



Article

Exploring Human Emotions: A Virtual Reality-Based Experimental Approach Integrating Physiological and Facial Analysis

Leire Bastida ^{1,*}, Sara Sillaurren ², Erlantz Loizaga ¹, Eneko Tomé ¹ and Ana Moya ¹

¹ TECNALIA Research and Basque Research and Technology Alliance (BRTA), 48160 Derio, Spain; erlantz.loizaga@tecnalia.com (E.L.); eneko.tome@tecnalia.com (E.T.); ana.moya@tecnalia.com (A.M.)

² TECNALIA, Basque Research and Technology Alliance (BRTA), 01510 Vitoria, Spain; sara.sillaurren@tecnalia.com

* Correspondence: leire.bastida@tecnalia.com

Abstract: This paper researches the classification of human emotions in a virtual reality (VR) context by analysing psychophysiological signals and facial expressions. Key objectives include exploring emotion categorisation models, identifying critical human signals for assessing emotions, and evaluating the accuracy of these signals in VR environments. A systematic literature review was performed through peer-reviewed articles, forming the basis for our methodologies. The integration of various emotion classifiers employs a ‘late fusion’ technique due to varying accuracies among classifiers. Notably, facial expression analysis faces challenges from VR equipment occluding crucial facial regions like the eyes, which significantly impacts emotion recognition accuracy. A weighted averaging system prioritises the psychophysiological classifier over the facial recognition classifiers due to its higher accuracy. Findings suggest that while combined techniques are promising, they struggle with mixed emotional states as well as with fear and trust emotions. The research underscores the potential and limitations of current technologies, recommending enhanced algorithms for effective interpretation of complex emotional expressions in VR. The study provides a groundwork for future advancements, aiming to refine emotion recognition systems through systematic data collection and algorithm optimisation.

Keywords: human emotions; human factors; physiological signals; facial analysis; multimodal perception; emotion recognition



Citation: Bastida, L.; Sillaurren, S.; Loizaga, E.; Tomé, E.; Moya, A. Exploring Human Emotions: A Virtual Reality-Based Experimental Approach Integrating Physiological and Facial Analysis. *Multimodal Technol. Interact.* **2024**, *8*, 47. <https://doi.org/10.3390/mti8060047>

Academic Editor: Alexey Karpov

Received: 14 April 2024

Revised: 31 May 2024

Accepted: 1 June 2024

Published: 4 June 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Human emotions are responses to external stimuli (matters or situations) that bear significant personal relevance [1]. Emotions, defined as a complex pattern of experiential, behavioural, and physiological elements, are fundamental to human experiences [2]. Typically arising spontaneously, these mental states are linked to notable physical and physiological changes across various human systems, including the brain, heart, muscles, and facial expressions, among others [3]. Advances in sensor technology have enabled the measurement of psychophysiological signals, providing an avenue for recognising and categorising human emotions systematically [4].

The primary objective of this study is to explore and evaluate psychophysiological signals and facial expressions within a virtual reality (VR) environment to assess human emotions accurately [5,6]. To effectively address this overarching goal, the investigation also delves into complementary research areas. These include a review and analysis of prevailing models for categorising emotions to establish a comprehensive theoretical framework and an examination of crucial human signals and features that are instrumental in assessing emotions. The study aims to bridge existing knowledge gaps identified in prior research, such as the limitations observed in emotion recognition in immersive environments.

The paper is structured as follows: Section 1 introduces the concept of emotion recognition and its significance, highlighting its general framework and outlining the major research questions this paper intends to explore. Section 2 delves into the theoretical background, underscoring prior research and elucidating the relevance of emotion detection via physiological signals and facial analysis. Section 3 details the experimental environment and elements. Section 4 outlines the experimental results. In Section 5, a comprehensive discussion is included, thoroughly exploring the implications of the research findings for emotion detection. Lastly, Section 6 offers the conclusion and provides answers to the research questions.

