Valeria Pérez Gil

Comprehensive Analysis of Human Emotions using Artificial Intelligence Techniques

Treball de Fi de Grau

dirigit per Edgar Batista de Frutos
i Antoni Martínez Ballesté

Grau en Enginyeria Biomèdica

UNIVERSITAT ROVIRA i VIRGILI

Tarragona

2024

Resum

Les emocions són fonamentals tant per als processos d'aprenentatge com d'ensenyament, ja que impacten directament en el rendiment acadèmic dels estudiants i la motivació dels docents. Aquest treball explora la viabilitat d'utilitzar senyals fisiològiques per al reconeixement d'emocions mitjançant tècniques d'intel·ligència artificial, especialment en aules intel·ligents. Específicament, es caracteritzen els canvis en l'activitat electrodermal, la freqüència cardíaca i la temperatura de la pell durant experiències emocionals per extreure característiques que descriguin diversos estats emocionals. Els experiments realitzats amb dos conjunts de dades reals indiquen que dependre únicament de senyals fisiològiques és insuficient per distingir amb precisió les emocions, fet que subratlla la necessitat de complementar-les amb informació addicional, com expressions facials, postura i informació contextual. Els nostres descobriments, a més de contribuir a la investigació acadèmica sobre l'ús de senyals fisiològiques per al reconeixement d'emocions, podrien permetre el desenvolupament d'eines complexes per als educadors, permetent-los ajustar les seves metodologies pedagògiques en temps real segons les necessitats emocionals dels estudiants.

Resumen

Las emociones son fundamentales tanto para los procesos de aprendizaje como de enseñanza, impactando directamente en el rendimiento académico de los estudiantes y la motivación de los docentes. Este trabajo explora la viabilidad de utilizar señales fisiológicas para el reconocimiento de emociones mediante técnicas de inteligencia artificial, especialmente en aulas inteligentes. Específicamente, se caracterizan los cambios en la actividad electrodermal, la frecuencia cardíaca y la temperatura de la piel durante experiencias emocionales para extraer características que describan diversos estados emocionales. Los experimentos realizados con dos conjuntos de datos reales indican que depender únicamente de señales fisiológicas es insuficiente para distinguir con precisión las emociones, lo que subraya la necesidad de complementarlas con información adicional, como expresiones faciales, postura e información contextual. Nuestros hallazgos, además de contribuir a la investigación académica sobre el uso de señales fisiológicas para el reconocimiento de emociones, podrían permitir el desarrollo de herramientas complejas para los educadores, permitiéndoles ajustar sus metodologías pedagógicas en tiempo real según las necesidades emocionales de los estudiantes.

Abstract

Emotions are vital to both learning and teaching processes, directly impacting students' academic performance and teachers' motivation. This work explores the feasibility of using physiological signals for emotion recognition through artificial intelligence techniques, particularly in smart classroom environments. Specifically, it characterises changes in electrodermal activity, heart rate, and skin temperature during emotional experiences to extract features that describe various emotional states. Experiments using two real-world datasets indicate that relying solely on physiological signals is insufficient for accurately distinguishing emotions, highlighting the need to complement them with additional information, such as facial expressions, posture, and contextual information. Our findings, in addition to contributing to academic research on the use of physiological signals for emotion recognition, could open the door for the development of complex tools for educators, enabling them to adjust their pedagogical methodologies in real-time according to the emotional needs of the students.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Motivation

Emotions play a fundamental role in human life, influencing survival, decision-making, communication, and relationships. They guide responses to danger, assist in making choices by assigning value to experiences, and facilitate nonverbal communication through expressions and body language. Furthermore, emotions strengthen social bonds by encouraging love, empathy, and trust, while driving motivation by providing purpose and direction. Additionally, emotions contribute to personal identity and are closely linked to both mental and physical well-being, making them essential for a balanced and fulfilling existence.

In the educational context, emotions significantly impact student engagement, motivation, and overall academic performance [1]. Positive emotional states, such as joy and surprise, enhance learning experiences, while negative emotions, such as stress and anxiety, can harm cognitive abilities and lead to poor educational outcomes. Emotions can hence help track students' overall mood during specific activities or analyse which activities are more effective at different times of the day. Therefore, understanding and managing emotions is crucial for both educators and students. To this end, new methods for monitoring and assessing emotional states in real-time are being explored. Most approaches for detecting emotions rely on video analysis, conducted either in real-time or post-lecture [2]. However, video-based approaches can be limited by factors such as lighting conditions, occlusions, or even individual differences in expressing emotions. Significantly enough, these approaches require strong privacy-preserving measures to protect the privacy and confidentiality of underage students in accordance to ethical and legal standards. To relax these constraints, emotions could also be detected using alternative data sources, such as physiological sensing. Physiological signals, such as heart rate, skin conductance, skin temperature, and brain activity, can provide valuable insights into people's emotional states [3]. These signals can be seamlessly captured through wearable devices in an unobtrusive and user-friendly manner, such as smartwatches, fitness trackers, and bracelets, among others [4]. These devices offer the possibility of real-time emotion monitoring in dynamic contexts, such as educational environments.

Particularly, the educational field is experiencing a groundbreaking revolution driven by the integration of digital, contextualised, and adaptive technologies in order to foster more efficient and effective learning environments [5]. Smart classrooms, which integrate technology with pedagogy, use advanced tools and systems to offer more adaptable and personalised learning experiences for students, while also enhancing the work environment for educators [6]. Incorporating emotion recognition technology into these smart classrooms has the potential to further improve this adaptability by providing real-time feedback on students' emotional states. Such information enables educators to adjust their teaching strategies and provide more targeted support to students who may be experiencing difficulties. Additionally, addressing students' emotional needs can contribute to better educational outcomes by creating a supportive learning environment and enhancing the overall learning experience.

The successful implementation of emotion recognition in smart classrooms lies in the use of artificial intelligence (AI) techniques [7]. These techniques may range from heuristics and rule-based AI to more sophisticated models based on machine learning and deep learning. Regardless of the specific technique employed, AI systems must be capable of discerning patterns within physiological data that correspond to specific emotional states. Consequently, when real-time physiological data is received, the AI model can identify these patterns and accurately classify them into the appropriate emotions. This process requires collecting large amounts of data, training the models, and validating their accuracy in recognising different emotional states.

Despite the promising potential of emotion recognition technology in smart classrooms, several challenges remain to be addressed. First, there are limited studies focusing on the use of physiological data for emotion recognition, particularly within the educational domain. Second, most existing studies involve adults rather than children, whose physiological responses to emotions may differ significantly. Third, publicly available datasets that include physiological signals for emotion recognition are scarce and often limited by the aforementioned factors. Fourth, emotions are often context-dependent, so physiological data alone may be insufficient to interpret emotions accurately, thus requiring complementary information, such as video footage or contextual data. Fifth, wearable devices used for physiological sensing must ensure accurate data collection and real-time transmission for effective analysis. Finally, ethical concerns regarding surveillance, intrusion, and potential misuse must be thoroughly considered.

1.2 Objectives

The main goal of this work is to validate the use of physiological signals data for emotion recognition through AI techniques and to evaluate the feasibility of implementing these methods in smart classroom settings. Specifically, the research objectives are as follows:

1. Develop a **comprehensive theoretical understanding of emotions** from a psychological perspective by examining prominent theories of emotion. The aim is to understand their role in education, specifically how they influence learning outcomes, student engagement, and classroom dynamics.

2. Identify the most effective **physiological signals for emotion detection** by reviewing existing studies on emotion recognition. This objective includes evaluating the strengths and weaknesses of these signals and assessing their feasibility for real-work application in educational contexts using wearable devices. Key considerations will include usability, comfort, and non-invasiveness.

3. Examine **existing datasets in the literature** that include physiological signals for emotion recognition. This involves identifying datasets with diverse, accurate, and scientifically validated data. The study will assess factors such as sample size, population characteristics, signal quality, and data diversity to determine their suitability for further testing.

4. Develop a **methodology to discern unique patterns** associated with different emotional states. To this end, signal pre-processing, characterisation, and

feature extraction techniques will be applied to the physiological data. Adequate strategies for allowing comparisons, both numerically and graphically, will be developed.

5. Provide a **quantitative and graphical comparison** of the different emotions to evaluate the effectiveness of the patterns uncovered from the physiological data. Results are compared using different comparison strategies, either with randomised groups of individuals and with groups of similar individuals based on demographic variables, such as age group and gender.

6. Set the ground for **further research developments** in the field of emotion recognition in smart classrooms.

2 Background

2.1 Theory of Emotion

Emotions are an integral part of the human experience, playing a crucial role in both the learning process and social interactions. Positive emotions contribute to improved health and work efficiency, while negative emotions can lead to health problems, with their prolonged accumulation being a predisposing factor for depression. An emotion is a mental state that arises spontaneously, rather than through conscious effort, and is often accompanied by physical and physiological changes. These changes impact various parts of the body, including the brain, heart, skin, blood flow, muscles, facial expressions, voice, and other tissues [8].

Among the numerous psychological models designed to define and categorise emotions theoretically, two of the most well-known are the Six Basic Emotions Model proposed by Paul Ekman [9], which identifies universal emotions such as joy, sadness, fear, surprise, disgust, and anger, and the Circumplex Model of Affect proposed by James Russell [10], which organises emotions within a two-dimensional space based on valence (pleasant-unpleasant) and arousal (high-low).

2.1.1 Ekman's Six Basic Emotions Model

Paul Ekman's Six Basic Emotions Model is a theory that identifies six universal emotions that are recognised and expressed similarly across different cultures [9]. These six emotions are as follows:

- **Joy**: A feeling of happiness and pleasure, often associated with smiling or laughing.
- **Sadness**: A feeling of sorrow or unhappiness, typically reflected in a down turned mouth, tears, and a sombre expression.
- **Fear**: An emotional response to a perceived threat, often characterised by widened eyes and a tense body.
- **Surprise**: A reaction to unexpected events, usually shown by raised eyebrows, wide eyes, and an open mouth.
- **Disgust**: A feeling of aversion or revulsion, often expressed by a wrinkled nose, a curled lip, or a grimace.
- **Anger**: A strong feeling of displeasure or hostility, commonly associated with furrowed brows, clenched teeth, and a tense jaw.

The popularity of this model is worldwide, as it is the one employed in the cinematic industry to represent the emotions in the Pixar's film "Inside Out" (Figure 1), where such emotions are represented with their characteristic colour and personified as characters, each governing the thoughts and behaviours of a little girl.

The groundbreaking aspect of Ekman's research was its cross-cultural validity, demonstrating that these emotions are universally recognised, regardless of cultural

Figure 1: Ekman's Basic Emotions in the "Inside Out" film.

background. He discovered that facial expressions associated with these emotions are not culturally learned, but are biologically innate, supporting the idea that these basic emotions evolved as part of human survival mechanisms [11]. Ekman's model is foundational in the field of emotion research, influencing numerous theories and practices across disciplines, including psychology, behavioural sciences, media, and entertainment, where understanding and portraying emotions is crucial. Furthermore, Ekman's studies also included the development of the Facial Action Coding System (FACS), a comprehensive framework for categorising physical expressions of emotions by analysing specific facial muscle movements [12]. This system has been widely adopted in fields such as psychology, anthropology, and even AI research to understand and interpret human emotions.

2.1.2 Russell's Circumplex Model of Emotion

The Circumplex Model of Emotion, originally known as the Circumplex Model of Affect, is a theory proposed by James A. Russell that organises emotions in a two-dimensional space [10]. The model represents emotions along two main dimensions, valence and arousal, described as follows:

- **Valence**: This dimension measures how pleasant or unpleasant an emotion is. Emotions with positive valence are considered pleasant, while those with negative valence are seen as unpleasant. For example, joy has a high positive valence, while sadness has a low positive valence.
- **Arousal**: This dimension indicates the level of activation or intensity of the emotion. Emotions with high arousal are intense and energetic, such as excitement, while those with low arousal are calm or subdued, such as relaxation.

In this model, emotions are plotted within a circular space where the horizontal axis represents valence (pleasant-unpleasant), and the vertical axis represents arousal (high-low). Emotions that fall close to each other in this space share similar characteristics, thus helping in understanding the relationships across different emotions. For example, emotions with both high valence and arousal include excitement and joy, while emotions with both low valence and arousal include sadness and boredom. Emotions with high valence and low arousal include contentment and relaxation while, conversely, emotions with low valence and high arousal include anger and anxiety. A comprehensive summary of such emotions in such circular space is depicted in Figure 2.

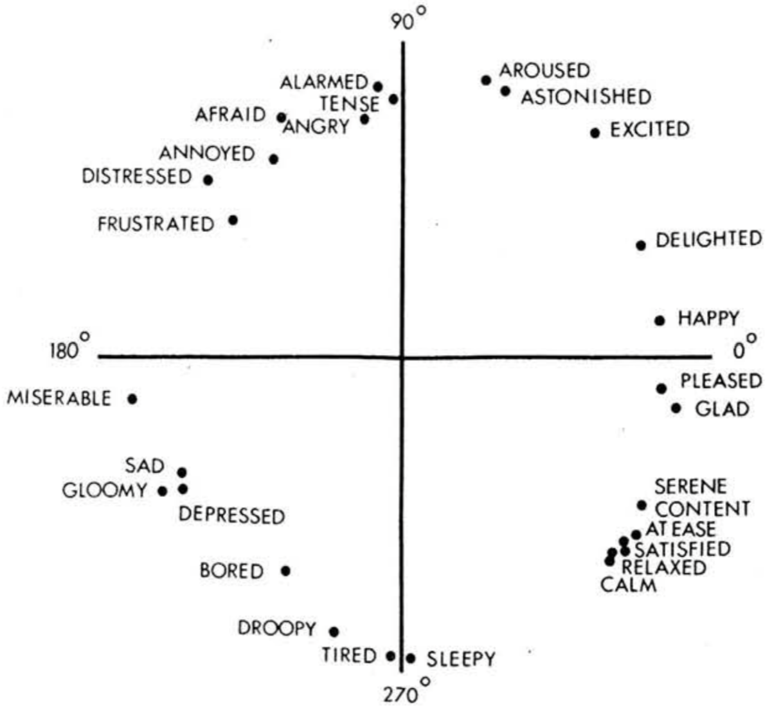


Figure 2: Russell's Circumplex Model of Affect.

Like the Six Basic Emotions model, Russell also conducted a cross-cultural study of his model [13]. Moreover, this model has been influential in emotion research because it provides a structured framework for categorising and understanding the complexity of emotional experiences. It emphasises that emotions are not discrete entities but rather exist on a continuum, allowing for a more detailed analysis of emotional states. This model is widely used in psychological research, emotion theory, and practical applications such as emotional regulation and interpersonal communication [14].

2.2 Physiological Signals

Various physiological signals have been considered in the study of emotion detection, providing valuable insights into the autonomic nervous system's responses to emotional stimuli. These signals, serving as essential tools for understanding emotional processes, are crucial in affective computing and psychophysiology research. Several studies in the literature have explored different physiological signals to detect and classify emotions [15–17].

The importance of physiological signals lies in their ability to provide an objective and continuous measure of emotional state, which can play a crucial role in educational settings where emotions fluctuate rapidly and significantly impact the learning process and outcomes. By monitoring these signals, it is possible to identify mental states and adapt educational strategies to optimise the learning experience. Moreover, it is noteworthy that these signals can be easily captured using wearables, such as bracelets, that are generally comfortable for children.

Three of the most popular physiological signals employed for emotion detection are considered in this work: heart rate, skin temperature, and electrodermal activity.

2.2.1 Heart Rate

A person's heart rate (HR), also known as their pulse, is the number of heartbeats per minute. It can fluctuate based on various factors, such as temperature, body position, exercise, emotions, body size, and medication use. For instance, HR typically increases during exercise or stress and decreases during resting periods [18, 19]. A normal resting HR for adults generally ranges from 60 to 100 beats per minute (bpm), but individual variations can occur due to factors like age, fitness level, and overall health. Younger individuals tend to have a higher HR, whose resting HR values typically range as follows [20]:

- **Newborns (birth to 4 weeks)**: 100 to 205 bpm.
- **Infant (4 weeks to 1 year)**: 100 to 180 bpm.
- **Toddler (1 to 3 years)**: 98 to 140 bpm.
- **Preschool (3 to 5 years)**: 80 to 120 bpm.
- **School-age (5 to 12 years)**: 75 to 118 bpm.
- **Adolescents (13 to 18 years)**: 60 to 100 bpm.

HR can be measured using various methods, ranging from basic to advanced techniques [4]. A simple yet reliable method involves checking the pulse at the wrist or neck, counting the beats for 15 seconds, and then multiplying by four to determine the beats per minute. However, this manual approach is not suitable for continuous monitoring. One of the most widely accepted methods for HR measurement is electrocardiography (ECG), which records the electrical activity and rhythm of the heart through electrocardiograms. While traditional ECG monitoring often relies on Holter monitors, these devices can be obstructive and invasive. In contrast, wearable technologies, such as fitness trackers, chest straps, and finger sensors, offer less invasive alternatives [21]. Another popular method for measuring HR is photoplethysmography (PPG), which uses pulse oximeter sensors to measure optical signals. Although PPG is significantly cheaper than ECG, its accuracy may be lower [22].

2.2.2 Skin Temperature

Skin temperature (TEMP) is the temperature of the outermost layer of the human body. It is influenced by factors such as the internal body temperature, external environment, blood flow, and metabolic activity. Normal TEMP for healthy adults ranges between 33-37 $^{\text{o}}$C. However, these values are subject to small fluctuations throughout the day and across different body regions [23]. Research studies have revealed that TEMP decrease involves higher arousal, while warmer TEMP values are generally associated to calm states [24, 25].

There are various methods for measuring TEMP [4]. Non-contact infrared thermometers are popular for quickly checking temperature on the forehead, while ear thermometers can do the same for the ear. Thermal imaging cameras provide detailed

TEMP maps in specialised settings. Contact methods attach skin sensors and digital thermometers directly to the skin for more precise and continuous readings. Wearable devices, such as smartwatches, are already used to track TEMP over time. Moreover, liquid crystal strips, often used on the forehead, offer a quick and simple way to estimate TEMP, especially in children [26].

2.2.3 Electrodermal Activity

Electrodermal activity (EDA), also known as skin conductance or galvanic skin response, aims to detect electrical changes on the skin's surface, triggered by signals from the brain and influenced by the sympathetic nervous system. These changes occur during emotional arousal, increased cognitive workload, or physical exertion, which lead to increased sweating that enhances electrical conductance. This property allows detecting emotional responses, such as stress, anxiety, depression, and fatigue [27, 28].

