiological signals and the groups of participants considered.

In this work, two general aspects related to the information to be introduced in the emotional classification models were considered. One in which selected features are introduced directly as model input and the other in which the principal components resulting from principal component analysis are introduced over the selected features. It was possible to observe that for the same principal component, the features that had the largest contribution on it were different between the defined groups, as well as between these groups and the global case, constituting an indication that the separation of individuals into groups made sense.

Moreover, considering the analysis of the three principal components that explained the greatest variability of the data, distributions were obtained which, although indicating overlap between the groups formed, allowed us to perceive that the smaller groups encompassed more information belonging to the tails of the distribution when compared to the larger group, acting as a further indication that it could be important to separate by groups, creating different models for the tails.

Furthermore, and even though this is not one of the focuses of the present work, it was also observed that the consideration of the division of the complete signals into excerpts, and therefore

the inclusion of a larger number of samples allowed for better results of the emotion classification models, pointing to the fact that the amount of information could have an impact on the model's ability to be able to correctly generalize and identify emotions.

It was also found that the process of feature selection on each emotion separately and subsequent union in a final matrix, and therefore the use of dataset 6, allows better results from the classification model, than when joining the data associated with the various emotions and doing feature selection after that, i.e. than dataset 5.

It was also possible to verify that the consideration of principal components allowed better results for some of the combinations. However, most of the best results were obtained when considering the features as model input. Based on this, the construction of optimized discrete models was considered for each combination between the sets of physiological signals and the groups of participants considered, and with the consideration of dataset 6. By considering this dataset, it was found that the use of physiological signals together generally allows better results to be obtained than considering them separately, constituting an indication that physiological signals complement each other in the definition of the biometric signature of emotion.

In the construction of the optimized models, it was obtained a performance, in terms of F1-Score, of 0.721, 0.654, 0.721, for the group composed of all individuals, 0.857, 0.889, 1.000, for group 1, 0.750, 0.769, 0. 889 for group 2 and 0.743, 0.714 and 0.791 for group 3, in the case of considering the emotions Fear, Happy and Neutral, respectively, and the information of the features associated to the ECG, EDA and EMG signals. The results in terms of F1-Score, precision, accuracy, and recall were also obtained for the remaining combinations. In general, it was possible to conclude that the establishment of emotional models for each group of participants formed allows for an improvement in their performance, when compared to the consideration of a global model that incorporates all individuals together. This may be the basis for the construction of systems adjusted and adapted to the individuals' state, capable of providing an increased efficiency in the construction of affective computing systems in real environments.

With this work it was possible to draw conclusions related to emotional recognition through physiological signals and the consideration of groups of participants in this process. The importance of this study lies in its innovative character compared to what is found in the literature, as well as in the clarification of its feasibility as a resource in this area. The literature only resorts to the information collected from the questionnaires to understand why in certain cases, associated with certain individuals, the results are recurrently biased. This gives an extremely innovative and important character to the present work, since an individualization by context is carried out, in which an analysis of the questionnaires is initially performed, discriminating the information a priori and directing the models to the profiles of each subject, trying to avoid obtaining biased results. In the health sector, the high number of people with mental problems, such as depression, schizophrenia, anxiety, among others, is an increasingly evident problem. It is therefore important to help these people to understand and deal with their own emotions, as well as to help the people around them and health professionals to understand and perceive their emotional state. This, not only in the health sector, but also at the economic, educational, and other levels, constitutes an important factor in the emotional management of individuals, maintaining a healthy state, improving their performance, well-being and quality of life. This social goal can be achieved through systems capable of performing emotional recognition, as intended through this work.

By identifying the participants' profiles and defining groups based on their contextual similarities, this work allows the construction of adaptive tools to individual needs.

References

[1] K. Cherry, “The 6 Types of Basic Emotions,” *Verywell Mind*, 2022, https://www.verywellmind.com/an-overview-of-the-types-of-emotions-4163976 (accessed May 12, 2022).

[2] M. M. Bradley and P. J. Lang, “Emotion and Motivation,” in *Handbook of Psychophysiology*, J. T. Cacioppo, L. G. Tassinary, and G. Berntson, Eds., 3rd ed. Cambridge University Press, 2007, pp. 581–607.

[3] P. Tarnowski, M. Kołodziej, A. Majkowski, and R. J. Rak, “Emotion recognition using facial expressions,” *Procedia Comput. Sci.*, vol. 108, pp. 1175–1184, 2017, doi: 10.1016/j.procs.2017.05.025.