2. Background

At the beginning of our research, a systematic literature review (SLR) was conducted to gather and analyse relevant studies focused on the current state of the art in emotion, psychophysiological responses, and facial emotion recognition. This systematic approach ensured a comprehensive examination of the latest research in these areas. Our literature search spanned multiple reputable databases: IEEE, Scopus, Springer Link, Semantic Scholar, ResearchGate, ScienceDirect, and MDPI Journals, with further cross-referencing through citations and references via Google Scholar. The detailed process of our systematic literature review, grounded in the PRISMA model, is illustrated in Figure 1.

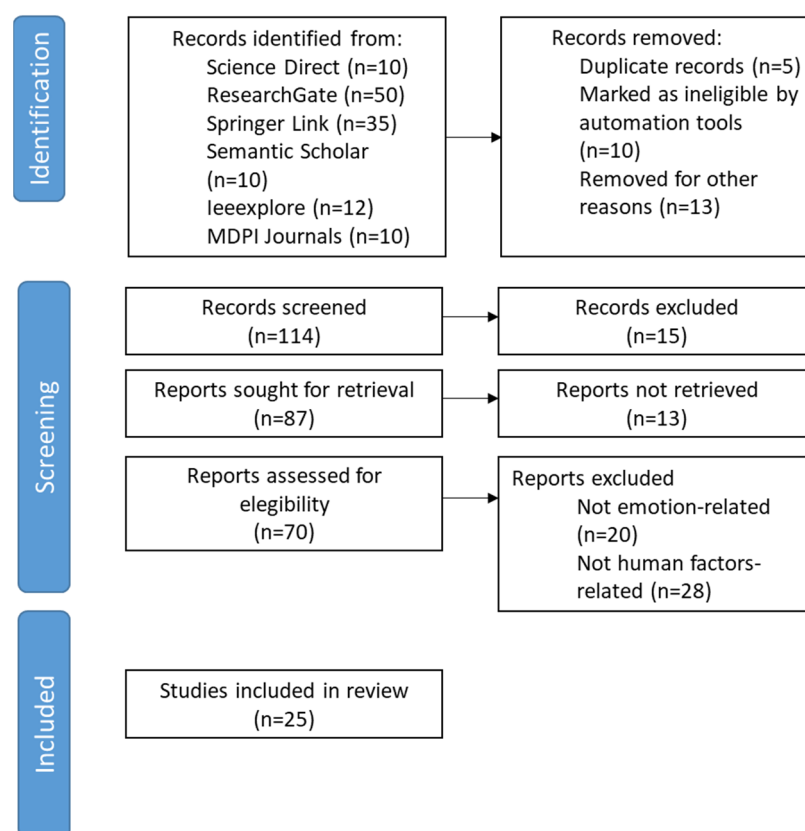


Figure 1. Report of the SLR derived from PRISMA model.

Initially, non-peer-reviewed content, such as editorials or call-for-papers announcements, was excluded to maintain a focus solely on peer-reviewed journals and conference articles. Subsequently, articles were further filtered based on their abstracts to ensure relevance to our research themes. The search criteria and methods based on specific keywords yielded 127 articles published between 2014 and 2022. After applying our inclusion and exclusion criteria, 70 studies were selected for deeper review. This was further refined, resulting in 25 articles that were directly relevant to our study.

The references and methodologies detailed in the following subsections are drawn from these carefully selected articles, forming the foundation of our analysis and discussion in this paper.

2.1. Human Emotion Models

Human emotions, arising from interactions with environmental stimuli, significantly influence motivation, behaviour, cognition, decision-making, and interpersonal relationships. The capacities for emotional response have both physiological and psychological components that are crucial for overall well-being; positive emotions enhance health and productivity, while prolonged negative emotions can impair them. Unlike mood states, which are conscious and sustained emotional responses, emotions are typically spontaneous and are accompanied by physiological changes that affect organs and tissues such as the brain, heart, and skin.