Simple instrumentation is needed to measure EDA, requiring only a couple of electrodes placed next to each other on the skin surface, generally at the wrist or fingertips [4]. These electrodes detect fluctuations in electrical resistance, which are used to assess changes in skin conductance. Normal human EDA values typically range from 1 to 20 μS. Although common devices for measuring EDA are wired, many wearable solutions already implement wireless solutions in the form of wristbands and finger straps [29].

2.3 Datasets for Emotion Recognition

AI-based emotion recognition methods require high-quality, relevant, and accurate data, as the effectiveness of the results largely depends on the quality of the data used. In this work, which focuses on students and teachers within educational environments, it is essential that the data accurately reflect this specific setting as far as possible.

Ideally, our system should be trained using data collected in a controlled classroom environment, where both students and educators wear wearable devices that gather physiological signals in real-time participating in various group and individual activities throughout the day. This approach would allow for the monitoring of multiple factors, including physiological signals (*e.g.,* HR, TEMP, and EDA), environmental conditions (*e.g.,* air temperature and humidity, CO_2 concentration, particulate matter, lighting, and noise level), and contextual activities (*e.g.,* learning activities conducted at different times, variations in academic load, and subjects taught). This approach would comprehensively capture the complexity and diversity of emotional and stress states in educational environments. However, in practice, finding datasets that meet all these criteria is often idealistic. Most of the datasets currently available do not perfectly align with this optimal scenario, and experiments must be conducted with the available data that best reflects the conditions of interest.

A comparison of the available datasets, either open-source and request-for-use datasets, is provided in Table 1. Both datasets WESAD [30] and SWELL [31] were sourced from Kaggle, an online community for data scientists and machine learning professionals. Dataset CLAS [32] was found in a Mendeley database, while datasets

Table 1: Comparison of datasets containing physiological signals for emotion recognition.

Dataset	Participants	Modalities	Acquisition context	Activities	Description
WESAD [30]	15 adults	EDA, ECG, BVP, EMG, respiration, body temperature, and acceleration	Daily life with scheduled activities	Neutral reading while sitting or standing, watching funny videos, and public speaking or doing arithmetic tasks to induce three states: stress, fun, and neutrality	Dataset associated with different affective states, collected using wearables
SWELL [31]	25 adults	EDA, HRV, body postures, facial expression, and computer logging	Typically knowledge work in a simulated office	Writing reports making presentations, reading e-mails and searching for information. Manipulating working conditions with stressors	Dataset for research on user stress and modelling
CLAS [32]	62 adults	ECG, PPG, EDA, and acceleration	Daily life with scheduled activities	Engaging in interactive activities (*e.g.,* math problems, logic puzzles,. . .) and perceptual activities (*e.g.,* viewing images and vídeos)	Dataset for automatic recognition of some specific mental states
CASE [33]	30 adults	BVP, ECG, EDA, EMG, respiration, and body temperature	Evaluation of audiovisual stimuli	Watching various videos randomly among participants	Dataset reflecting the continuous real-time annotation of participants' emotions
EmoWear [34]	49 adults	EDA, ECG, BVP, respiration, TEMP, acceleration, and gyroscope	Evaluation of audiovisual stimuli	Watching various videos randomly among participants	Dataset recorded for emotion recognition and context awareness
DEAP [35]	32 adults	EEG, EDA, BVP, EMG, EOG, respiration, and body temperature	Evaluation of audiovisual stimuli	Watching videos	Dataset for analysing human affective states
AMIGOS [36]	40 adults	EDA, ECG, EEG, and body videos	Evaluation of audiovisual stimuli	Watching videos individually and in groups	Dataset for research on affect, personal traits and mood
MAHNOB-HCI [37]	27 adults	EDA, ECG, EEG, respiration, body temperature facial videos, audio cues, and gaze data	Evaluation of audiovisual stimuli	Watching videos	Dataset recorded in response to affective stimuli for emotion recognition and implicit labelling research

CASE [33], EmoWear [34], and DEAP [35] were discovered from related literature on emotion recognition. The authors of the latter also provided another dataset, AMIGOS [36], which includes excerpts of multiple datasets, including MAHNOB [37].

These datasets contain physiological measurements that are suitable for detecting emotions, with particular emphasis on EDA data. All the data provided is exclusively objective, this is, captured by sensors, and information about the participants' feelings either during or after the experiment is not included.

Moreover, it is noteworthy that all datasets involve only adults and not children, showing a lack of research in the field of emotion recognition through physiological sensing in children. Research by [38] demonstrated age-related differences between 7 to 9 year old children and young adults in ratings of arousal and valence. Children tend to rate positive and neutral images as more positive and stimulating compared to adults, while no significant age differences were found for negative or aversive images. This tendency suggests that children perceive certain situations more positively or optimistically, opening the door for a specific emotion recognition systems tailored for them. Such a system, however, would require training data that accurately reflects children's emotional characteristics and responses. Unfortunately, this seems unrealistic with the available datasets in the state of the art.

With regards to teachers, stress levels might be presented differently between men and women. The study in [39] analysed gender differences in stress-related factors among university professors in Spain, using questionnaires to assess psychological, nutritional, physical activity, and oral health aspects. The results showed that female university professors exhibited higher levels of perceived stress, emotional exhaustion, and neuroticism compared to their male colleagues, as well as more physical symptoms associated with stress, such as dry mouth and digestive problems. Additionally, the work in [40] investigated the feasibility of using EDA signals to predict stress and determine factors affecting the accuracy of its classification. In their experiment, when gender information was included, the performance analysis showed significant differences between men and women. With these results, both objective and subjective, it can be concluded that the difference in stress between men and women is an important aspect that should be considered when developing the predictive model.

3 Methodology

In this section, the methodological framework of our work is described, detailing the datasets used and the rationale for their selection, the signal characterisation process conducted before analysis, and the strategies employed to examine each dataset.

3.1 Datasets Selection

Two of the eight datasets outlined in Section 2.3 have been used in this work: WESAD and EmoWear. These datasets were selected for two main reasons. First, the devices employed for data collection in both datasets were wearable devices, specifically the Empatica E4 [41], which closely simulates how data might be collected in smart classroom environments. Second, unlike some other datasets, WESAD and EmoWear include a range of elicited emotions beyond just stress and non-stress. Additionally, these datasets were created at different time periods: WESAD, which is about six years old and widely used in previous research, and EmoWear, which has been published recently. This temporal difference is noteworthy as it could reflect changes over time in physiological signals and, consequently, in the emotions captured.

WESAD (Wearable Stress and Affect Detection) is a high-quality, multi-modal dataset that includes both physiological and motion data for detecting stress and affective states, specifically neutral, stress, and amusement. The data is recorded using two devices: the Empatica E4, worn on the wrist of the non-dominant hand, and the RespiBAN Professional, worn on the chest. The sensor modalities include ECG, EDA, TEMP, blood volume pulse (BVP), electromyogram (EMG), respiration rate, and three-axis acceleration. All signals from the RespiBAN Professional are sampled at 700 Hz, while signals from the Empatica E4 are sampled at different rates depending on the type of signal and do not include every signal recorded by the RespiBAN Professional. The sampling rates for the signals are 64 Hz for BVP, 4 Hz for both EDA and TEMP, and 32 Hz for acceleration. In this dataset, participants are mainly graduate students from the authors' research facility. Exclusion criteria for participation were applied, including pregnancy, heavy smoking, mental disorders, and chronic or cardiovascular diseases. As a result, a total of 17 subjects were initially included in the study, but data from two participants had to be discarded due to sensor malfunctions, thus resulting in 15 participants (3 females, 12 males) with a mean age of 27.5 ± 2.4 years old. In order to elicit the three affective states on the participants, the research protocol was conducted as follows:

- **Preparation**: Participants were instructed to avoid caffeine and tobacco for at least an hour before the experiment and to avoid excessive exercise on the day of the study. Prior to the experiment, they read and signed a consent form. Upon arrival at the study location, participants were fitted with the necessary sensors, and a brief sensor test was conducted. The RespiBAN and E4 devices were manually synchronised using a double tap gesture.

- **Baseline condition**: Once the participants were equipped with the sensors, a 20-minute baseline recording was conducted. During this period, participants were seated or standing at a table and were provided with neutral reading material, specifically magazines. The purpose of the baseline condition was to induce a neutral affective state in the subjects.

- **Amusement condition**: During the amusement condition, participants watched eleven funny videos. Between each clip, there was a brief neutral sequence of five seconds. The entire amusement condition lasted 392 seconds.

- **Stress condition**: In the Trier Social Stress Test (TSST), participants first gave a five minute speech about their personal traits to a panel of three people who were introduced as human resources specialists. They had three minutes to prepare their speech but were not allowed to use notes. After the speech, participants were asked to count backward from 2023 in steps of 17, restarting each time a mistake was made. The entire TSST lasted approximately ten minutes, followed by a ten minute break.

- **Meditation**: After the amusement and stress conditions, which were designed to stimulate the participants, participants took part in a guided meditation to help return to a neutral emotional state. This seven minute meditation included a controlled breathing exercise guided by an audio track. Participants followed the instructions with their eyes closed while sitting comfortably.

- **Recovery**: At the end of the protocol, the sensors were again synchronised with a double tap gesture. Then, the sensors were removed and the subjects were informed that the panel members were actually just regular researchers.

The study lasted about two hours in total. The protocol included two main conditions: an amusement condition and a stress condition, which were alternated between subjects in order to avoid effects of order. In addition to these conditions, baseline measurements and two meditation periods were recorded. To introduce variability in posture, the baseline, amusement, and stress conditions were conducted with participants either standing or sitting, with roughly half of the subjects in each posture. However, all subjects were seated during the meditation. Between the recordings of each condition, participants completed self-reports. The study protocol included two different versions:

- **Version A:** Baseline, Amusement, Meditation I, Stress, Rest and Meditation II.
- **Version B:** Baseline, Stress, Rest, Meditation I, Amusement, and Meditation II.

EmoWear (Wearable Physiological and Motion Dataset for Emotion Recognition and Context Awareness) is a multi-modal dataset designed for emotion recognition using seismocardiography, which measures small cardio-respiratory induced vibrations on the chest wall through Inertial Measurement Units (IMUs). The dataset includes signals, such as ECG, EDA, TEMP, BVP, respiration rate, three-axis acceleration, and gyroscope. The data is recorded using three devices: the Empatica E4, Zephyr BioHarness 3 (BH3), and ST SensorTile.box (STb), each worn at different locations on the body and recording various physiological and motion signals at different sampling rates, as follows:

- **ST SensorTile.box (STb)**, worn on the sternum, back, and cup-side, records:
  - Accelerometer 1 LIS2DW12 at 1600 Hz, ±2 g
  - Accelerometer 2 LIS3DHH at 1100 Hz, ±2.5 g
  - Accelerometer 3 LSM6DSOX at 208 Hz, ±2 g

  - Gyroscope LSM6DSOX at 208 Hz, ±250 dps

- **Zephyr BioHarness 3 (BH3)**, worn on the chest, records:
  - ECG at 250 Hz
  - R wave to R wave at a variable rate
  - HR at 1 Hz
  - Respiration rate at 25 Hz
  - Breath to Breath at a variable rate
  - Breath rate at 1 Hz
  - Accelerometer at 100 Hz

- **Empatica E4**, worn on the wrist of the non-dominant hand, records:
  - EDA at 4 Hz
  - TEMP at 4 Hz
  - BVP at 64 Hz
  - HR at 1 Hz
  - Inter-beat Interval at a variable rate
  - Accelerometer at 32 Hz

Data were recorded from 49 participants who watched emotionally stimulating videos that had been previously validated for their emotional content. Participants were recruited through social network advertisements (*e.g.,* Facebook, Instagram, X –formerly Twitter–, and LinkedIn) targeting healthy adults aged 18 to 65 with no history of cardiovascular diseases, movement disorders, or psychological impairments. For the sake of completeness, additional information about participants' long and short term conditions before the experiments was gathered, including demographic details (*e.g.,* gender, birth year, education level, dominant hand, vision status, and use of vision aids), alertness level (*e.g.,* typical nightly sleep hours, last night's sleep duration, any physical or psychological disorders), and substance consumption the day before the experiments (*e.g.,* coffee, tea, alcohol, tobacco, drugs). Each participant was then assigned a unique 4-character ID along with a sequential code from 1 to 49. However, due to a software error, data from one participant (code: 35, ID: 9W29) were not recorded, resulting in a final sample size of 48 participants (21 females, 27 males), aged between 21 and 45 with a mean of 29.3 ± 4.5 years old.

The elicitation of the emotional stimuli was based on the Russell's Circumplex Model, by dividing the valence-arousal space into four quadrants: high arousal-high valence (HAHV), low arousal-high valence (LAHV), low arousal-low valence (LALV), and high arousal-low valence (HALV). Each quadrant was used as a reference label for the video clips. A total of 38 stimuli were presented (10 HAHV, 9 LAHV, 10 LALV, and 9 HALV) across 38 experiment cycles, with the content displayed on a 24-inch monitor at a resolution of 1920x1080 pixels. To minimise biases, the order of the stimuli was at random. However, to ensure that participants concluded the data collection session on a positive mood, the last stimuli was always a HAHV one. More specifically, the research protocol was conducted as follows:

- **Preparation**: Upon arrival, participants were given a detailed explanation on the purpose, scope, and details of the experiments. They were informed explicitly of their right to withdraw from the study at any point during the session. Any questions or uncertainties were addressed, and all necessary information was provided in written form. After explicit consent to participate was obtained, participants were given the opportunity to use the restroom before starting the experiments. Additionally, they were asked to mute any electronic devices, such as smartphones and smartwatches.

- **Pre-experiment questionnaire**: Participants completed an electronic pre-experiment questionnaire using the ColEmo software, an open-source software designed specifically for this study, while the experimenter initiated the wearable devices. The participants were assisted in wearing the sensors. To synchronise the sensors, participants were asked to perform three consecutive jumps with about a one-second pause between each jump, holding the third STb sensor in their hand. The experimenter pressed a button in the ColEmo software immediately after the participant's third jump to indicate the completion of the synchronisation jumps, with the moment of this button press recorded in the GUI logs.

- **Phase 1**: Participants recorded vocal vibrations for potential future use in voice activity detection. They read 10 sentences while all sensors were active, with each sentence shown for 6 seconds and a progress bar displayed. Participants were instructed to maintain a steady position and were recorded using a built-in microphone, with the audio segmented and reviewed for accuracy before deletion. The sentences were selected from the Common Voice dataset, featuring various types and lengths.

- **Phase 2**: Participants underwent a series of 38 elicit-assess-walk cycles. Each cycle included a baseline recording, a fixation cross, a 1-minute video stimulus, a self-assessment survey, a walking task, and an optional water break. The self-assessment surveys asked for the participants' emotional valence, arousal, and dominance, along with responses to questions about the videos. This cyclic procedure was designed to assess the impact of emotional states on gait and other physiological parameters. Participants were familiarised with the procedure through a trial sequence and walked a set route during each cycle. The total session lasted about 1 hour and 30 minutes.

- **Conclusion**: At the end of the experiment, the sensors were removed, and the researchers offered a drink to the participants.

3.2 Signal Characterisation

The data from both datasets are organised similarly. For the WESAD dataset, after downloading the ZIP file, there are 15 folders, each corresponding to a study participant. Within each participant's folder, the data recorded by the Empatica E4 and RespiBAN Professional devices are included, along with a README file. Additionally, it contains the participants' questionnaire responses, which include timestamps for each recorded condition based on the assigned study protocol version. It is worth noting that this data has not been pre-processed. For the EmoWear dataset, both raw and pre-processed data are found, being the latter the set used in this work. As before, this dataset is organised into 48 folders, each corresponding to a participant, and each folder contains the data from the BH3 and Empatica E4 along with the timestamps for each phase, the answers to the questionnaires, and some additional parameters.

For the sake of simplicity, only the data of the physiological signals of interest (EDA, HR, and TEMP) have been extracted for further analysis. To facilitate this analysis, the data for each signal has been organised by specific emotional states. Python and Jupyter notebooks have been used for this process.

For the WESAD dataset, the data for each of the three signals are converted into individual dataframes. These dataframes are then divided according to the timestamp and the sampling rate of each signal, resulting in new dataframes for each emotional state. It is noteworthy that the original timestamp values are provided in minutes and seconds, but they are converted into seconds to ease the division process. As a result, a series of CSV files is generated, following the naming pattern: $\texttt{s}_i$`_signal_emotion.csv`, where $\texttt{s}_i$ represents the subject's ID, `signal` refers to one of the following signals {EDA, HR, TEMP}, and `emotion` refers to one of the following emotions {baseline, amusement, stress}.

For the EmoWear dataset, a similar but slightly different process is followed. First, the information of the timestamps, videos and quadrant values (*i.e.,* the video's emotion) is provided in different files. So, it is merged into a single dataframe based on the video's ID, ensuring that the quadrants are correctly assigned. After this procedure, the three CSV files corresponding to the three physiological signals are converted dataframes with two columns: one for the timestamp and the other for the signal's values. To determine the data for each emotion, a function locates the closest timestamp value within the signals dataframe. Once the values are found, the signals dataframe is divided by emotions. As a result, a series of CSV files is generated, following the naming pattern: $\texttt{s}_i$`_signal_video_quadrant.csv`, where $\texttt{s}_i$ represents the subject's ID, `signal` refers to one of the following signals {EDA, HR, TEMP}, `video` refers to the video's ID, and `quadrant` refers to one of the following quadrants {HAHV, LAHV, LALV, HALV}.

After the files have been divided by signal and emotion, features can be calculated. These features summarise the properties and characteristics of a signal. In this work, a function called `calculate_features()` has been implemented to compute the mean (μ), median (Me), standard deviation (σ), minimum (Mi), and maximum (Ma) values of a signal. However, given that the duration of these signals varies across datasets, we opted to segment the signals into 10-second windows (Figure 3). Consequently, features are calculated for each of these windows rather than for the entire signal as a whole: *e.g.,* set $\{\mu_i, \mathrm{Me}_i, \sigma, \mathrm{Mi}_i, \mathrm{Ma}_i\}$ indicates the features of the i-th window. For instance, for a 1-minute signal, six sets of features are computed (one for each window) instead of a single set. This strategy allows observing the evolution of a signal's pattern throughout the entire emotional state.