[4] C. Min Lee, S. Member, S. S. Narayanan, and S. Member, “Toward Detecting Emotions in Spoken Dialogs,” *IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING*, vol. 13, no. 2, p. 293, 2005, doi: 10.1109/TSA.2004.838534.

[5] C. L. Lisetti and F. Nasoz, “Using noninvasive wearable computers to recognize human emotions from physiological signals”, *EURASIP J. Appl. Signal Processing*, vol. 2004, no. 11, pp. 1672–1687, 2004, doi: 10.1155/S1110865704406192.

[6] L. Shu, Y. Yu, W Chen, H. Hua, Q. Li, J. Jin, X. Xu, “Wearable Emotion Recognition Using Heart Rate Data from a Smart Bracelet”. Sensors (Basel). 2020 Jan 28;20(3):718. doi: 10.3390/s20030718. PMID: 32012920; PMCID: PMC7038485.

[7] P. Nandwani and R. Verma, “A review on sentiment analysis and emotion detection from text,” *Soc. Netw. Anal. Min.*, vol. 11, no. 1, pp. 1–19, 2021, doi: 10.1007/s13278-021-00776-6.

[8] Y. Wang *et al.*, “A systematic review on affective computing: emotion models, databases, and recent advances”, *Information Fusion*, vol. 83–84, pp. 19–52, Jul. 2022, doi: 10.1016/j.inffus.2022.03.009.

[9] D. Das Chakladar and S. Chakraborty, “EEG based emotion classification using “Correlation Based Subset Selection””, *Biologically Inspired Cognitive Architectures*, vol. 24, pp. 98–106, 2018, doi: 10.1016/j.bica.2018.04.012.

[10] E. A. Veltmeijer, C. Gerritsen and K. V. Hindriks, "Automatic Emotion Recognition for Groups: A Review," in *IEEE Transactions on Affective Computing*, vol. 14, no. 1, pp. 89-107, 1 Jan.-March 2023, doi: 10.1109/TAFFC.2021.3065726.

[11] L. B. Krithika and G. G. Lakshmi Priya, “Student Emotion Recognition System (SERS) for e-learning Improvement Based on Learner Concentration Metric”, *Procedia Computer Science*, vol. 85, pp. 767–776, 2016, doi: 10.1016/j.procs.2016.05.264.

[12] A. Dzedzickis, A. Kaklauskas, V. Bucinskas, “Human Emotion Recognition: Review of Sensors and Methods”. Sensors (Basel). 2020 Jan 21;20(3):592. doi: 10.3390/s20030592. PMID: 31973140; PMCID: PMC7037130.

[13] D. Setiono, D. Saputra, K. Putra, J. V. Moniaga, and A. Chowanda, “Enhancing Player Experience in Game with Affective Computing,” *Procedia Computer Science*, vol. 179, pp. 781–788, 2021, doi: 10.1016/j.procs.2021.01.066.

[14] J. Tao, T. Tan, “Affective Computing: A Review”. In: Tao, J., Tan, T., Picard, R.W. (eds) Affective Computing and Intelligent Interaction. ACII 2005. Lecture Notes in Computer Science, vol. 3784. Springer, Berlin, Heidelberg. https://doi.org/10.1007/11573548_125

[15] I. T. Meftah, N. L. Thanh and C. B. Amar, "Multimodal Approach for Emotion Recognition Using an Algebraic Representation of Emotional States," *2012 Eighth International Conference on Signal Image Technology and Internet Based Systems*, Sorrento, Italy, 2012, pp. 541-546, doi: 10.1109/SITIS.2012.84.

[16] UWA, “The Science of Emotion: Exploring the Basics of Emotional Psychology”, 2019, https://online.uwa.edu/news/emotional-psychology/ (accessed May 12, 2022).

[17] C. Dias, J. F. Cruz, and A. M. Fonseca, “Rep-Emoções Passado, presente e futuro,” vol. XXII, no. 2, pp. 11–31, 2008.

[18] WFU, “The Difference Between Feelings and Emotions”, https://counseling.online.wfu.edu/blog/difference-feelings-emotions/ (accessed May 12, 2022).