To enhance accuracy in emotion recognition and promote a reliable system design [4], a profound comprehension of emotion modelling and processing is essential. Yet, the precise categorisation of basic emotions remains a contentious topic among psychologists. They often adopt two main approaches [7–10]: the discrete or categorical emotion model, which segregates emotions into distinct categories, and the multidimensional model, which is based on affective dimensions for emotion labelling.

Under the discrete model, basic emotions such as anger, joy, sadness, disgust, fear, and surprise—categorised by Ekman [11]—are considered foundational to all human emotional responses. Plutchik's Wheel of Emotions [12] expands on this, adding anticipation and trust to the basic set, depicting emotions with varying intensities. Izard [2] further identifies ten basic emotions, noting that these evolved simple neural circuits are devoid of complex cognitive components.

The multidimensional models, such as Russell's circumplex [13] and Mehrabian's [14] PAD (pleasure, arousal, dominance) model [15], describe emotions on continuums of valence, arousal, and, occasionally, dominance. These models allow for nuanced identification and analysis of emotions based on their dimensional characteristics, though they may struggle to differentiate closely related emotional states.

In studying and understanding human emotions, the selection and application of appropriate emotion models are critical. Each model brings its own distinct theoretical underpinnings and practical implications. By conducting a comparative analysis of these models, researchers and practitioners can gain richer insights into the complexities of emotional expressions, significantly advancing the development of emotion recognition systems.

Discrete/categorical models are straightforward and user-friendly, categorising emotions into distinct groups. This simplifies the process of identifying and labelling emotions. A major advantage is their ability to provide clear, easy-to-communicate emotion labels which are intuitive for both researchers and laypeople. However, they struggle with the analysis of complex emotions that do not fit neatly into predefined categories, such as mixed emotions or subtler emotional nuances, making them less flexible in handling the complexity of human emotional states. Another limitation is that discrete models typically cover only a limited range of universally acknowledged basic emotions, which may not encompass the full spectrum of human emotions, leading to potential oversimplifications.

Multidimensional models offer a more nuanced view of emotions by plotting them on continuous scales, such as valence and arousal. This approach allows for a richer description of emotional states, capturing variations in intensity and subtlety that categorical models might miss. Multidimensional models are particularly valuable for examining the interrelationships between emotions and for exploring emotions that are not easily classified into discrete categories. The main drawback is their complexity: interpreting the multidimensional data requires more sophisticated analysis and might be less intuitive for users not trained in these methods.

In summary, understanding the strengths and limitations of each model can aid in selecting the appropriate model based on the specific needs of a study or application.

For instance, discrete models may be best suited for applications requiring quick and straightforward emotion recognition, such as user training or interface adaptation in real time. In contrast, multidimensional models might be more appropriate for therapeutic settings where a deeper understanding of emotional processes is needed. Moreover, the fusion of both models may offer a promising strategy to capitalise on the unique strengths of each model; however, the significant disparity in their underlying principles poses challenges to their seamless integration.

2.2. Human Signals Monitoring

Analysing emotional responses involves monitoring spontaneous psychophysiological processes via internal signals (such as brain activity, heart rate, and skin conductance) and external physical signals (like facial expressions, speech, and gestures). Internal signals offer a window into the workings of the central and autonomic nervous systems, leading to accurate emotion detection from a physiological standpoint [16,17]. Conversely, physical signals are more readily observed but can be less reliable, as they are easier for individuals to consciously control or disguise.

Identification and analysis of these signals hinge on sophisticated feature extraction techniques, which are critical for constructing predictive models of human emotions. A detailed exploration of signal types, measurement methodologies, and feature relevance is forthcoming in the following subsections.

2.2.1. Psychophysiological Signals

The assessment of human emotions can leverage various physiological metrics that signify specific affective states. Electrical activity in the brain, measured through EEG, scrutinises frequencies across delta, theta, alpha, beta, and gamma ranges to echo mental states from inattention to alertness [18]. Cardiovascular signals, captured via electrocardiogram (ECG) and photoplethysmography (PPG), analyse facets like heart rate variability and blood volume pulse to infer emotional arousals, reflecting shifts in heart dynamics and blood flow associated with emotional changes [19–21].