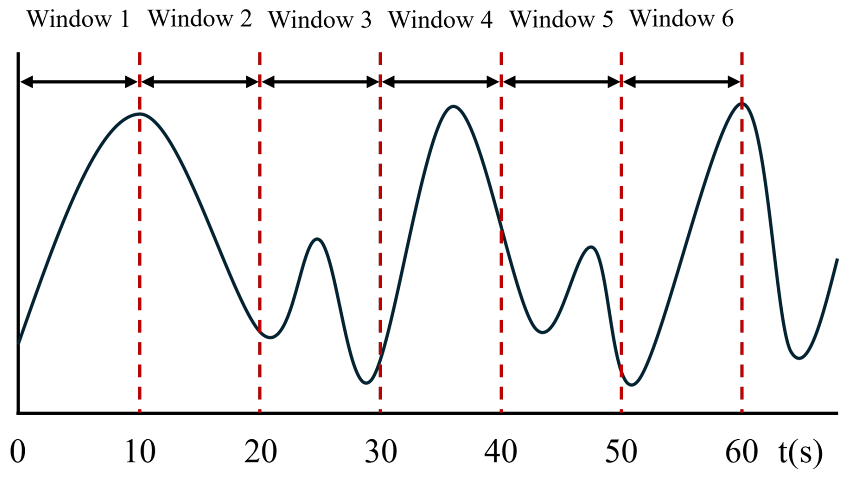


Figure 3: Windows-based signal segmentation.

3.3 Comparison Strategies

Two strategies have been developed to compare samples of signals corresponding to different emotional states: a randomised sample and a segmented sample.

The randomised sample aims to create p groups of n subjects selected at random, where p and n are user-predefined values. To this end, a function called `divide_participants()` has been implemented. In this work, we have opted to create three groups of subjects to ease comparison. Consequently, each group will encompass 5 and 16 subjects each for the WESAD and EmoWear datasets, respectively. Within each group, a virtual *average* subject, which averages the signals of all the subjects within the group to denote the average group behaviour, is also calculated. The random nature of this approach allows assessing whether the behaviour of the physiological signals is homogeneous across participants.

The segmented sample aims to group similar subjects based on specific user-centric variables, such as demographic data. In this work, these variables refer to the participants' age, gender, and body mass index (BMI). Unfortunately, this information has only been found in the WESAD dataset. In this case, since there are only 15 participants, the creation of these groups has been performed manually. This strategy allows assessing whether the behaviour of the physiological signals varies with the contextual conditions of the participants.

To graphically depict the evolution of the physiological signals patterns to enable visual comparison, both strategies employ the `plot_mean_with_average()` function that create a multi-line chart with the signal's windows means (feature) of all the subjects belonging to a specific group.

3.4 Indicators

Beyond graphical comparisons, a quantitative analysis is appropriate for properly comparing the patterns of all the signals within the same group across different groups. Specifically, four indicators have been proposed:

- **Window Variability Index (WVI)**: WVI, which measures the overall variability of all signals belonging to the same group, is the sum of all standard deviations of all windows a signal has. The standard deviation for each window indicates how spread out the data points are around the mean. Hence, if the data points are close together, WVI tends to 0. Conversely, if the data point are widely dispersed, WVI will increase.

$$\text{WVI} = \sum_{i=1}^{n} \sigma_i, \text{ where } n \text{ indicates the number of windows}$$

- **Windows Min-Max (WMM)**: WMM quantifies the range of mean values across different windows by showing the difference between the maximum and minimum mean values. It is calculated as the difference between the highest and lowest mean values across all windows. A greater difference in the mean values across the windows is indicated by a larger WMM value, highlighting significant changes in the signal's level over time.

$$\text{WMM} = \max_{i=1}^{n}(\mu_i) - \min_{i=1}^{n}(\mu_i), \text{ where } n \text{ indicates the number of windows}$$

- **Amplitude Min-Max (AMM)**: AMM provides a relative measure of the peak-to-trough changes in the mean values across windows. It is calculated as the ratio of the maximum mean value to the minimum mean value across the windows, with the growth between the valley and peak values being indicated by this ratio. A value close to 1 is indicated to show that the means across the windows are relatively similar, with less pronounced fluctuations.

$$\text{AMM} = \frac{\max_{i=1}^{n}(\mu_i)}{\min_{i=1}^{n}(\mu_i)}, \text{ where } n \text{ indicates the number of windows}$$

- **Sum of Consecutive Distances (SCD)**: SCD, used to quantify the total change between consecutive windows, offers insights into how the mean values evolve from one window to the next, indicating the amount of increase and decrease in the signal (in absolute value). SCD is calculated as the sum of the absolute differences between the mean values of consecutive windows. A higher SCD value is suggested to indicate more dynamic changes in the signal over time, which can be indicative of rapid transitions or trends at a physiological level.

$$\text{SCD} = \sum_{i=1}^{n-1} |\mu_{i+1} - \mu_i|, \text{ where } n \text{ indicates the number of windows}$$

It is important to note that these indicators are not exhaustive and have been specifically developed for the purposes of this work.

4 Results

This section presents the results obtained after applying the two comparison strategies of the physiological signals in the datasets selected. Specifically, the results of the WESAD and EmoWear datasets are provided in Section 4.1 and Section 4.2, respectively, each of them providing both graphical and quantitative results.

4.1 WESAD Dataset

Next, the results of the WESAD dataset are provided. These results are described according to the two aforementioned comparison strategies: a randomised sample in Section 4.1.1 and a segmented sample in Section 4.1.2. The quantitative analysis through indicators is provided in Section 4.1.3. Finally, a summary of the main findings is provided in Section 4.1.4.

4.1.1 Randomised Sample

The analysis of physiological signals of each individual emotion across the three random groups reveals several trends.

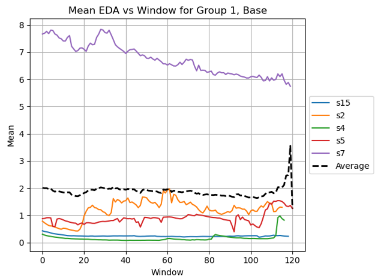
First, regarding the baseline affective state (Figure 4), the TEMP signal generally increases gradually by +1 ºC over time across all three groups, which is considered favourable due to the consistent trend behaviour. Similarly, significant fluctuations over time in all groups are observed with the HR signal. However, the EDA signal does not exhibit a uniform behaviour across the groups: it stays more constant in groups 1 and 3, while more fluctuation is noted in group 2, which makes this trend less reliable due to such inter-group inconsistencies.

With regards to the amusement emotion (Figure 5), TEMP remains steady with minimal variation over time ($<$0.5 ºC), proving to be a consistent behaviour across all groups. Similarly, in the HR signal, two noticeable peaks are generally observed over the time, being some of them more apparent in some subjects than in others. Notwithstanding, these peaks are shown around the same windows, proving a heartbeat acceleration during that time period. Finally, little variation is observed in the EDA signal, which generally remains stable over time.

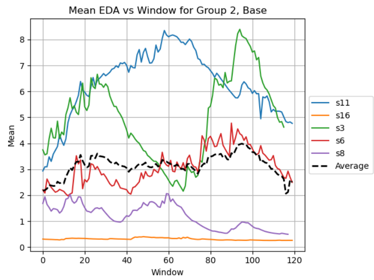
With regards to the stress condition (Figure 6), TEMP remains stable over time with minimal variation ($<$0.5 ºC) across all groups, thus it is considered a favourable trend. The HR signal shows considerable fluctuations over time in all groups, which is also viewed positively for its inter-group consistency. Also, EDA remains uniform with a slight decline in the second half of the period, which could also be seen favourable, although there is a notable exception with subject s6 from the second group, who displays a distinctly different pattern behaviour.

4.1.2 Segmented Sample

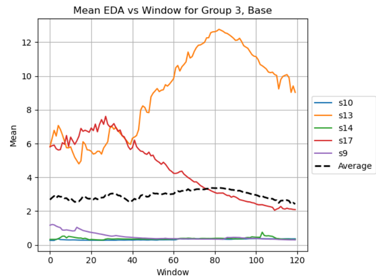
Next, the analysis of the physiological signals based on different demographic factors, such as age, gender, and BMI, is conducted across the different emotions.



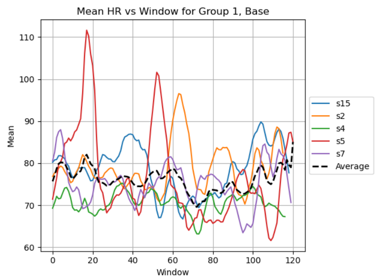
(a) EDA in group 1



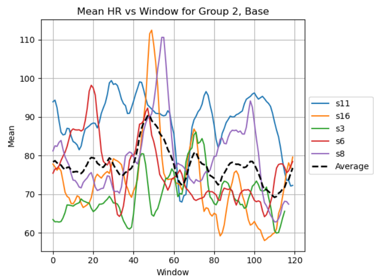
(b) EDA in group 2



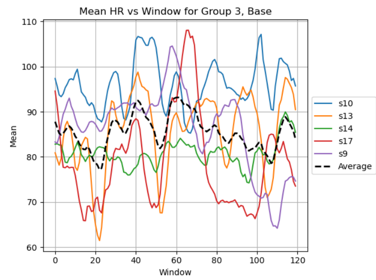
(c) EDA in group 3



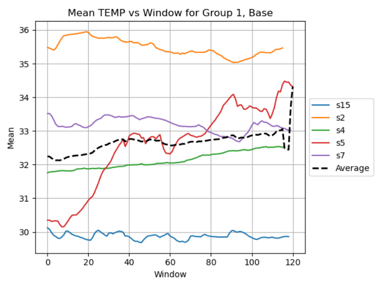
(d) HR in group 1



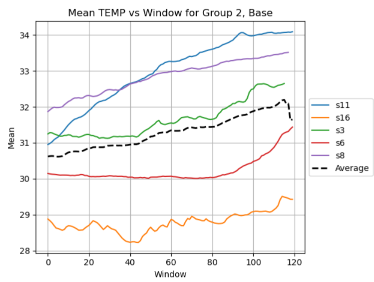
(e) HR in group 2



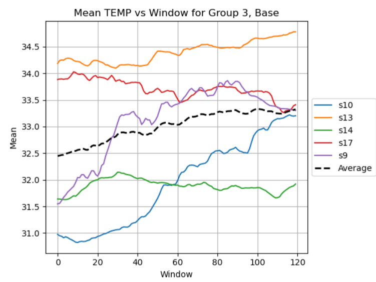
(f) HR in group 3



(g) TEMP in group 1



(h) TEMP in group 2

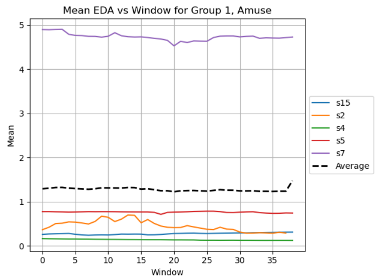


(i) TEMP in group 3

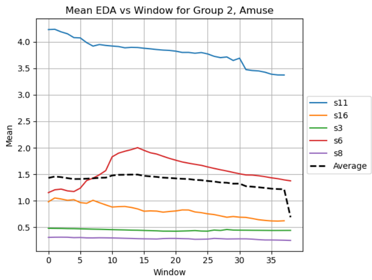
Figure 4: Results of the baseline emotion using a randomised sample strategy in the WESAD dataset.

With regards to age group comparison, a group of subjects aged 28 and older and another group of subjects under 28 have been created to have two similarly balanced groups (with 6 and 9 subjects each). By comparing the physiological signals of each emotion (Figures 7 to 9), a slight increase in the TEMP signal is noted over time for the older group during the amusement state, while the younger group shows a slight decrease in TEMP. In the stress state, an EDA peak is observed around window 10 for the older group, followed by a slight decline, while the younger group maintains a more constant EDA signal. Overall, no significant behavioural differences are evident between the age groups, which may be expected given the relatively narrow age range of the dataset's population (24-35 years), suggesting minimal age-related variability at these ages.

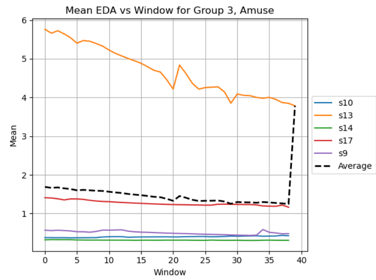
Regarding gender differences (Figures 10 to 12), it is observed that in the baseline state, the variation in the range of TEMP is greater in women compared to men. Sim-



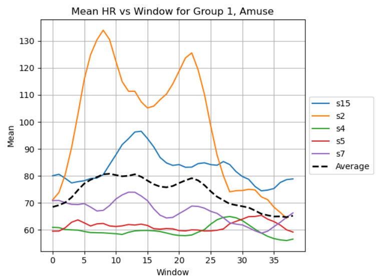
(a) EDA in group 1



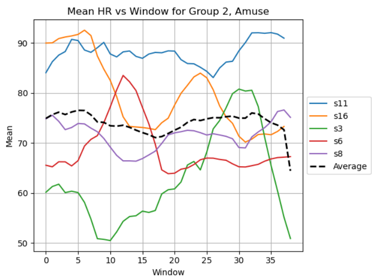
(b) EDA in group 2



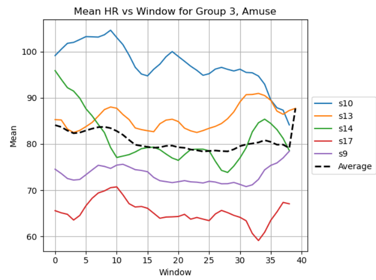
(c) EDA in group 3



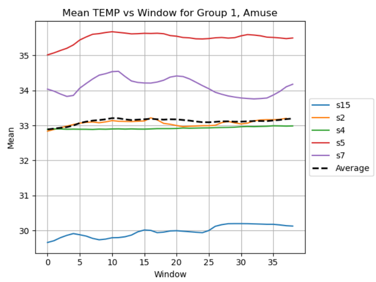
(d) HR in group 1



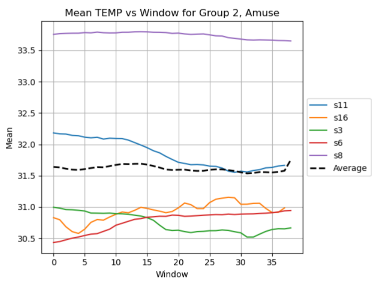
(e) HR in group 2



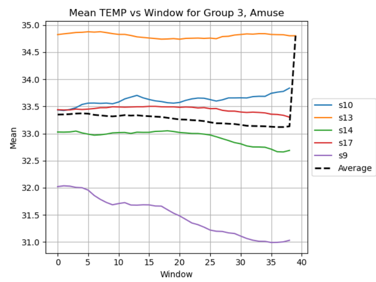
(f) HR in group 3



(g) TEMP in group 1



(h) TEMP in group 2

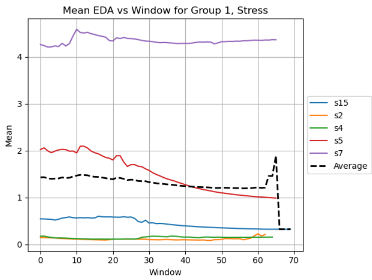


(i) TEMP in group 3

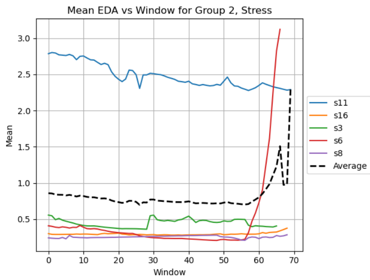
Figure 5: Results of the amusement emotion using a randomised sample strategy in the WESAD dataset.

ilarly, women show greater variation in HR. Additionally, the EDA in women tends to decrease after the midpoint of the period, while in men, it either remains stable or slightly increases. In the stress state, women exhibit HR peaks around windows 42 and 58, a behaviour not observed in men during those windows. However, these observations cannot be statistically significant due to the unbalanced sample size, with only 3 females compared to 12 males.

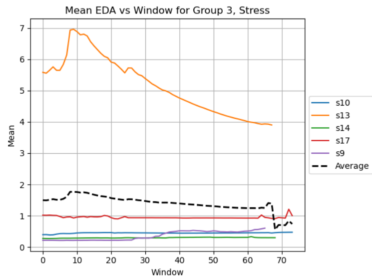
When examining the physiological signals based on BMI (Figures 13 to 15), distinct patterns can be observed. Three groups have been created, distinguishing between underweight subjects, normal-weight subjects, and overweight subjects. However, these groups are not balanced, having only one underweight subject and three overweight subjects, whereas the rest of the 11 subjects have a normal BMI. Therefore, these observations cannot be statistically significant due to such imbalance. In the baseline state, a significant and accelerated increase in TEMP is observed in the underweight group, while the other two groups show a more gradual increase. The underweight