[19] L. A. González, “Do you know the difference between feelings and emotions?”, 2021, https://ifeelonline.com/en/ifeel-therapies/do-you-know-the-difference-between-feelings-and-emotions/ (accessed May 12, 2022).
[20] L. Shu *et al.*, “A Review of Emotion Recognition Using Physiological Signals,” *Sensors (Basel)*, vol. 18, no. 7, 2018, doi: 10.3390/s18072074. PMID: 29958457; PMCID: PMC6069143.
[21] Smile and Learn, “As emoções básicas para crianças - Alegria, tristeza, medo, raiva, nojo e surpresa”, https://www.youtube.com/watch?v=_6J1fO2wUAw (accessed May 12, 2022).
[22] Song Z., “Facial Expression Emotion Recognition Model Integrating Philosophy and Machine Learning Theory”, *Front Psychol*, vol.12, Sep. 2021, doi: 10.3389/fpsyg.2021.759485. PMID: 34646223; PMCID: PMC8503687.
[23] C. F. Alves, “Discrete to dimensional physiological emotion classification”, Universidade de Aveiro, 2021. Accessed: May 26, 2023. [Online]. Available: http://hdl.handle.net/10773/33692
[24] 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.
[25] M. González-Arias, A. Martínez-Molina, S. Galdames, and A. Urzúa, “Psychometric properties of the 20-Item Toronto Alexithymia Scale in the Chilean Population”, *Front Psychol*, vol. 9, Jun. 2018, doi: 10.3389/fpsyg.2018.00963.
[26] A. S. Mendes, “STICSA: psychometric analysis in a sample of Portuguese university students”, Universidade de Aveiro, 2018. Accessed: May 26, 2023. [Online]. Available: http://hdl.handle.net/10773/25865
[27] I. K. Tindall, G. J. Curtis, and V. Locke, “Dimensionality and Measurement Invariance of the State-Trait Inventory for Cognitive and Somatic Anxiety (STICSA) and Validity Comparison With Measures of Negative Emotionality”, *Front Psychol*, vol. 12, Jun. 2021, doi: 10.3389/fpsyg.2021.644889. PMID: 34163400; PMCID: PMC8215200.
[28] A. L. Vaz, “Normalizar ou padronizar as variáveis?”, Data Hackers, https://medium.com/data-hackers/normalizar-ou-padronizar-as-variáveis-3b619876ccc9 (accessed May 16, 2022).
[29] S. Sharma, “*Applied Multivariate Techniques”.* Wiley, 1996. doi: 10.2307/2291485.
[30] K. R. Gabriel, “The Biplot Graphic Display of Matrices with Application to Principal Component Analysis,” *Biometrika*, vol. 58, no. 3, p. 453, 1971, doi: 10.2307/2334381.
[31] D. W. Wichern, R. A. Johnson, “*Applied Multivariate Statistical Analysis”*, 2007, 6th Edition, Pearson Prentice Hall, Upper Saddle River.
[32] I. D. E. Battisti, and F. M. da Silva Smolski. *Capítulo 2 Análise Fatorial | Software R: curso avançado*, 2019, http://rstub.github.io/bookdown-chapterbib/ (accessed May 12, 2022).
[33] “*KMO test (for measuring sampling adequacy) - All Things Statistics*”, All Things Statistics, 2021, https://allthingsstatistics.com/inferential-statistics/kmo-test/ (accessed May 12, 2022).
[34] B. Everitt, “Cluster analysis,” *Qual Quant*, vol. 14, no. 1, pp. 75–100, 1980, doi: 10.1007/BF00154794.
[35] S. S. Shapiro and M. B. Wilk, “An Analysis of Variance Test for Normality (Complete Samples)”, *Biometrika*, vol.52, no.3/4, 1965, pp. 591-611. https://doi.org/10.2307/2333709.
[36] A. Ghasemi and S. Zahediasl, “Normality tests for statistical analysis: A guide for non-statisticians,” *Int J Endocrinol Metab*, vol. 10, no. 2, pp. 486–489, 2012, doi: 10.5812/ijem.3505.
[37] P. Mishra, C. M. Pandey, U. Singh, A. Gupta, C. Sahu, and A. Keshri, “Descriptive statistics and normality tests for statistical data”, *Ann Card Anaesth*, vol. 22, no. 1, pp. 67–72, Jan. 2019, doi: 10.4103/aca.ACA_157_18.
[38] “R: Shapiro-Wilk normality test.” https://search.r-project.org/CRAN/refmans/psyntur/html/shapiro_test.html (accessed May 27, 2023).
[39] Laerd Statistics, “One-way ANOVA in SPSS Statistics - Step-by-step procedure including testing of assumptions.” https://statistics.laerd.com/spss-tutorials/one-way-anova-using-spss-statistics.php (accessed May 27, 2023).
[40] Laerd Statistics, “One-way ANOVA - An introduction to when you should run this test and the test hypothesis.” https://statistics.laerd.com/statistical-guides/one-way-anova-statistical-guide.php (accessed May 27, 2023).