Electrodermal activity (EDA), or the measurement of skin conductance, reveals fluctuations in sweat glands' activity that correlate with emotional intensity; this method is especially sensitive to phasic changes reflecting immediate emotional reactions. In muscular assessments, electromyography (EMG) evaluates electrical activity produced by muscle contractions, offering cues about emotional states through facial expressions or body language. Similarly, respiratory patterns, detected through methods like capnography or thermal imaging of exhaled breath, change with emotional states. This manifests itself in variations of breathing rates and depths that can indicate excitement or stress.

Additionally, measuring eye movements through electro-oculography (EOG) complements emotion analysis by capturing corneo-retinal standing potential variations tied to eye position changes. These biometrics, together with sophisticated signal processing techniques, enable nuanced and complex interpretations of emotional states, supporting more accurate and effective emotional recognition [22].

2.2.2. Physical Signals

Human physical signals, such as facial expressions and stance, are key to detecting emotions, adding valuable dimensions to emotional recognition systems:

- Facial expressions are crucial in social interactions and closely linked to emotions [23]. The movements of facial muscles, supplied by the facial nerve, create expressions that are universally recognised indicators of emotions, regardless of a person's gender, age, cultural background, or social training. While humans are capable of creating thousands of different expressions, only a few specific facial configurations are universally associated with particular emotions [11].
- Human stance: Body language, including posture and gestures, plays a significant role in expressing emotions [24]. Observing how an individual positions their body

can reveal subtle emotional states. By integrating analysis of human stance with other data, researchers can enhance emotion detection systems, providing a more nuanced understanding of both overt and subtle emotional expressions.

2.3. Relevant Emotional Features

This section delves into the nuanced ways in which emotions influence and manifest through psychophysiological signals and facial expressions. The analysis begins by examining how emotions impact the autonomic nervous system (ANS) [17], presenting data through systematic literature review findings that map specific emotions to observable changes in various physiological responses, including electroencephalography, cardiovascular systems, electrodermal activity, and respiratory patterns.

Subsequently, the discussion shifts to facial expressions, a crucial component in conveying emotions within social interactions. Different methodologies employed to analyse facial expressions are explored, ranging from facial electromyography (fEMG) to advanced computer vision techniques. This exploration sheds light on the technological advancements in this field and addresses the challenges and complexities involved in distinguishing genuine emotional expressions from posed ones. Each point discussed sets the stage for a deeper understanding of the multifaceted nature of emotional expression, highlighting cutting-edge techniques and their implications for both psychological research and practical applications.

2.3.1. Psychophysiological Signals

An in-depth analysis based on SLR findings demonstrates how different emotions impact activities of the autonomic nervous system. These impacts occur across various psychophysiological signals and are quantified for each emotion, showing changes from a baseline state (increased, neutral, or decreased). Variation is defined as the response direction reported by the majority of studies (unweighted), with at least three studies indicating the same response direction. Icons indicate increased (+), decreased (−), or no change in activation from baseline (blank), or both increases and decreases between studies (\equiv). This section features mapping tables that correlate the seven most widely studied emotions with the features derived from psychophysiological response measurements: anger, disgust, fear, joy, sadness, anticipation, and surprise. Trust, as defined by Plutchik, is also featured in this section. These correlated tables systematically outline the associations between psychophysiological features and emotional states derived from electroencephalography (see Table 1), cardiovascular responses (see Table 2), electrodermal activity (see Table 3), and respiratory responses (see Table 4). When specific emotions lack associated psychophysiological data in the systematic literature review, these emotions do not appear in the summary tables.

Table 1. Relationship between emotions and features of electroencephalography responses [2–4,7,15,18,19,25,26].

Feature	Anger	Fear	Joy
Power spectral density of α wave	+	−	+
Power spectral density of β wave	−		
Power spectral density of γ wave		−	+
Differential entropy	+	−	+
Differential asymmetry	\equiv	+	−
Rational asymmetry	+	+	−

Table 2. Relationship between emotions and features of cardiovascular responses [2,4,7,15,18,19,25,26].