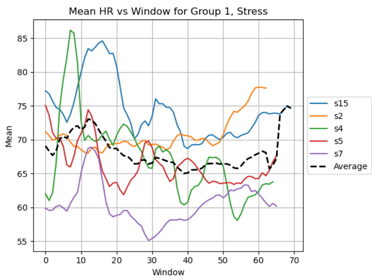
(a) EDA in group 1



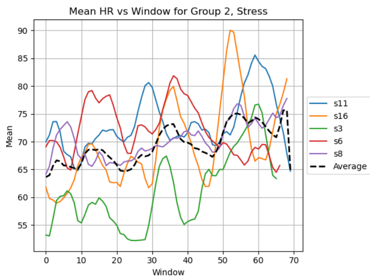
(b) EDA in group 2



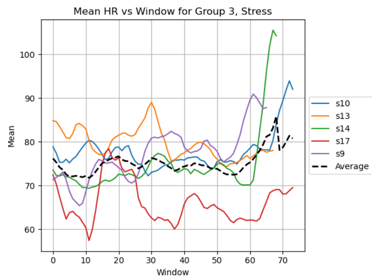
(c) EDA in group 3



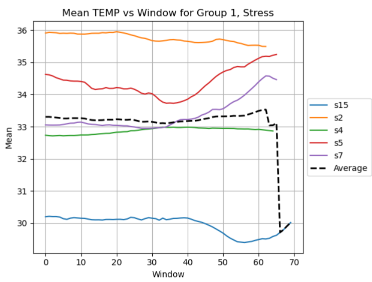
(d) HR in group 1



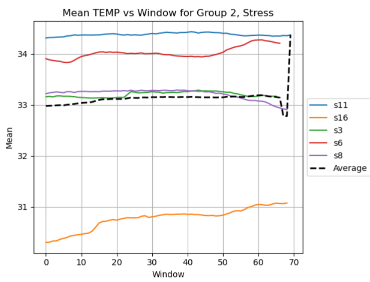
(e) HR in group 2



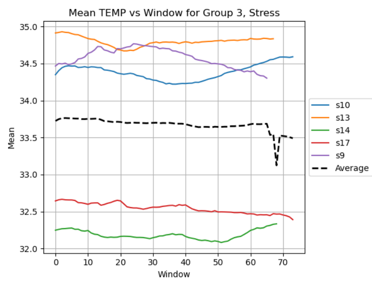
(f) HR in group 3



(g) TEMP in group 1



(h) TEMP in group 2



(i) TEMP in group 3

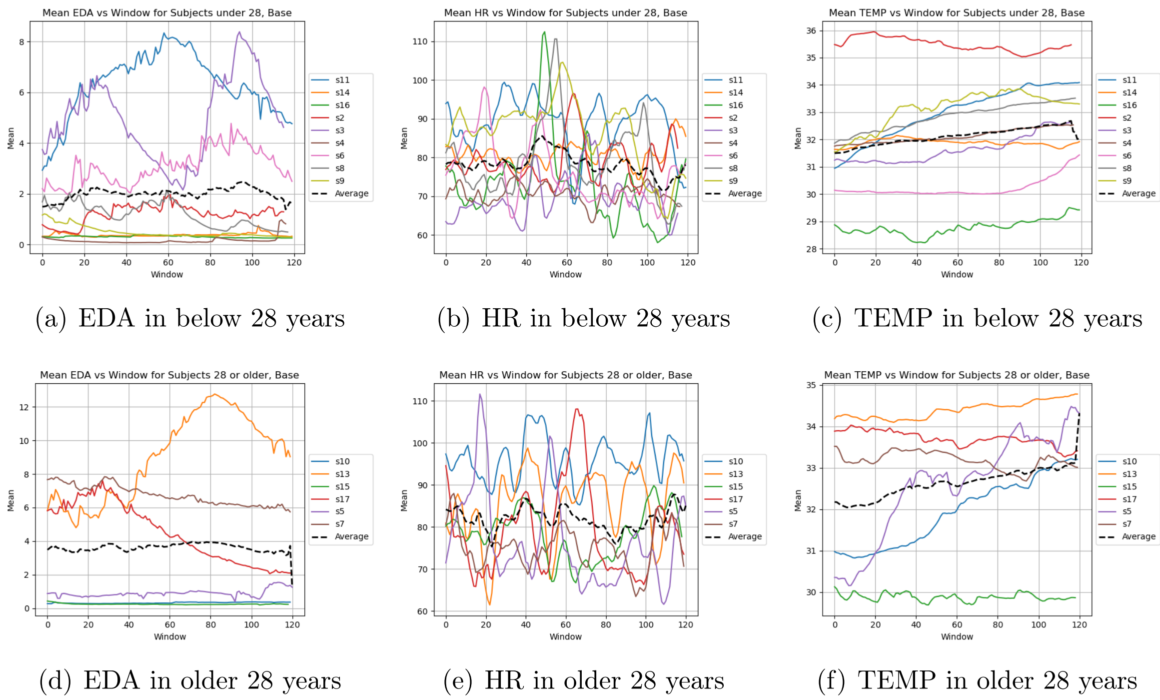
Figure 6: Results of the stress emotion using a randomised sample strategy in the WESAD dataset.

group also exhibits a pronounced peak in EDA at the middle of the period, which is not seen in the other groups, where EDA remains more constant. During the amusement state, the underweight group shows a constant decrease in both TEMP and EDA, while the other two groups either remain stable or experience a slight increase in these signals. The HR of individuals with normal weight remains relatively constant, whereas more fluctuations are observed in the underweight and overweight groups. In the stress state, the HR of individuals with normal weight increases, while the other two groups experience more fluctuations.

4.1.3 Quantitative Analysis

Tables 2 to 4 present the quantitative results of the described indicators for the EDA, HR, and TEMP signals, respectively. The numerical results are aligned with the observations from the graphics. EDA signals have high WVI and SCD indicators in the baseline state, while amusement and stress states result in lowers WVI and

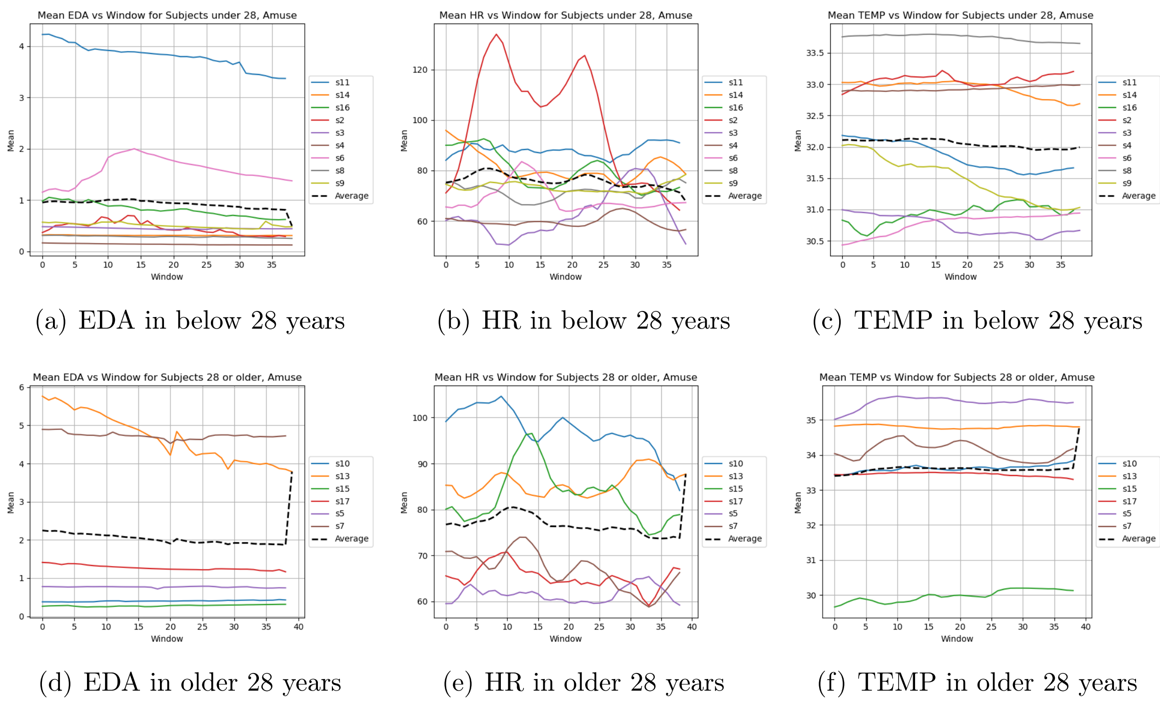
SCD values, suggesting that non-neutral states have less EDA variability. Similarly, HR-related indicators show a similar behaviour. With regards to TEMP-related indicators, a similar pattern is noticeable when considering the WVI indicator, but not



(a) EDA in below 28 years (b) HR in below 28 years (c) TEMP in below 28 years

(d) EDA in older 28 years (e) HR in older 28 years (f) TEMP in older 28 years

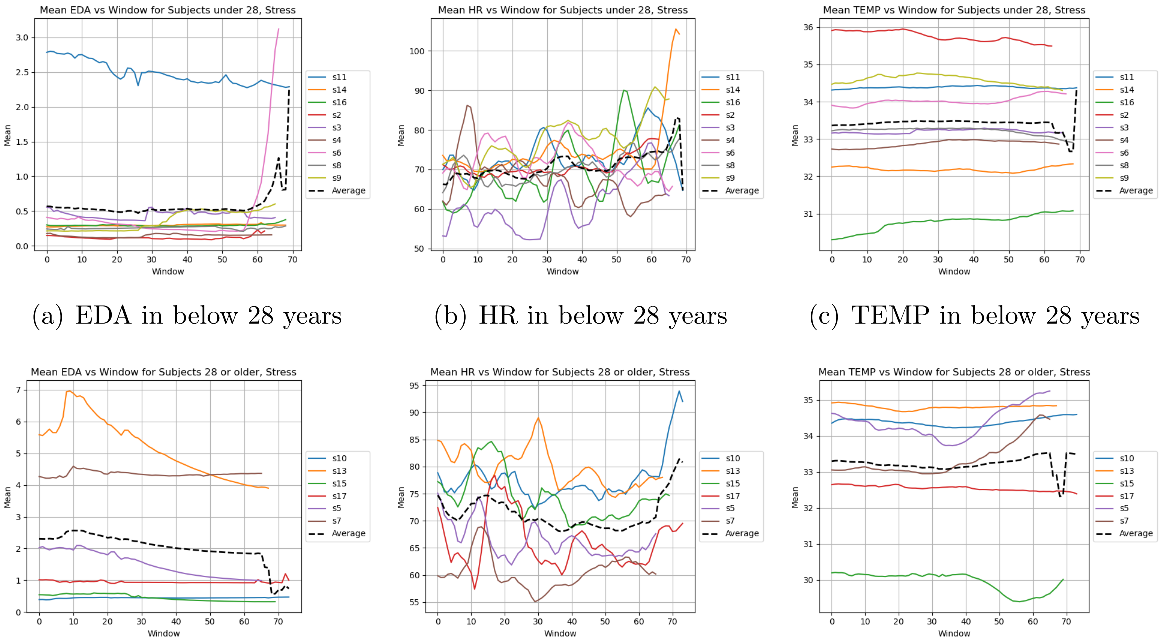
Figure 7: Results of the baseline emotion by age group in the WESAD dataset.



(a) EDA in below 28 years (b) HR in below 28 years (c) TEMP in below 28 years

(d) EDA in older 28 years (e) HR in older 28 years (f) TEMP in older 28 years

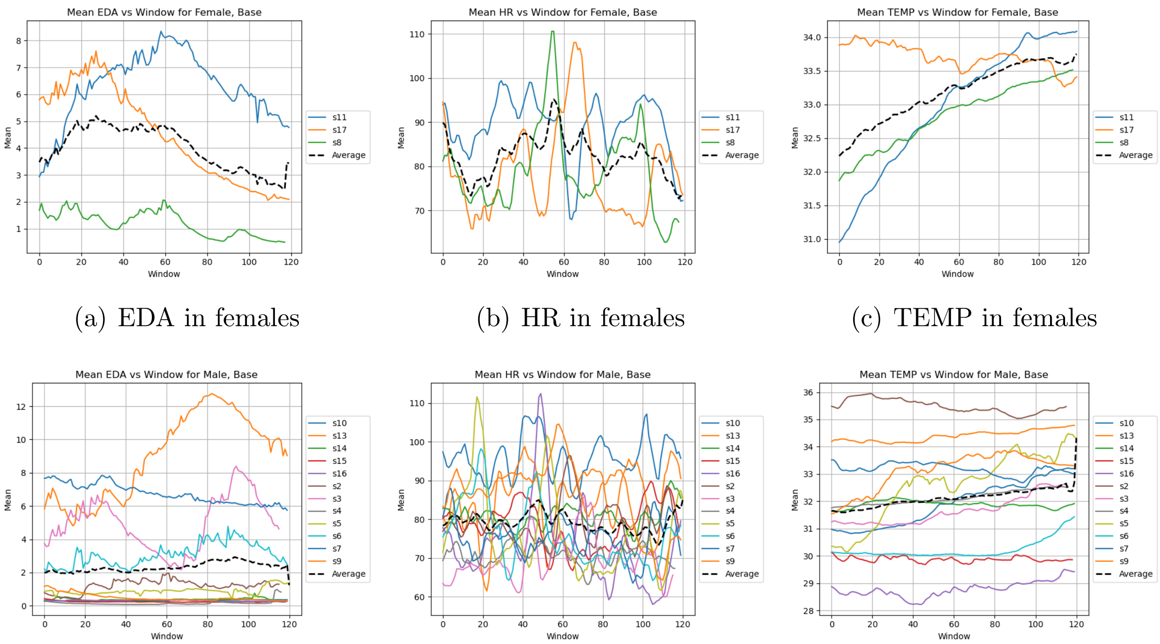
Figure 8: Results of the amusement emotion by age group in the WESAD dataset.



(a) EDA in below 28 years (b) HR in below 28 years (c) TEMP in below 28 years

(d) EDA in older 28 years (e) HR in older 28 years (f) TEMP in older 28 years

Figure 9: Results of the stress emotion by age group in the WESAD dataset.

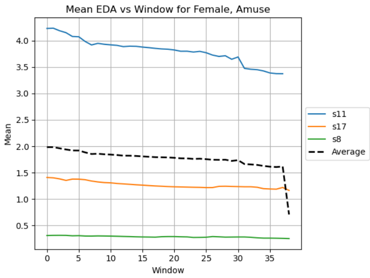


(a) EDA in females (b) HR in females (c) TEMP in females

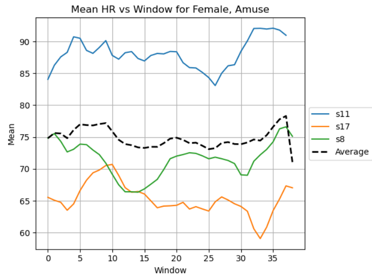
(d) EDA in males (e) HR in males (f) TEMP in males

Figure 10: Results of the baseline emotion by gender in the WESAD dataset.

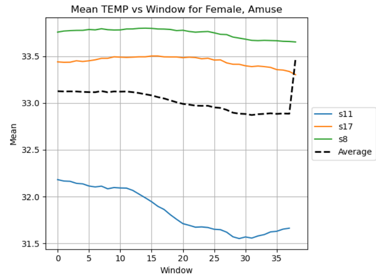
the SCD one. Conversely, WMM and AMM do not seem suitable for distinguishing among emotional states, thus suggesting that working with absolute values might not be relevant for this purpose.



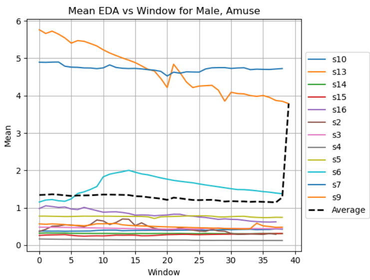
(a) EDA in females



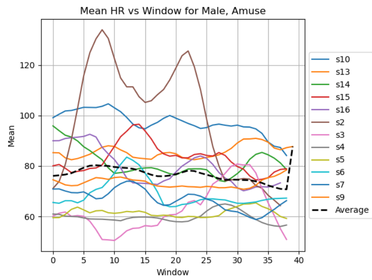
(b) HR in females



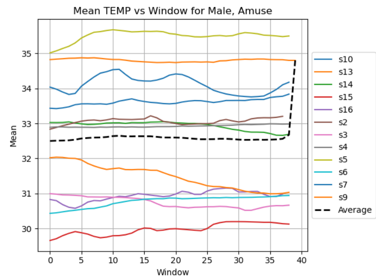
(c) TEMP in females



(d) EDA in males

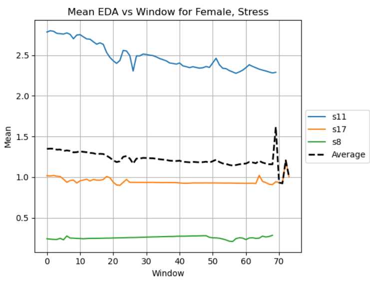


(e) HR in males

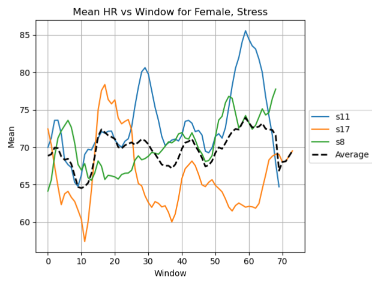


(f) TEMP in males

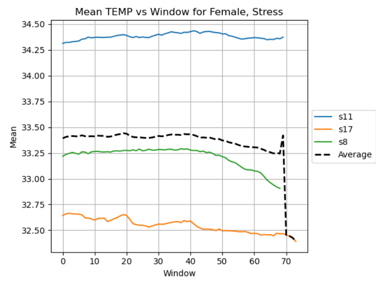
Figure 11: Results of the amusement emotion by gender in the WESAD dataset.



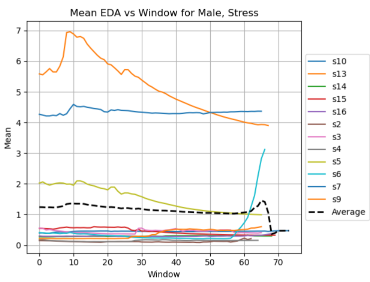
(a) EDA in females



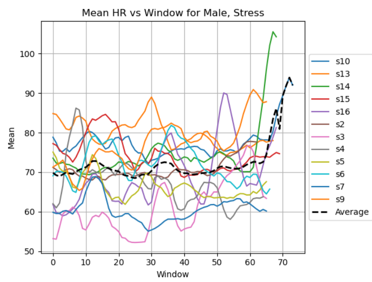
(b) HR in females



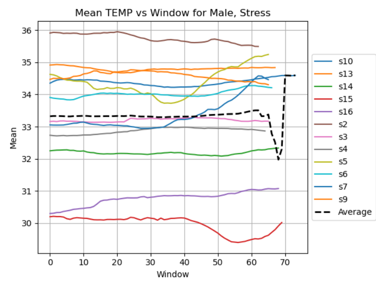
(c) TEMP in females



(d) EDA in males

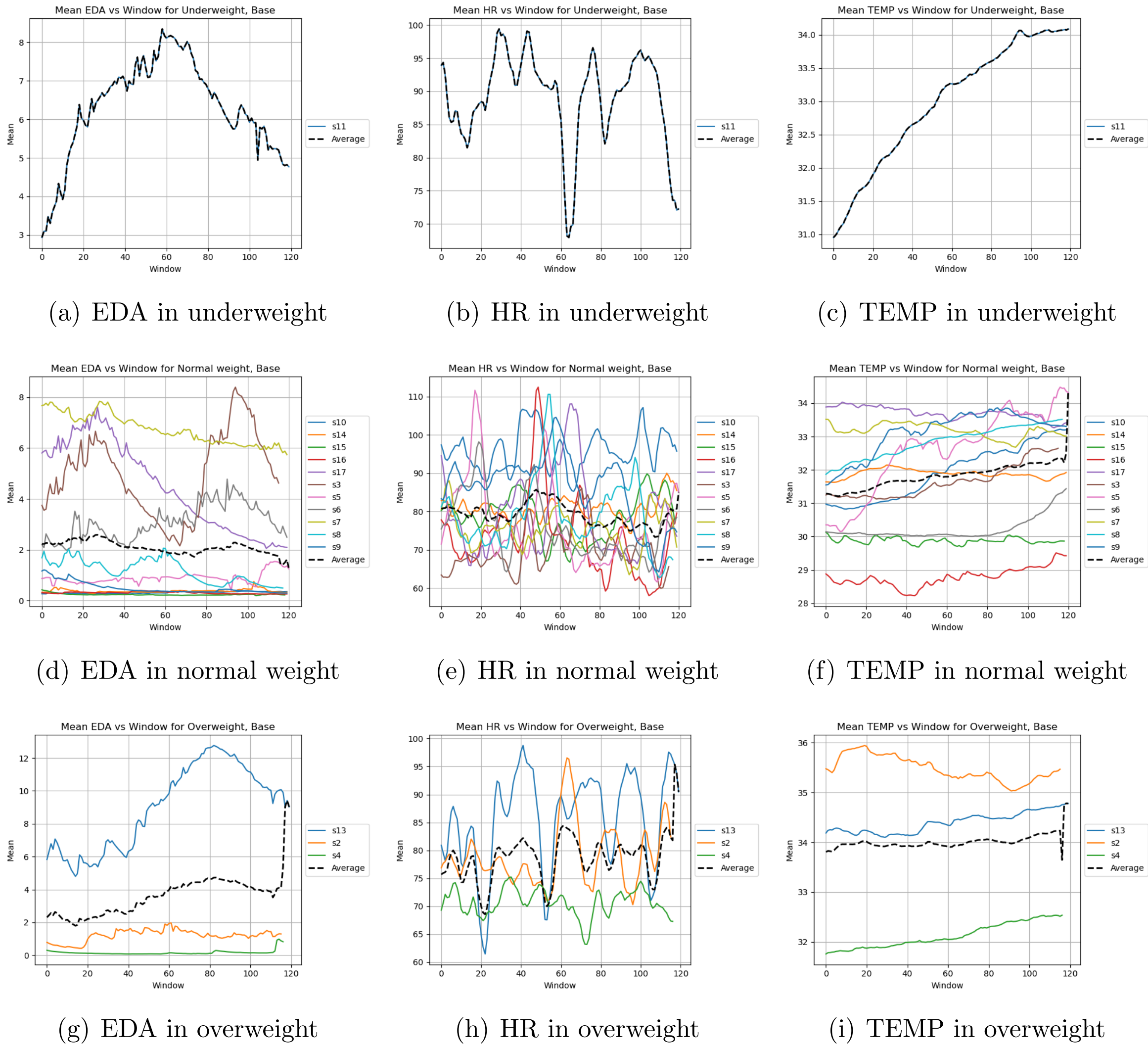


(e) HR in males



(f) TEMP in males

Figure 12: Results of the stress emotion by gender in the WESAD dataset.