[41] Laerd Statistics, “Repeated Measures ANOVA - Understanding a Repeated Measures ANOVA.” https://statistics.laerd.com/statistical-guides/repeated-measures-anova-statistical-guide.php (accessed May 27, 2023).
[42] Laerd Statistics, “One-way ANOVA - How to report the significance results, homogeneity of variance and running post-hoc tests.” https://statistics.laerd.com/statistical-guides/one-way-anova-statistical-guide-4.php (accessed May 27, 2023).
[43] STHDA, “One-Way ANOVA Test in R.” http://www.sthda.com/english/wiki/one-way-anova-test-in-r (accessed May 27, 2023).
[44] Laerd Statistics, “Kruskal-Wallis H Test in SPSS Statistics.” https://statistics.laerd.com/spss-tutorials/kruskal-wallis-h-test-using-spss-statistics.php (accessed May 27, 2023).
[45] V. Bewick, L. Cheek, and J. Ball, “Statistics review 10: Further nonparametric methods,” *Critical Care*, vol. 8, no. 3. pp. 196–199, Jun. 2004. doi: 10.1186/cc2857.
[46] “kruskal.test function - RDocumentation.” https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/kruskal.test (accessed May 27, 2023).
[47] “pairwise.wilcox.test function - RDocumentation.” https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/pairwise.wilcox.test (accessed May 28, 2023).
[48] M. Xu, D. Fralick, J. Z. Zheng, B. Wang, X. M. Tu, and C. Feng, “The differences and similarities between two-sample t-test and paired t-test,” *Shanghai Arch Psychiatry*, vol. 29, no. 3, pp. 184–188, Jun. 2017, doi: 10.11919/j.issn.1002-0829.217070.
[49] E. Alpaydin, “*Introduction to Machine Learning,”* third edition, pp. 349-383. in Adaptive Computation and Machine Learning series. MIT Press, 2014.
[50] Y. Cai, X. Li, and J. Li, “Emotion Recognition Using Different Sensors, Emotion Models, Methods and Datasets: A Comprehensive Review,” *Sensors*, vol. 23, no. 5. MDPI, Mar. 01, 2023. doi: 10.3390/s23052455.
[51] F. Pedregosa *et al.*, “Scikit-learn: Machine Learning in {P}ython,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
[52] B. Roy, “All about Feature Scaling,” *Medium*, 2020. https://towardsdatascience.com/all-about-feature-scaling-bcc0ad75cb35 (accessed May 27, 2023).
[53] K. Choudhury, “Feature Engineering — Recursive Feature Elimination With Cross-Validation”, *Medium*, 2020.” https://towardsdatascience.com/feature-engineering-recursive-feature-elimination-with-cross-validation-745d594df730 (accessed May 27, 2023).
[54] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, 2nd ed. Springer New York, NY, 2009. doi: https://doi.org/10.1007/978-0-387-84858-7.
[55] I. Düntsch and G. Gediga, “Confusion Matrices and Rough Set Data Analysis,” in *Journal of Physics: Conference Series*, Institute of Physics Publishing, May 2019. doi: 10.1088/1742-6596/1229/1/012055.
[56] M. Sokolova, N. Japkowicz, and S. Szpakowicz, “Beyond accuracy, F-score and ROC: A family of discriminant measures for performance evaluation,” in *AAAI Workshop - Technical Report*, 2006, pp. 24–29. doi: 10.1007/11941439_114.
[57] D. Choudhary and M. Cescon, “EDA-sense: Dynamic Feedback Control of Sympathetic Arousal,” *IFAC-PapersOnLine*, vol. 53, no. 5, pp. 238–243, Jan. 2020, doi: 10.1016/J.IFACOL.2021.04.103.
[58] F. Agrafioti, D. Hatzinakos, and A. K. Anderson, “ECG Pattern Analysis for Emotion Detection,” *IEEE Trans Affect Comput*, vol. 3, no. 1, pp. 102–115, Jan. 2012, doi: 10.1109/T-AFFC.2011.28.