Feature	Anger	Disgust	Fear	Joy	Sadness	Anticipation	Surprise
Heart rate	+	≡	+	+	—	—	+
Heart rate variability (HRV)	—	+	—	—	—	+	
Low-frequency spectral HRV			≡	≡			
P-wave amplitude			+				
Pre-ejection period	—	—	—	+			
Stroke volume	≡		—	≡			
Cardiac output	≡	—	+	≡			
Systolic blood pressure	+	+	+	+			
Diastolic blood pressure	+	+	+	+ ≡			
Mean arterial pressure		+	+	+			
Total peripheral resistance	+	+	—	+ ≡			
Finger pulse amplitude	—	—	—	≡	—		
Finger pulse transit time	—	≡	—	+			
Ear pulse transit time		≡	—	+			
Finger temperature	—	≡	—	+	≡	+	≡

Table 3. Relationship between emotions and features of electrodermal responses [2,4,7,15,18,19,25,26].

Feature	Anger	Disgust	Fear	Joy	Sadness	Anticipation	Surprise
Skin conductance response	+	+	+	+	—	+	≡
Nonspecific skin conductance response rate	+	+	+	+	—		
Skin conductance level	+	+	+	≡ +	—	+	+

Table 4. Relationship between emotions and features of respiratory responses [2,4,7,15,18,19,25,26].

Feature	Anger	Disgust	Fear	Joy	Sadness	Anticipation	Surprise
Respiration rate	+	+	+	+	+	—	—
Inspiratory time	—	—	≡ —	—			
Expiratory time	—	+	—	—			
Post-inspiratory pause time	+		≡	—			
Inspiratory duty cycle		—	+				
Tidal volume	≡	—	≡ +	≡	+	—	
Inspiratory flow rate or inspiratory drive		—		+			

2.3.2. Facial Expressions

Facial expressions are intricately intertwined with human social interactions and are pivotal in conveying emotions [27–29]. Analysing these expressions helps deepen our understanding of social dynamics. Moreover, it is essential for distinguishing genuine emotional expressions from posed ones, which is key in evaluating the authenticity of emotional displays.

There are several methods for analysing facial expressions, each with distinct advantages and challenges:

- Facial electromyography (fEMG) tracks facial muscle activity using skin-mounted electrodes [30]. This method is excellent for capturing subtle muscle movements, meaning it can detect suppressed emotional expressions. However, it is invasive and sensitive to motion and electrical interferences, requiring specialised biosensor processing expertise, thereby adding to its complexity.
- Manual coding involves trained experts observing and coding facial movements into action units (AUs) based on anatomical features [31]. It is a non-intrusive and highly reliable method that offers considerable face validity. Nevertheless, it demands high-quality video equipment and significant expert involvement, making it costly and labour-intensive.

- Automatic facial expression analysis uses computer vision technologies to analyse expressions algorithmically [32]. This approach is easy to access and has been validated for standard expressions, but its accuracy can vary depending on the particular expressions and contexts being analysed.

The role of facial expressions in understanding social interactions has evolved, shifting the emphasis from merely identifying personal emotional states to analysing broader social dynamics. This shift underlines the importance of distinguishing spontaneous (genuine) from posed (deliberate) facial expressions, often referred to as the SVP (spontaneous versus posed) challenge. With advances in computer vision and machine learning, significant progress has been made in the automated detection of these differences. Nonetheless, early studies often relied on posed expressions. These are easier to collect as subjects are asked to mimic specific emotions.

Additionally, the increasing availability and capabilities of image databases have made them a valuable resource in facial expression research. These databases, which include both posed and spontaneous expressions induced by various stimuli, provide a rich dataset for analysis [33]. This method allows for a broader and more diversified examination of expressions but poses challenges in maintaining the authenticity and variability of the expressions, which have been documented.