(a) EDA in underweight (b) HR in underweight (c) TEMP in underweight

(d) EDA in normal weight (e) HR in normal weight (f) TEMP in normal weight

(g) EDA in overweight (h) HR in overweight (i) TEMP in overweight

Figure 13: Results of the baseline emotion by BMI in the WESAD dataset.

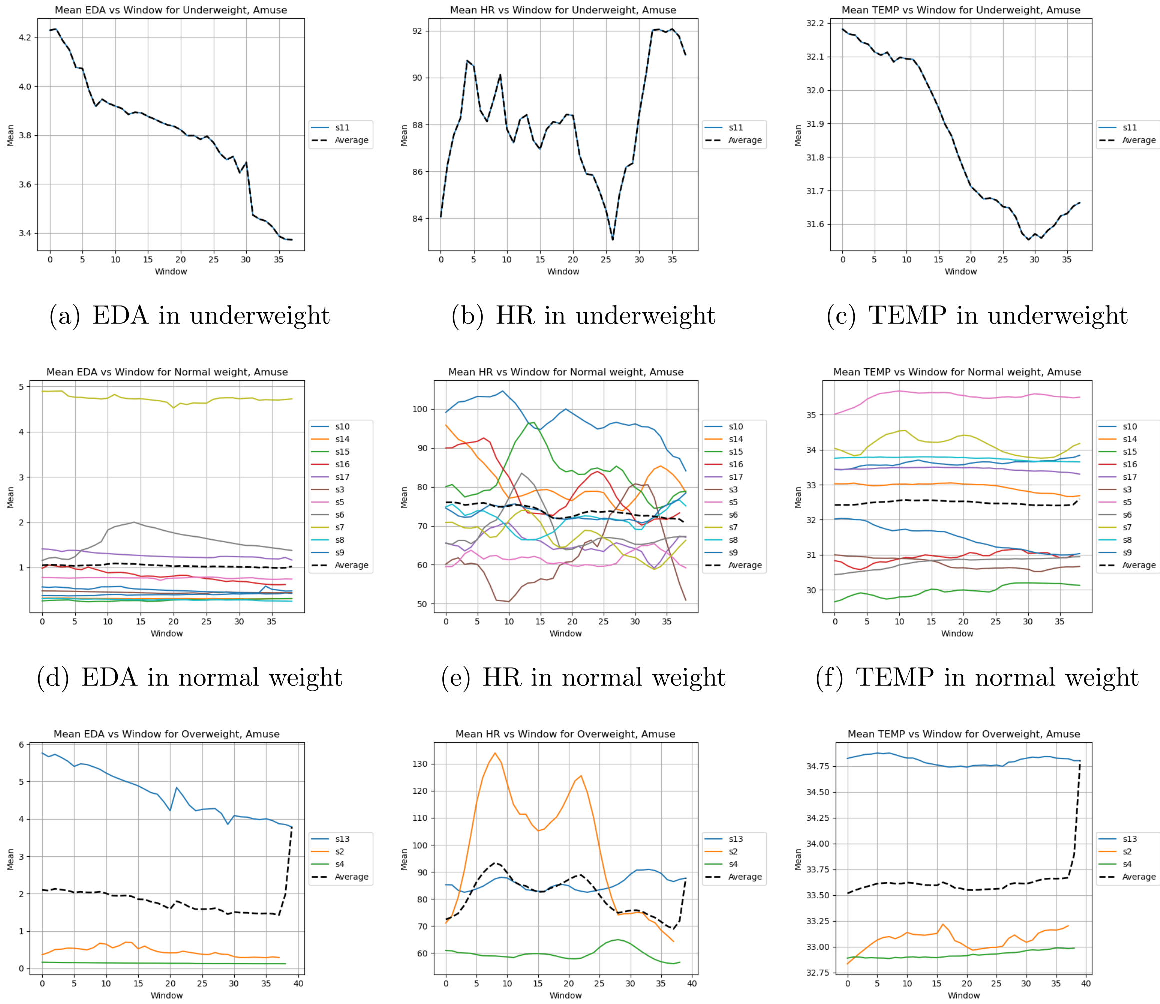
Table 2: Indicators for the EDA signals in the WESAD dataset.

		Randomised Sample			Segmented Sample						
	Index	**Group 1**	**Group 2**	**Group 3**	**Females**	**Males**	**< 28**	**≥ 28**	**Underweight**	**Normal weight**	**Overweight**
Amusement	**WVI**	2.015	2.247	3.538	1.187	6.613	3.702	4.098	0.840	3.307	3.652
	WMM	0.251	0.804	2.539	1.274	2.640	0.521	1.915	0.861	0.098	2.356
	AMM	1.205	2.164	3.037	2.799	3.306	2.047	2.024	1.255	1.099	2.649
	SCD	0.604	0.962	3.423	1.345	3.120	0.663	2.708	1.083	0.266	3.805
Baseline	**WVI**	18.575	56.846	28.725	31.590	72.556	67.047	37.099	15.433	66.459	22.254
	WMM	2.303	1.923	0.962	2.722	1.672	1.079	2.706	5.413	1.341	7.616
	AMM	2.844	1.932	1.397	2.100	2.339	1.778	3.167	2.846	2.074	5.240
	SCD	7.067	10.838	6.093	10.512	6.156	6.316	8.589	20.425	4.997	14.785
Stress	**WVI**	3.310	4.425	4.889	2.916	9.707	5.726	6.898	1.475	7.684	3.465
	WMM	1.569	1.591	1.209	0.692	1.088	1.818	2.004	0.524	0.638	2.492
	AMM	5.826	3.278	3.179	1.749	3.986	4.857	4.554	1.230	2.466	2.728
	SCD	2.812	3.139	2.355	2.425	2.168	3.005	2.791	1.970	1.655	4.181

4.1.4 Summary of Findings

The physiological signals across different emotional states are analysed in both the randomised and segmented samples, revealing several key trends.

In the randomised samples, TEMP exhibits greater fluctuations with frequent rises



(a) EDA in underweight (b) HR in underweight (c) TEMP in underweight

(d) EDA in normal weight (e) HR in normal weight (f) TEMP in normal weight

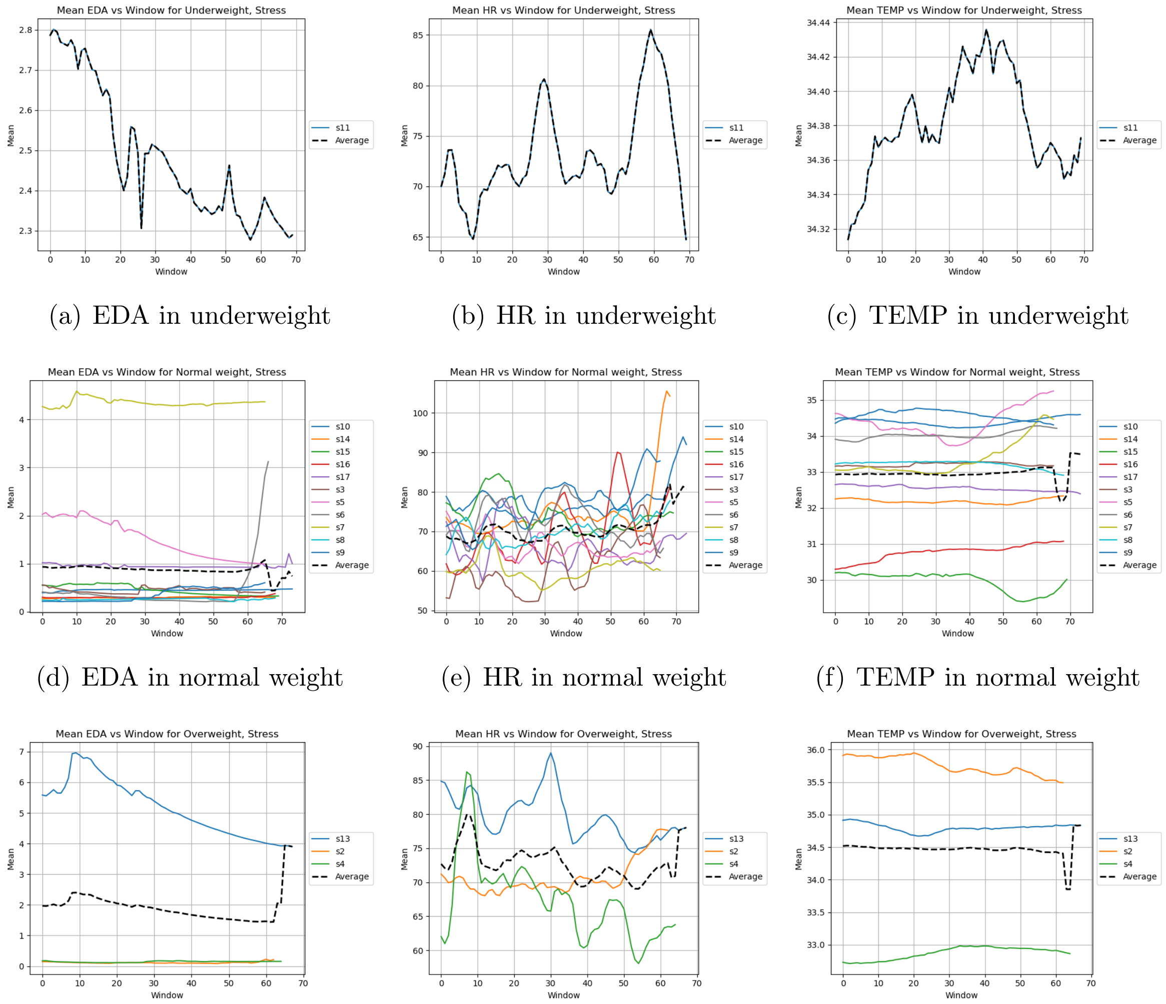
(g) EDA in overweight (h) HR in overweight (i) TEMP in overweight

Figure 14: Results of the amusement emotion by BMI in the WESAD dataset.

Table 3: Indicators for the HR signals in the WESAD dataset.

		Randomised Sample			**Segmented Sample**						
	Index	**Group 1**	**Group 2**	**Group 3**	**Females**	**Males**	**< 28**	**≥ 28**	**Underweight**	**Normal weight**	**Overweight**
Amusement	**WVI**	100.335	79.816	60.219	36.283	204.088	162.436	77.934	12.964	157.128	70.277
	WMM	16.050	12.100	9.317	7.221	16.950	12.963	14.027	8.995	5.446	24.517
	AMM	1.247	1.187	1.118	1.101	1.239	1.191	1.190	1.108	1.077	1.355
	SCD	37.126	26.104	22.494	25.186	35.445	27.591	28.409	34.143	11.907	78.781
Baseline	**WVI**	241.734	293.189	285.429	182.383	637.970	445.323	375.030	55.973	606.210	158.169
	WMM	15.488	21.244	16.033	22.469	11.712	14.006	12.438	31.458	12.384	26.821
	AMM	1.222	1.307	1.207	1.309	1.159	1.195	1.165	1.463	1.168	1.391
	SCD	83.257	95.317	100.660	106.450	68.248	56.971	81.900	171.503	58.651	137.131
Stress	**WVI**	92.641	150.710	114.767	81.051	277.067	234.810	123.308	29.552	270.941	57.625
	WMM	9.970	12.006	14.065	9.398	25.430	18.404	13.366	20.863	14.878	10.869
	AMM	1.153	1.188	1.196	1.145	1.371	1.284	1.196	1.322	1.221	1.157
	SCD	36.708	57.254	47.783	48.381	56.123	51.709	35.601	90.458	44.629	47.702

and falls during the baseline state, while amusement and stress states show a more gradual trend. However, TEMP alone does not effectively differentiate between amusement and stress. HR shows more fluctuation and frequent changes during baseline and stress states, while a more gradual peak-valley-peak pattern is observed during amusement. For EDA, constant fluctuations are observed in the baseline state, with a more gradual trend in both amusement and stress states. However, distinguishing between amuse-



(a) EDA in underweight (b) HR in underweight (c) TEMP in underweight

(d) EDA in normal weight (e) HR in normal weight (f) TEMP in normal weight

(g) EDA in overweight (h) HR in overweight (i) TEMP in overweight

Figure 15: Results of the stress emotion by BMI in the WESAD dataset.

Table 4: Indicators for the TEMP signals in the WESAD dataset.

		Randomised Sample			**Segmented Sample**						
	Index	**Group 1**	**Group 2**	**Group 3**	**Females**	**Males**	**< 28**	**≥ 28**	**Underweight**	**Normal weight**	**Overweight**
Amusement	**WVI**	3.403	2.921	2.748	1.467	7.606	5.198	3.874	0.568	6.923	1.581
	WMM	0.321	0.217	1.682	0.604	2.301	0.183	1.400	0.629	0.186	1.285
	AMM	1.009	1.006	1.050	1.018	1.070	1.005	1.041	1.019	1.005	1.038
	SCD	0.668	0.601	2.016	0.923	2.583	0.337	1.660	0.815	0.513	1.521
Baseline	**WVI**	10.819	9.004	9.332	5.101	24.054	16.371	12.784	1.783	22.381	4.991
	WMM	2.189	1.583	0.884	1.516	2.722	1.179	2.274	3.139	3.109	1.134
	AMM	1.068	1.051	1.027	1.047	1.086	1.037	1.070	1.101	1.099	1.033
	SCD	4.657	2.366	1.586	1.986	3.781	2.026	3.319	3.425	3.990	2.926
Stress	**WVI**	5.139	4.488	4.580	2.721	11.486	7.886	6.321	0.967	10.816	2.424
	WMM	3.820	1.593	0.639	1.050	2.618	1.704	1.208	0.122	1.332	0.986
	AMM	1.128	1.048	1.019	1.032	1.081	1.052	1.037	1.003	1.041	1.029
	SCD	5.059	2.317	1.329	1.626	4.647	2.814	3.292	0.474	2.714	1.887

ment and stress using EDA alone is challenging, as clearly seen in subject s13 from the third group. Given these observations, both TEMP and EDA are useful for distinguishing between baseline states and emotional responses, as they tend to change with any emotional arousal. However, HR is crucial for specifically differentiating between stress and amusement, as it provides more distinct information for these emotional states. A summary of these findings is provided in Table 5.

In the segmented samples, distinct differences based on demographic and physiological factors are observed. Age-related differences are noted, with a slight increase in TEMP for older individuals during amusement and varied EDA patterns in stress. Sex differences are observed, showing greater TEMP and HR variability in women compared to men, with EDA trends diverging after the middle point of the exercise. Analysis based on BMI indicate that underweight individuals exhibit more pronounced changes in TEMP and EDA, while normal-weight individuals show more stable HR during stress. However, due to imbalances in group sizes, these findings are not considered statistically significant. A summary of these findings is provided in Table 6.

Overall, valuable insights are provided by both samples, though the segmented analysis suggests that demographic factors may influence physiological responses to emotions, with statistical significance limited by sample size imbalances.

4.2 EmoWear Dataset

Next, the results of the EmoWear dataset are provided. These results are again described according to the two aforementioned comparison strategies: a randomised sample in Section 4.2.1 and a segmented sample in Section 4.2.2. The quantitative analysis through indicators is provided in Section 4.2.3. Finally, a summary with the main findings is provided in Section 4.2.4.

4.2.1 Randomised Sample

Several patterns are revealed by analysing the physiological signals of each emotion (*i.e.,* quadrant) across the three random groups of individuals. For the sake of clarity, a mnemonic name that properly summarises each quadrant has been given: "Happy" for HAHV, "Relaxed" for LAHV, "Sad" for LALV, and "Angry" for HALV.

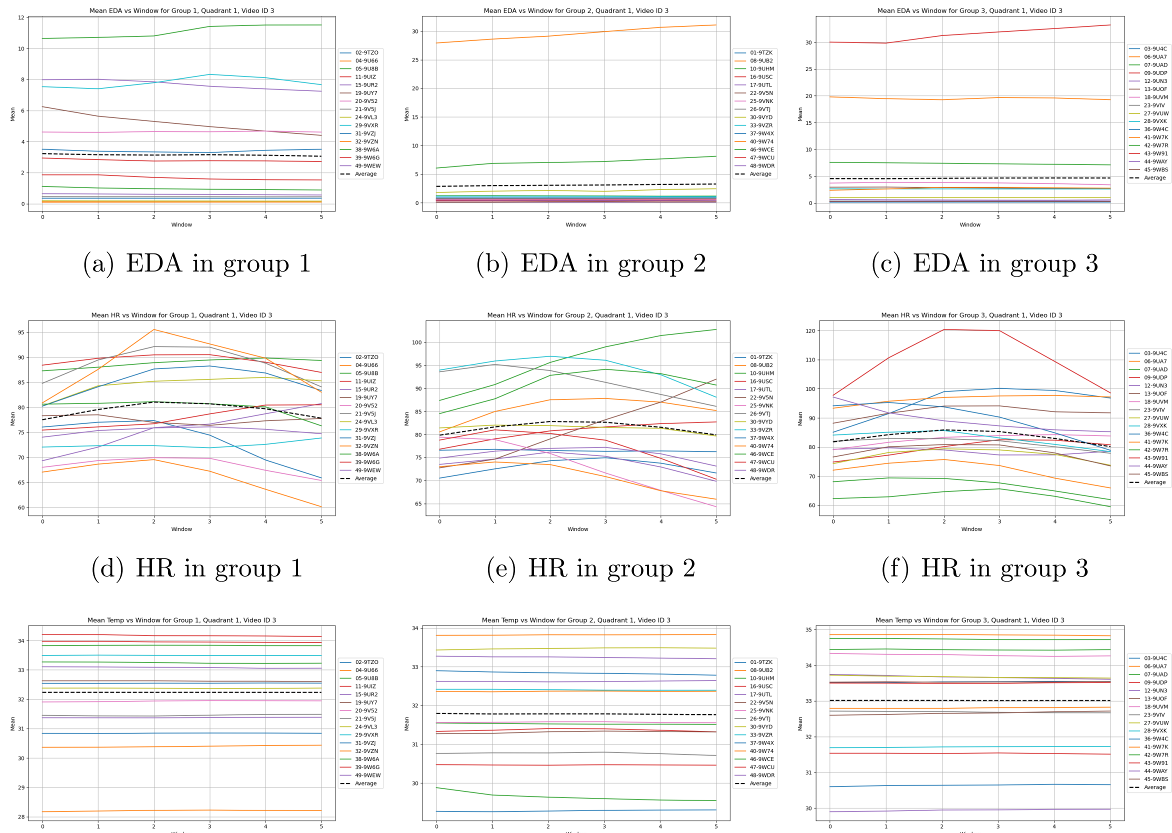
For the happy emotion (Figure 16), the TEMP signal remains completely constant over time, which is considered favourable due to the similar trend across subjects. In contrast, HR exhibits a gradual increase towards the midpoint of the time period, followed by a gradual decrease, returning to approximately the initial value. This consistency is viewed positively. EDA also remains largely constant over time, although

Table 5: Summary of findings using a randomised sample strategy in the WESAD dataset.

	TEMP	HR	EDA
Baseline	Gradual increase (+1°C), higher fluctuation in signal view	Significant fluctuations, frequent changes	Constant in groups 1 & 3, fluctuates in group 2; constant fluctuations in signal view
Amusement	Steady with minimal variation (<0.5°C)	Two noticeable peaks; gradual peak-valley-peak	Uniform
Stress	Consistent with minimal variation (<0.5°C)	Considerable fluctuations	Uniform with slight decline in second half, challenging to distinguish from amusement

Table 6: Summary of findings using a segmented sample strategy in the WESAD dataset.

	Age group	Gender	BMI
TEMP	Amusement: >28 shows slight increase, $\leq$ 28 shows slight decrease	Baseline: Greater variation in women	Baseline: Significant increase in underweight group; gradual in others
HR	Stress: Notably constant or variable depending on age	Baseline: Greater variation in women	Amusement: Constant in normal weight; more fluctuations in under and overweight groups
EDA	Stress: >28 shows peak around window 10, followed by slight decline; $\leq$ 28 constant	Baseline: EDA decreases after midpoint in women; stable or slightly increasing in men	Baseline: Pronounced peak midway in underweight group, constant in others
General observations	Minimal differences due to narrow age range	Observations not statistically significant due to sample imbalance	Observations not statistically significant due to sample imbalance



(a) EDA in group 1 (b) EDA in group 2 (c) EDA in group 3

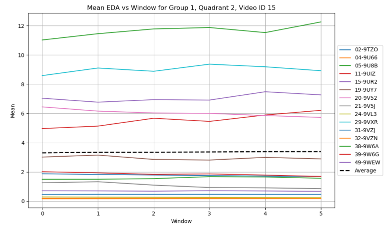
(d) HR in group 1 (e) HR in group 2 (f) HR in group 3

(g) TEMP in group 1 (h) TEMP in group 2 (i) TEMP in group 3

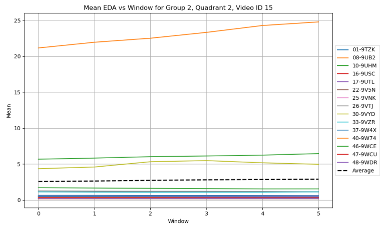
Figure 16: Results of the happy emotion using a randomised sample strategy in the EmoWear dataset.

there are some subjects who show slight increases or decreases starting from the second window. This variation is not considered strictly positive or negative but remains neutral.

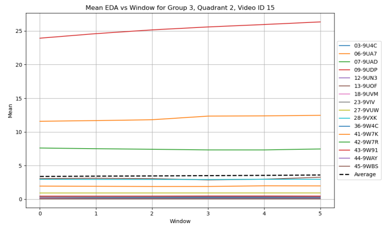
With regards to the relaxed state (Figure 17), TEMP also remains completely constant over time across all groups, a favourable and consistent trend. Similarly, EDA



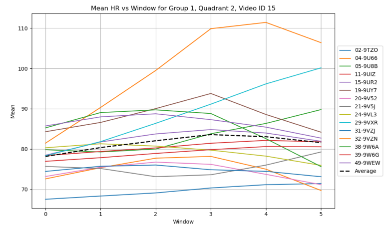
(a) EDA in group 1



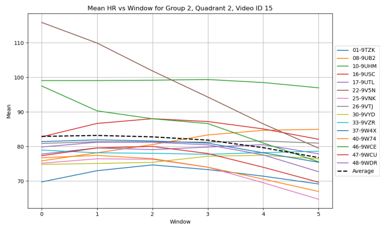
(b) EDA in group 2



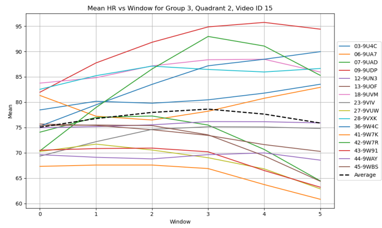
(c) EDA in group 3



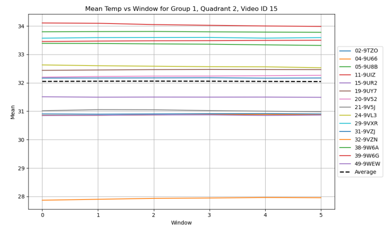
(d) HR in group 1



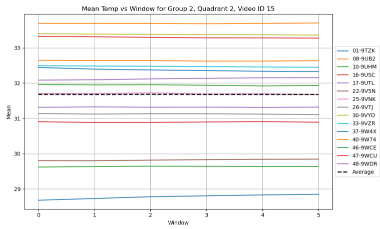
(e) HR in group 2



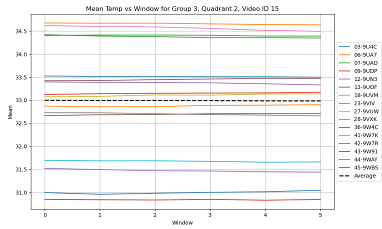
(f) HR in group 3



(g) TEMP in group 1



(h) TEMP in group 2



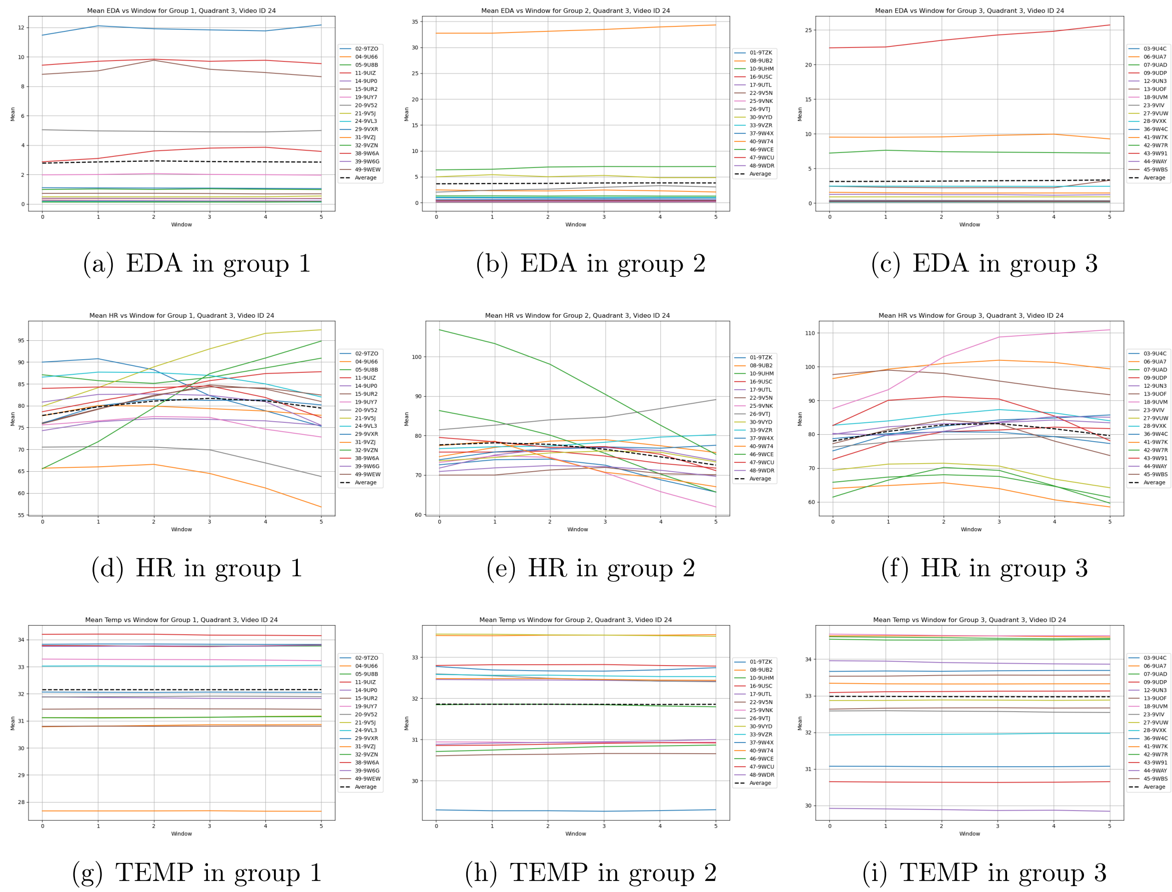
(i) TEMP in group 3

Figure 17: Results of the relaxed emotion using a randomised sample strategy in the EmoWear dataset.

remains completely constant over time, which is seen as positive for its consistency. However, HR shows a variety of behaviours among subjects: some subjects experience a constant increase, others a constant decrease, and some display a peak at the mid-point of the signal. This variability in HR trends is less suitable for our purposes due to the lack of uniformity.

Focusing on the sadness emotion (Figure 18), the patterns are pretty similar to those of the relaxed state. This is, TEMP and EDA signals remain completely constant over time, but HR exhibits a diverse range of behaviours across subjects, with some showing a constant increase, others a decrease, and some having a peak at the midpoint. These patterns are suitable as sadness subjects show similar behaviour patterns, but it hinders distinguishing between relaxed and sad subjects.

Last, in the angry state (Figure 19), TEMP and EDA signals continue to remain completely constant over time, showing similar behaviours and patterns. In the case of the HR signal, it shows a similar pattern to the happy emotion, with a gradual increase towards the midpoint, followed by a gradual decrease to approximately the initial value. Therefore, these patterns are uniform across angry subjects, but distinguishing between happy and angry subjects would be challenging.



(a) EDA in group 1 (b) EDA in group 2 (c) EDA in group 3

(d) HR in group 1 (e) HR in group 2 (f) HR in group 3

(g) TEMP in group 1 (h) TEMP in group 2 (i) TEMP in group 3

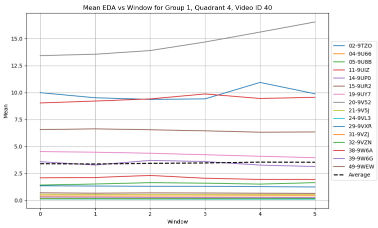
Figure 18: Results of the sad emotion using a randomised sample strategy in the EmoWear dataset.

4.2.2 Segmented Sample

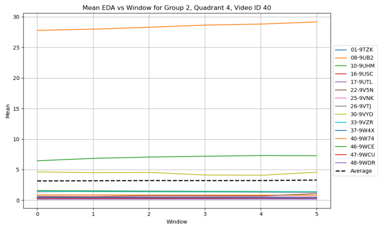
Next, the analysis of the physiological signals based on different demographic factors, such as age and gender, is conducted across the four emotional states.

Regarding age group comparison, two balanced groups of subjects have been created for such comparison, one of them with 27 subjects aged below 30, and another with 19 subjects aged 30 or above. By comparing the physiological signals of each emotion (Figures 20 to 23), TEMP and EDA signals do not show distinguishable patterns across different age groups. In contrast, slight differences in HR are observable. Subjects below 30 years experience a less pronounced peak when they are happy, while subjects above 30 have a more pronounced peak around the second window with a more noticeable decrease. Overall, no significant behavioural differences are evident between the age groups, which may be expected given the relatively narrow age range of the dataset's population (young adults between 21 and 45 years), suggesting minimal age-related variability at this population cohort.

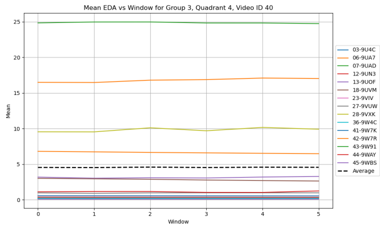
With regards to gender-based comparison (Figures 24 to 27), no different patterns among men and women are observed in any of the signals. This suggests that gender does not result in different physiological patterns in this dataset.



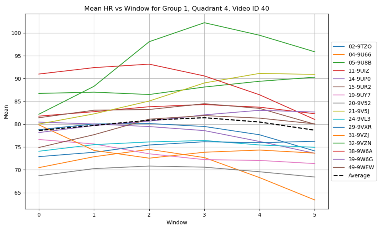
(a) EDA in group 1



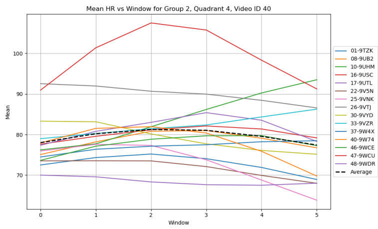
(b) EDA in group 2



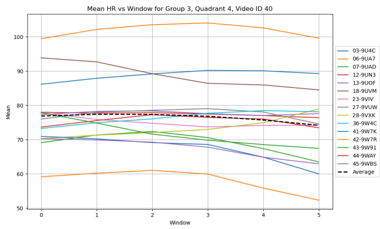
(c) EDA in group 3



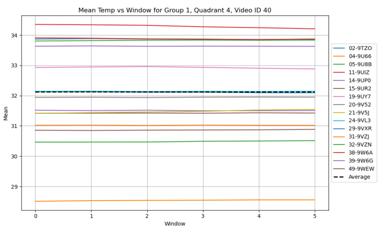
(d) HR in group 1



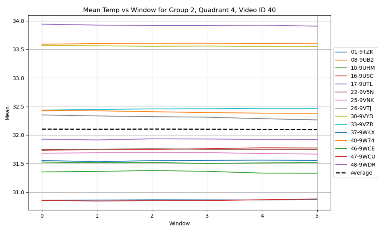
(e) HR in group 2



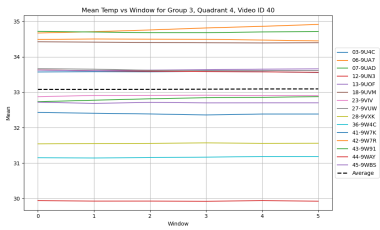
(f) HR in group 3



(g) TEMP in group 1

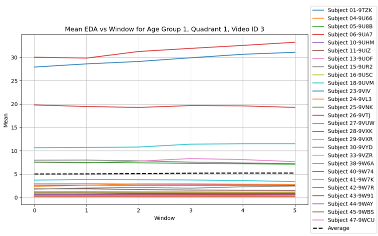


(h) TEMP in group 2

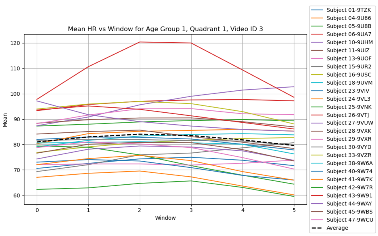


(i) TEMP in group 3

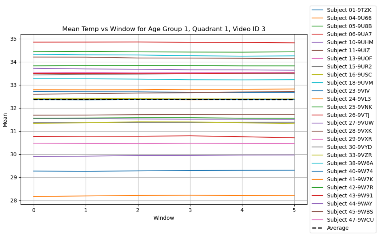
Figure 19: Results of the angry emotion using a randomised sample strategy in the EmoWear dataset.



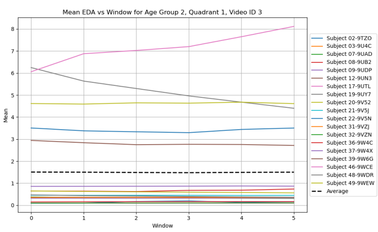
(a) EDA in below 30 years



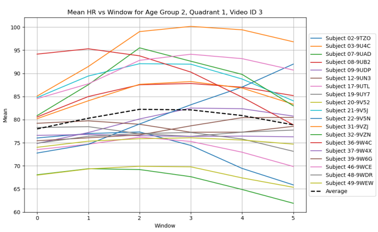
(b) HR in below 30 years



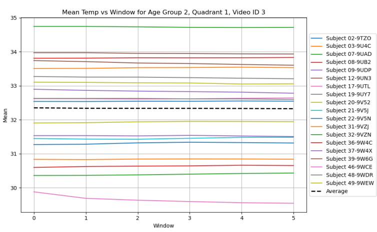
(c) TEMP in below 30 years



(d) EDA in older 30 years

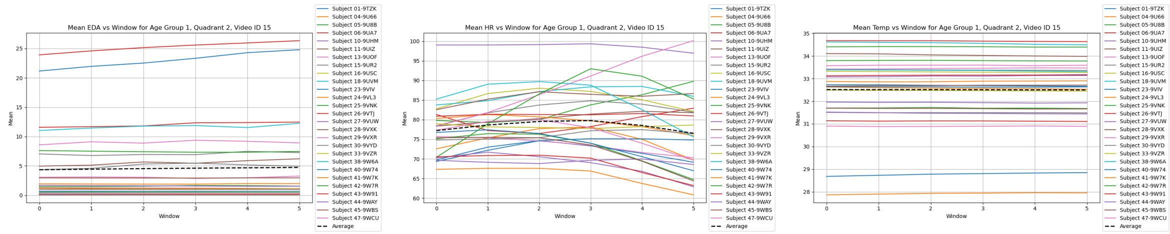


(e) HR in older 30 years

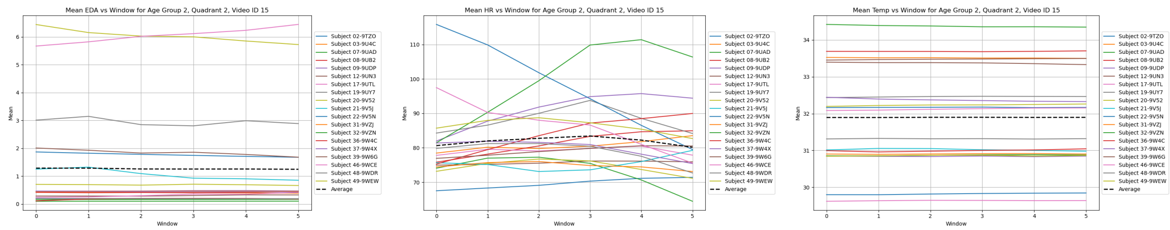


(f) TEMP in older 30 years

Figure 20: Results of the happy emotion by age group in the EmoWear dataset.

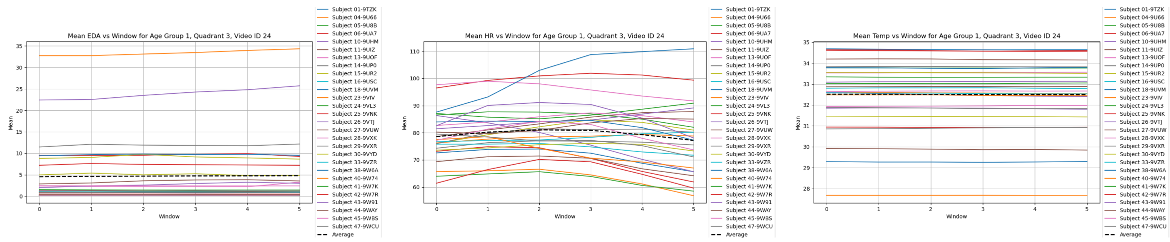


(a) EDA in below 30 years (b) HR in below 30 years (c) TEMP in below 30 years

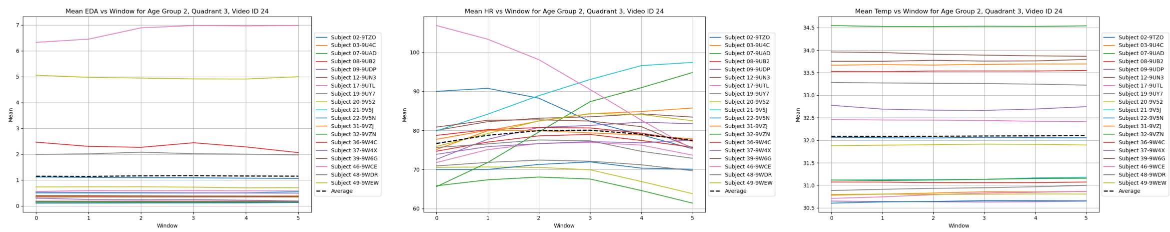


(d) EDA in older 30 years (e) HR in older 30 years (f) TEMP in older 30 years

Figure 21: Results of the relaxed emotion by age group in the EmoWear dataset.



(a) EDA in below 30 years (b) HR in below 30 years (c) TEMP in below 30 years

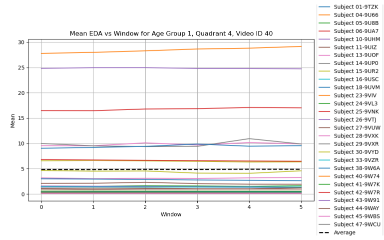


(d) EDA in older 30 years (e) HR in older 30 years (f) TEMP in older 30 years

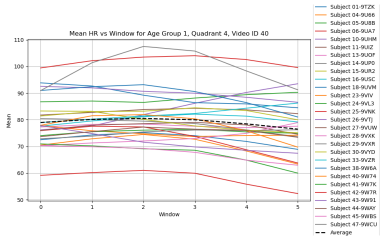
Figure 22: Results of the sad emotion by age group in the EmoWear dataset.

4.2.3 Quantitative Analysis

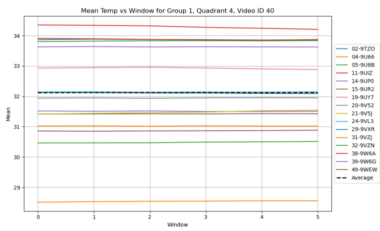
Tables 7 to 9 describe the quantitative results of the described indicators for the EDA, HR, and TEMP signals, respectively. Numerical results are aligned with the aforementioned graphical observations. TEMP-related indicators cannot distinguish among emotional states in the randomised samples, suggesting that this signal only cannot be used to classify emotions. Similarly, a non-homogeneous behaviour is appreciated when analysing EDA-related indicators. Notwithstanding, some indicators, such



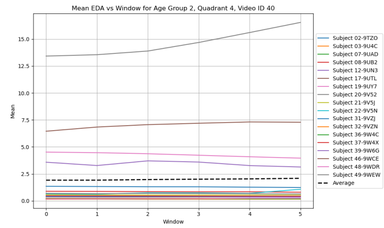
(a) EDA in below 30 years



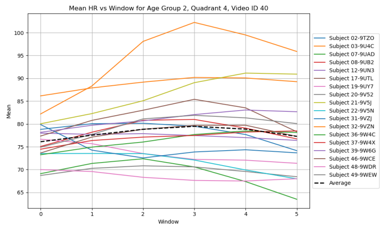
(b) HR in below 30 years



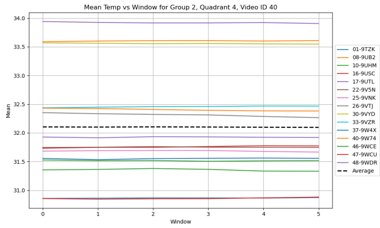
(c) TEMP in below 30 years



(d) EDA in older 30 years

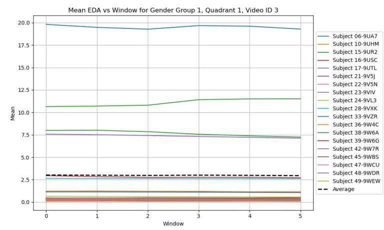


(e) HR in older 30 years

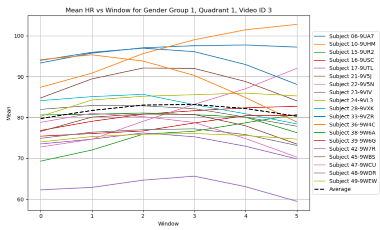


(f) TEMP in older 30 years

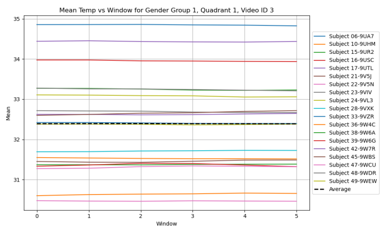
Figure 23: Results of the angry emotion by age group in the EmoWear dataset.



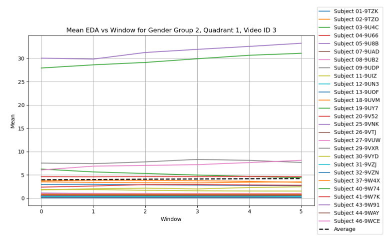
(a) EDA in females



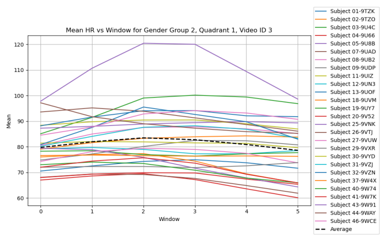
(b) HR in females



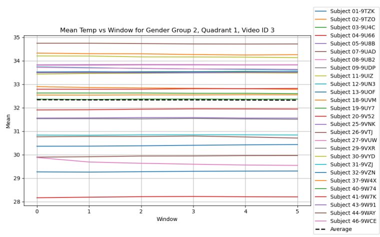
(c) TEMP in females



(d) EDA in males



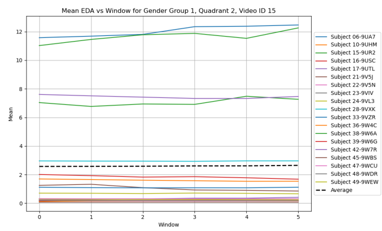
(e) HR in males



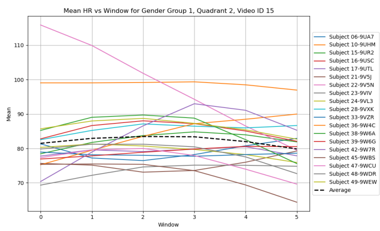
(f) TEMP in males

Figure 24: Results of the happy emotion by gender in the EmoWear dataset.

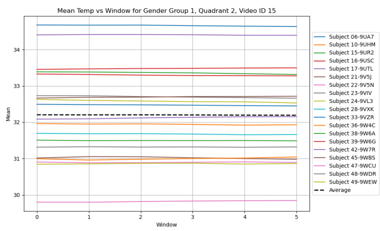
as WMM for sadness, present similar results in the randomised sample. However, they might not be enough to differentiate them from other emotions. Moreover, HR-related indicators present some minor patterns. For instance, WVI results of the relaxed state are generally similar and they are quite distant in comparison to the other WVI results of the other emotional states. Likewise, WMM results of the angry state are relatively similar among the different random groups and distant from the other emotions.



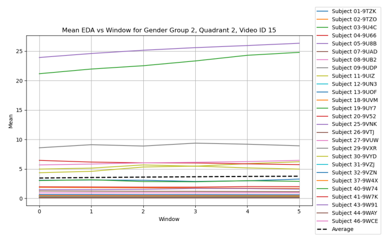
(a) EDA in females



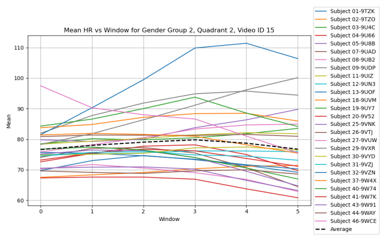
(b) HR in females



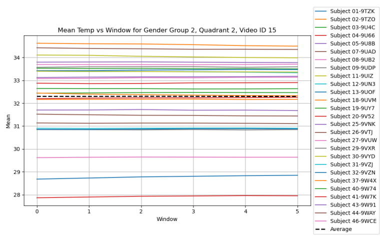
(c) TEMP in females



(d) EDA in males

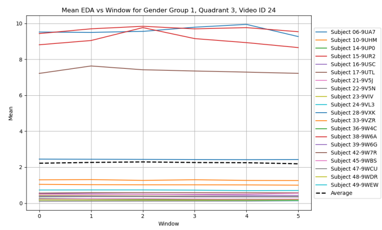


(e) HR in males

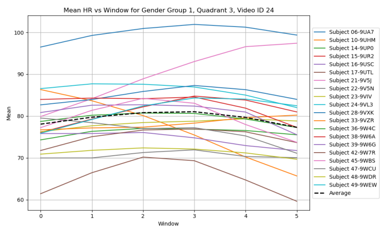


(f) TEMP in males

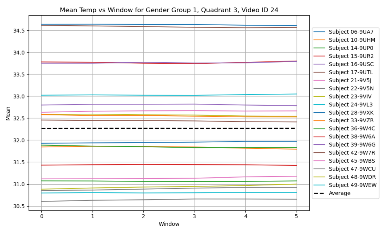
Figure 25: Results of the relaxed emotion by gender in the EmoWear dataset.



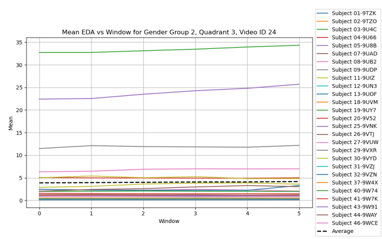
(a) EDA in females



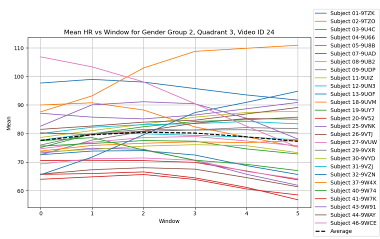
(b) HR in females



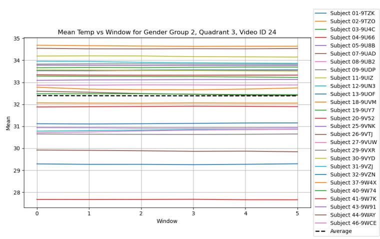
(c) TEMP in females



(d) EDA in males



(e) HR in males



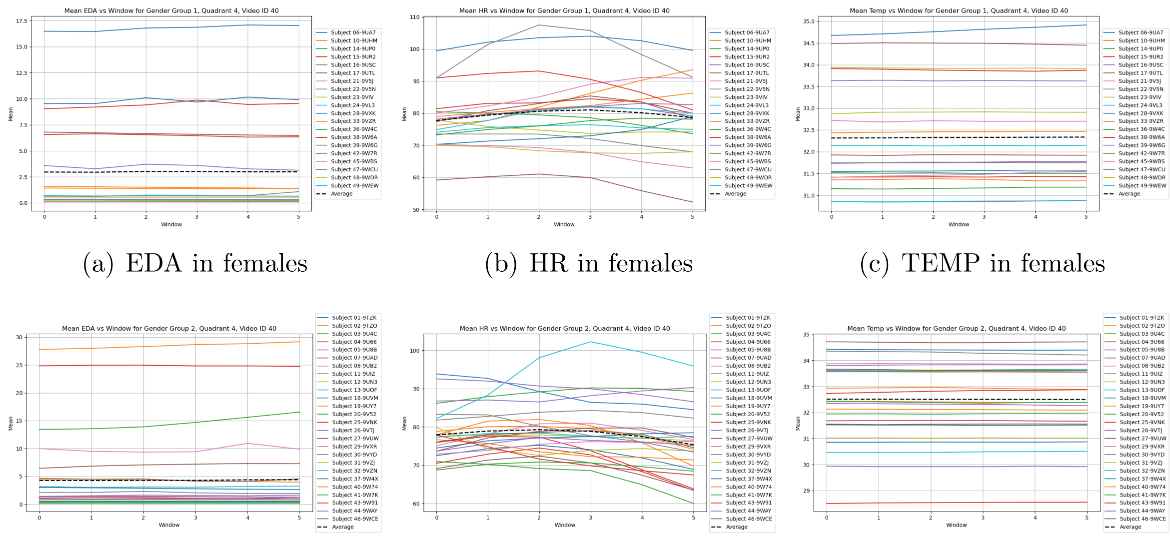
(f) TEMP in males

Figure 26: Results of the sad emotion by gender in the EmoWear dataset.

4.2.4 Summary of Findings

The physiological signals across the different emotional states reveal several key trends in both the randomised and segmented samples.

In the randomised samples, it is observed that TEMP does not vary and remains constant across all emotions, making it difficult to distinguish between different emotional states using this signal alone. The general trend of HR also presents challenges in distinguishing between the four emotions easily. However, individual (non-aggregated) behaviours reveal different trends, although these variations are more individual than



(a) EDA in females (b) HR in females (c) TEMP in females

(d) EDA in males (e) HR in males (f) TEMP in males

Figure 27: Results of the angry emotion by gender in the EmoWear dataset.

Table 7: Indicators for the EDA signals in the EmoWear dataset.

		Randomised Sample			**Segmented Sample**			
	Index	**Group 1**	**Group 2**	**Group 3**	**Females**	**Males**	**< 30**	**≥ 30**
Happy	**WVI**	5.081	4.324	5.009	3.657	10.757	11.499	2.915
	WMM	0.165	0.404	0.136	0.067	0.249	0.206	0.028
	AMM	1.054	1.141	1.030	1.023	1.063	1.041	1.018
	SCD	0.211	0.404	0.156	0.129	0.249	0.206	0.050
Relaxed	**WVI**	9.713	3.949	4.187	6.425	11.425	14.491	3.358
	WMM	0.085	0.323	0.219	0.076	0.303	0.386	0.041
	AMM	1.025	1.126	1.064	1.029	1.088	1.088	1.033
	SCD	0.085	0.323	0.219	0.076	0.303	0.386	0.043
Sad	**WVI**	7.243	6.842	6.085	6.350	13.846	17.343	2.852
	WMM	0.164	0.165	0.223	0.105	0.288	0.253	0.028
	AMM	1.055	1.045	1.071	1.048	1.074	1.055	1.025
	SCD	0.249	0.169	0.223	0.178	0.288	0.253	0.055
Angry	**WVI**	7.883	3.048	4.221	5.754	9.428	9.910	5.272
	WMM	0.184	0.134	0.059	0.072	0.152	0.069	0.173
	AMM	1.051	1.042	1.013	1.024	1.035	1.014	1.090
	SCD	0.222	0.145	0.189	0.120	0.153	0.156	0.173

group-based. Similarly, EDA does not vary and remains constant, making it challenging to differentiate between emotions based on this signal. A summary of findings is provided in Table 10.

For the segmented samples, there are no apparent distinguishable differences by comparing the signals either by age group or by gender.

Given the analysis, this dataset appears to be less effective at distinguishing emotions for several reasons. The dataset includes data for only one minute, which might not be sufficient for the signals to evolve meaningfully. In comparison, the WESAD

Table 8: Indicators for the HR signals in the EmoWear dataset.

		Randomised Sample			Segmented Sample			
	Index	**Group 1**	**Group 2**	**Group 3**	**Females**	**Males**	**< 30**	**≥ 30**
Happy	**WVI**	53.294	56.985	76.014	73.016	113.277	107.320	78.973
	WMM	3.526	2.956	5.729	3.436	4.871	4.432	4.236
	AMM	1.045	1.037	1.071	1.043	1.061	1.055	1.054
	SCD	6.746	5.845	9.848	6.306	8.596	7.494	7.584
Relaxed	**WVI**	68.011	62.749	61.991	84.714	108.038	101.776	90.975
	WMM	5.418	6.415	3.576	3.606	3.187	3.241	3.279
	AMM	1.069	1.083	1.047	1.045	1.041	1.042	1.040
	SCD	7.415	6.692	6.325	5.594	6.254	5.799	6.080
Sad	**WVI**	62.230	55.033	68.118	74.786	112.427	110.825	76.388
	WMM	4.008	5.655	5.180	3.722	3.178	3.884	3.390
	AMM	1.051	1.078	1.066	1.048	1.041	1.050	1.044
	SCD	6.286	6.175	8.756	6.687	6.023	6.367	5.986
Angry	**WVI**	50.162	64.544	44.151	73.850	87.561	103.944	57.467
	WMM	3.102	3.851	3.199	3.292	3.920	4.020	3.377
	AMM	1.039	1.049	1.043	1.042	1.052	1.052	1.044
	SCD	5.692	7.109	3.714	5.637	5.254	5.552	5.569

Table 9: Indicators for the TEMP signals in the EmoWear dataset.

		Randomised Sample			Segmented Sample			
	Index	**Group 1**	**Group 2**	**Group 3**	**Females**	**Males**	**< 30**	**≥ 30**
Happy	**WVI**	1.252	1.501	1.377	1.657	2.473	2.365	1.765
	WMM	0.002	0.032	0.004	0.004	0.019	0.006	0.026
	AMM	1.000	1.001	1.000	1.000	1.000	1.000	1.000
	SCD	0.004	0.036	0.006	0.007	0.019	0.011	0.026
Relaxed	**WVI**	1.259	1.335	1.402	1.643	2.354	2.305	1.691
	WMM	0.014	0.005	0.015	0.012	0.007	0.015	0.005
	AMM	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	SCD	0.024	0.012	0.018	0.014	0.013	0.019	0.010
Sad	**WVI**	1.248	1.444	1.379	1.756	2.418	2.467	1.706
	WMM	0.009	0.008	0.011	0.005	0.012	0.018	0.020
	AMM	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	SCD	0.009	0.013	0.012	0.009	0.014	0.018	0.022
Angry	**WVI**	1.246	1.355	1.346	1.702	2.322	2.490	1.534
	WMM	0.005	0.008	0.015	0.020	0.011	0.003	0.005
	AMM	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	SCD	0.011	0.014	0.016	0.020	0.011	0.006	0.009

dataset provides several minutes of data, allowing for a more comprehensive analysis of emotional states along the time.

While patterns could potentially be detected using the HR signal, these patterns appear to be more individual-specific rather than consistent across the group. This suggests that emotion detection might require a more personalised approach, necessitating the collection of more data from individual users and the development of custom models tailored to each person. However, this approach is much more complex and costly, making it less practical for broader applications.

Table 10: Summary of findings using a randomised sample strategy in the EmoWear dataset.

	TEMP	**HR**	**EDA**
Happy	Constant over time	Gradual increase to midpoint, then decrease	Largely constant, slight individual variations
Relaxed	Constant over time	Varied behaviours, some increase, decrease, or peak at midpoint	Constant over time
Sad	Constant over time	Varied behaviours, some increase, decrease, or peak at midpoint	Constant over time
Angry	Constant over time	Gradual increase to midpoint, then decrease	Constant over time
General observations	TEMP remains constant, not useful for differentiation	HR trends vary, individual behaviours reveal differences	EDA remains constant, not useful for differentiation

5 Conclusions

Emotions are complex physiological responses to stimuli triggered by our experiences and interactions with our environment. Understanding and managing emotions is crucial in the educational domain for fostering positive learning outcomes and creating pleasant workplaces. With the growing popularity of smart classrooms, large amounts of data are expected to be collected through wearable devices from both students and teachers. In this work, we have investigated the feasibility of using physiological signals to recognise emotions through AI techniques. To this end, we have identified the most relevant frameworks related to emotion theory from a psychological perspective, as well as the most relevant physiological signals for emotion recognition. A novel methodology has been proposed to characterise these signals and extract features that enable their differentiation. Our work provides both quantitative and graphical results for the EDA, HR, and TEMP signals, obtained from two real-world, open-source datasets.

5.1 Summary of Findings

The main concluding remarks of our work include the following key insights and observations:

1. **Contextualisation of physiological signals**: Physiological signals alone are not enough to fully capture emotions. While they provide valuable insights into the body's internal state, they often lack the context needed to accurately interpret emotional states. It is necessary to complement physiological signals with additional information, such as facial expressions, which can provide subtle changes in mood, body postures, which can reflect overall engagement, and contextual information like environmental conditions (*e.g.,* noise, air temperature, CO_2 concentration...). These additional layers of data help develop a more holistic representation of emotions.

2. **Heterogeneity on the evolution of the physiological signals' patterns**: No clear patterns have been identified that allow emotions to be distinguished numerically or graphically. This suggests that the data used are insufficient or the datasets may lack the diversity needed to capture the complexity of emotional states. Moreover, EDA and HR signals have proven to be more relevant than TEMP signals, particularly in the case of the EmoWear dataset, where only one minute of data is available for each signal. The short duration of these signals may not be sufficient to capture the induced emotions, as it does not allow enough time for emotional states to stabilise or for individual differences to manifest. To better capture these emotional variations, it is preferable to have, in addition to have a large number of participants, longer recordings that provide a more comprehensive view of the evolution of physiological signals over time.

3. **Inability to develop physiological signals-based machine learning models for emotion recognition**: The features and patterns extracted from the physiological signals unable the development of robust machine learning models as there are no discernible patterns among emotions. As a result, the models struggle to generalise or make accurate predictions.

4. **Lack of datasets for emotion recognition in children**: The datasets available for emotion recognition research relate to adults, leaving a significant gap in data for children. This limitation hinders the development of emotion recognition systems tailored for the population of smart classrooms, as these systems would be based on patterns identified in adults rather than children. As children may have different physiological and behavioural responses to emotions compared to adults, relying solely on adult data may lead to inaccurate interpretations of children's emotional states. Therefore, there is a clear need to create dedicated datasets that capture physiological signals associated with emotional elicitation in children.

5.2 Future Work

Future research and development efforts should focus on further enhancing the characterisation and analysis of the physiological signals, while also considering extended experiments to validate the suitability of the proposed methodology. Potential areas for improvement include:

1. **Enhancement of the datasets' quality**: Expanding the size and diversity of datasets, particularly by including samples from children in education contexts, is crucial. Although this may be expensive, conducting studies in real classrooms would provide more relevant and applicable data for the educational domain.

2. **Advanced signal analyses**: It is recommended to analyse the signals in the frequency domain, in addition to only evaluating their absolute values in the time domain. This approach could uncover more significant patterns that are not visible in time-domain analyses. Additionally, conducting studies with varying window sizes could reveal temporal dynamics that are critical for understanding emotional processes.

3. **Incorporation of contextual variables**: Considering additional contextual variables, such as CO_2 levels, air temperature, and noise, can be crucial. Since data acquisition settings can vary between datasets, these environmental factors may significantly influence emotional states and their physiological expressions. Incorporating these variables can enhance the understanding of the data and lead to more accurate emotion recognition.

4. **Application of machine learning techniques**: Applying machine learning techniques to emotion classification might be of interest when combining physiological signals with additional contextual variables. These methods could capture complex, non-linear relationships among these data and emotional states, offering a deeper understanding and more effective applications.

5. **Development of educational tools**: The analysis and interpretation of physiological signals are essential for the development of emotion-oriented tools to be integrated into smart classrooms. These tools have the potential to improve and adapt teaching methodologies, making classes more effective and responsive to students' emotions in real-time.

5.3 Research Publications

The following research publication has resulted from the study carried out in this work:

- Edgar Batista, Laia Cot, Valeria Pérez, and Antoni Martínez-Ballesté, "Real-Time Emotion Assessment System in Smart Classrooms using Wearable Bracelets", in the 18th International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies (UBICOMM), Venice, Italy, 2024 (to appear). *[Conference ranking: CORE C]*

Till the end of time.

References

[1] Jing Tan, Jie Mao, Yizhang Jiang, and Ming Gao. The Influence of Academic Emotions on Learning Effects: A Systematic Review. *International Journal of Environmental Research and Public Health*, 18(18):9678, 2021.

[2] Liujun Li, Chao Ping Chen, Lijun Wang, Kai Liang, and Weiyue Bao. Exploring Artificial Intelligence in Smart Education: Real-Time Classroom Behavior Analysis with Embedded Devices. *Sustainability*, 15(10):7940, 2023.

[3] Bohdan Myroniv, Cheng-Wei Wu, Yi Ren, Albert Christian, Ensa Bajo, and Yu-Chee Tseng. Analyzing User Emotions via Physiology Signals. *Data Science and Pattern Recognition*, 1(2):11–25, 2017.

[4] Edgar Batista, M. Angels Moncusi, Pablo López-Aguilar, Antoni Martínez-Ballesté, and Agusti Solanas. Sensors for Context-Aware Smart Healthcare: A Security Perspective. *Sensors*, 21(20):6886, 2021.

[5] Rob Koper. Conditions for effective smart learning environments. *Smart Learning Environments*, 1(1):1–17, 2014.

[6] Antoni Martínez-Ballesté, Edgar Batista, Elena Figueroa, Gabriela Fretes Torruella, Cèlia Llurba, José Quiles-Rodríguez, Oihane Unciti, and Ramon Palau. A Proposal for the Smart Classroom Infrastructure using IoT and Artificial Intelligence. In *Proceedings of the 48th Annual Computers, Software, and Applications Conference (COMPSAC)*, pages 109–114, Osaka, Japan, 2024. IEEE.

[7] Elena Figueroa, Edgar Batista, Ramon Palau, Oihane Unciti, Maria Ferre, and Antoni Martínez-Ballesté. The Use of Artificial Intelligence Techniques in Smart Classrooms is in its Infancy. *IEEE Access*, 2024. (In press).

[8] Lin Shu, Jinyan Xie, Mingyue Yang, Ziyi Li, Zhenqi Li, Dan Liao, Xiangmin Xu, and Xinyi Yang. A Review of Emotion Recognition Using Physiological Signals. *Sensors*, 18(7):2074, 2018.

[9] Paul Ekman, Robert W. Levenson, and Wallace V. Friesen. Autonomic Nervous System Activity Distinguishes Among Emotions. *Science*, 221(4616):1208–1210, 1983.

[10] James A. Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161–1178, 1980.

[11] Paul Ekman, Wallace V. Friesen, Maureen O'Sullivan, Anthony Chan, Irene Diacoyanni-Tarlatzis, Karl Heider, Rainer Krause, William Ayhan LeCompte, Tom Pitcairn, Pio E. Ricci-Bitti, Klaus Scherer, Masatoshi Tomita, and Athanase Tzavaras. Universals and Cultural Differences in the Judgments of Facial Expressions of Emotion. *Journal of Personality and Social Psychology*, 53(4):712–717, 1987.

[12] Paul Ekman and Wallace V. Friesen. Facial Action Coding System. *Environmental Psychology & Nonverbal Behavior*, 1978.

[13] James A. Russell, Maria Lewicka, and Toomas Niit. A Cross-Cultural Study of a Circumplex Model of Affect. *Journal of Personality and Social Psychology*, 57(5): 848–856, 1989.

[14] Jonathan Posner, James A. Russell, and Bradley S. Peterson. The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology*, 17(3):715—734, 2005.

[15] Niranjana Krupa, Karthik Anantharam, Manoj Sanker, Sameer Datta, and John Vijay Sagar. Recognition of emotions in autistic children using physiological signals. *Health and Technology*, 6:137–147, 2016.

[16] Stanisław Saganowski, Bartosz Perz, Adam G. Polak, and Przemysław Kazienko. Emotion Recognition for Everyday Life Using Physiological Signals From Wearables: A Systematic Literature Review. *IEEE Transactions on Affective Computing*, 14(3):1876–1897, 2022.

[17] Sarah Sarabadani, Larissa C. Schudlo, Ali Akbar Samadani, and Azadeh Kushski. Physiological Detection of Affective States in Children with Autism Spectrum Disorder. *IEEE Transactions on Affective Computing*, 11(4):588–600, 2018.

[18] Wagno Leão Sergio, Victor Ströele, Mário Dantas, Regina Braga, and Douglas Dyllon Macedo. Enhancing well-being in modern education: A comprehensive eHealth proposal for managing stress and anxiety based on machine learning. *Internet of Things*, 25:101055, 2024.

[19] Nathalie Michels, Isabelle Sioen, Els Clays, Marc De Buyzere, Wolfgang Ahrens, Inge Huybrechts, Barbara Vanaelst, and Stefaan De Henauw. Children's heart rate variability as stress indicator: Association with reported stress and cortisol. *Biological Psychology*, 94(2):433–440, 2013.

[20] Cleveland Clinic. Heart Rate. https://my.clevelandclinic.org/health/diagnostics/heart-rate, 2024. [Accessed 16-08-2024].

[21] Elisa Spanò, Stefano Di Pascoli, and Giuseppe Iannaccone. Low-PowerWearable ECG Monitoring System for Multiple-Patient Remote Monitoring. *IEEE Sensors Journal*, 16(13):5452–5462, 2016.

[22] Andriy Temko. Accurate Heart Rate Monitoring During Physical Exercises Using PPG. *IEEE Transactions on Biomedical Engineering*, 64(9):2016–2024, 2017.

[23] Yi Su, Chunsheng Ma, Jing Chen, Huiping Wu, Weixiang Luo, Yueming Peng, Zebang Luo, Lin Li, Yongsong Tan, Olatunji Mumini Omisore, Zhengfang Zhu, Lei Wang, and Hui Li. Printable, Highly Sensitive Flexible Temperature Sensors for Human Body Temperature Monitoring: A Review. *Nanoscale Research Letters*, 15:200, 2020.

[24] Min Woo Park, Chi Jung Kim, Mincheol Hwang, and Eui Chul Lee. Individual Emotion Classification between Happiness and Sadness by Analyzing Photoplethysmography and Skin Temperature. In *Proceedings of the 4th World Congress on Software Engineering*, pages 190–194. IEEE, 2013.

[25] Rie Nakanishi and Kyoko Imai-Matsumura. Facial skin temperature decreases in infants with joyful expression. *Infant Behavior and Development*, 31(1):137–144, 2008.

[26] Li-Yuan Chiang, Chun-Ta Wang, Ting-Syuan Lin, Steve Pappert, and Paul Yu. Highly sensitive silicon photonic temperature sensor based on liquid crystal filled slot waveguide directional coupler. *Optics Express*, 28(20):29345–29356, 2020.

[27] Arturo Martínez-Rodrigo, Antonio Fernández-Caballero, Fábio Silva, and Paulo Novais. Monitoring Electrodermal Activity for Stress Recognition Using a Wearable. In *Proceedings of the 12th International Conference on Intelligent Environments*, pages 416–425, London, United Kingdom, 2016. IOS Press.

[28] Roberto Zangróniz, Arturo Martínez-Rodrigo, José Manuel Pastor, María T López, and Antonio Fernández-Caballero. Electrodermal Activity Sensor for Classification of Calm/Distress Condition. *Sensors*, 17(10):2324, 2017.

[29] Gunnar C. Pope and Ryan J. Halter. Design and Implementation of an Ultra-Low Resource Electrodermal Activity Sensor for Wearable Applications. *Sensors*, 19 (11):2450, 2019.

[30] Philip Schmidt, Attila Reiss, Robert Dürichen, Claus Marberger, and Kristof Van Laerhoven. Introducing WESAD, a Multimodal Dataset for Wearable Stress and Affect Detection. In *Proceedings of the 20th ACM International Conference on Multimodal Interaction*, pages 400–408, Boulder, USA, 2018.

[31] Saskia Koldijk, Maya Sappelli, Suzan Verberne, Mark A. Neerincx, and Wessel Kraaij. The SWELL Knowledge Work Dataset for Stress and User Modeling Research. In *Proceedings of the 16th International Conference on Multimodal Interaction*, pages 291–298, Istanbul, Turkey, 2014.

[32] Valentina Markova, Todor Ganchev, and Kalin Kalinkov. CLAS: A Database for Cognitive Load, Affect and Stress Recognition. In *Proceedings of the International Conference on Biomedical Innovations and Applications*, pages 1–4, Varna, Bulgaria, 2019.

[33] Karan Sharma, Claudio Castellini, Egon L. van den Broek, Alin Albu-Schäeffer, and Friedhelm Schwenker. A dataset of continuous affect annotations and physiological signals for emotion analysis. *Scientific Data*, 6(1):196, 2019.

[34] Mohammad Hasan Rahmani, Michelle Symons, Omid Sobhani, Raf Berkvens, and Maarten Weyn. EmoWear: Wearable Physiological and Motion Dataset for Emotion Recognition and Context Awareness. *Scientific Data*, 11:648, 2024.

[35] Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. DEAP: A Database for Emotion Analysis Using Physiological Signals. *IEEE Transactions on Affective Computing*, 3(1):18–31, 2012.

[36] Juan Abdon Miranda-Correa, Mojtaba Khomami Abadi, Nicu Sebe, and Ioannis Patras. AMIGOS: A Dataset for Affect, Personality and Mood Research on Individuals and Groups. *IEEE Transactions on Affective Computing*, 12(2):479–493, 2021.

[37] Mohammad Soleymani, Jeroen Lichtenauer, Thierry Pun, and Maja Pantic. A Multimodal Database for Affect Recognition and Implicit Tagging. *IEEE Transactions on Affective Computing*, 3(1):42–55, 2012.

[38] Ingrid M. Cordon, Annika M. D. Melinder, Gail S. Goodman, and Robin S. Edelstein. Children's and adults' memory for emotional pictures: Examining age-related patterns using the Developmental Affective Photo System. *Journal of Experimental Child Psychology*, 114(2):339–356, 2013.

[39] Laura Redondo-Flórez, José Francisco Tornero-Aguilera, Domingo Jesús Ramos-Campo, and Vicente Javier Clemente-Suárez. Gender Differences in Stress-and Burnout-Related Factors of University Professors. *BioMed Research International*, 2020(1):6687358, 2020.

[40] Lili Zhu, Petros Spachos, Pai Chet Ng, Yuanhao Yu, Yang Wang, Konstantinos Plataniotis, and Dimitrios Hatzinakos. Stress Detection Through Wrist-Based Electrodermal Activity Monitoring and Machine Learning. *IEEE Journal of Biomedical and Health Informatics*, 27(5):2155–2165, 2023.

[41] Empatica Inc. Empatica E4. `https://www.empatica.com/research/e4/`, 2024. [Accessed 22-08-2024].