Overall, while each method of data collection offers unique benefits suitable for different research applications, the increasing reliance on image databases is a notable trend. These databases significantly enhance our ability to study a wide range of naturalistic facial expressions, a development that is likely to refine our understanding of the complexities of human emotional and social interactions.

3. Materials and Methods

In this experiment, virtual reality serves as a platform to study and evoke a broad spectrum of human emotions, including joy, sadness, anger, fear, disgust, surprise, anticipation, and trust. This is achieved by immersing participants in various virtual scenarios designed to induce specific emotional responses. By manipulating elements within the virtual reality environment, the stimuli to which respondents are controlled facilitates the capture of both psychophysiological signals and facial expressions in a dynamic yet regulated setting.

The integration of VR technology in this experiment allows for real-time monitoring and recording of physiological responses, such as brain signals, heart rate, and galvanic skin response, along with advanced facial recognition techniques to capture expressions. This dual approach of combining psychophysiological data and facial expression analysis provides a holistic view of how emotions manifest physiologically and behaviourally within a virtual context. The goal is to enhance our understanding of the intricate relationship between emotional experiences, physiological reactions, and facial expressions, which in turn could advance emotion recognition technologies applicable in psychology, human–computer interaction, and immersive virtual reality experiences.

In this experiment, specific objectives include the following:

- Recruiting a substantial number of participants to ensure significant data and demographic representation;
- Collecting high-quality psychophysiological data to map changes as participants experience various emotion-triggering stimuli;
- Establishing a standardised data collection procedure to minimise data disturbances and the impact of uncontrolled variables.

3.1. Conceptual Design of the Experiment

This experiment engages participants in a series of virtual reality (VR) environments, each meticulously designed to systematically elicit one of eight target emotions: joy, trust, fear, surprise, sadness, anticipation, anger, and disgust. This suite of ten distinct VR scenarios, constructed from 360-degree videos and 3D games, is specifically tailored to

provoke consistent emotional responses while accommodating the broad spectrum of personal beliefs and perceptual biases participants may hold.

Participants navigate a variety of settings, ranging from calm (such as a peaceful beach or a lush forest crafted to induce relaxation and joy) to more dynamic and challenging situations (such as finding oneself in an underwater cage surrounded by marine creatures or in an emergency room intended to evoke feelings of fear, surprise, and anticipation). Notably, two of the scenarios are deliberately designed to reinforce the emotions of joy and fear, each presented twice within the sequence (at the beginning and at the end of the experience). This strategic duplication aims to activate a broad spectrum of emotional responses early in the experiment, setting a clear demarcation between two emotional extremes. This approach is vital for two reasons: first, to solidly anchor the two polar ends of the emotional range from the onset and, second, to mitigate potential biases that could arise from uncontrolled experimental variables, such as participants' initial focus on the technological aspects, like the quality of the VR images.

In selected scenarios, participants also have opportunities to interact with the environment using VR controllers. This interactivity not only engages participants more deeply but also enhances the immersion, potentially intensifying the emotional impact of the experience.

It is crucial to acknowledge the inherent variability in individual responses; the same scenario might evoke different emotions among participants or the same emotion at varying intensities. Such variations are influenced by each participant's personal beliefs, previous experiences, and individual interpretations of the world. To manage the emotional intensity and avoid prolonged exposure to high-stress emotions, the sequence in which these scenarios are presented is strategically orchestrated throughout the course of the experiment.

3.2. Experimental Sample

The experimental sample was carefully selected to align with the demographic requirements necessary for validating the experiment. For this study on emotional responses elicited via virtual reality scenarios, a total of 51 participants were initially involved, with 49 ultimately providing valid data; the remaining two were excluded due to issues such as poor data recording.

Since the focus of this experiment is on human emotions, which are universal, the target population is broad, encompassing all individuals in the active working age group. However, to effectively analyse the impact of demographic factors on emotional response, the participant pool was segmented based on gender. These criteria were selected as they are significant human traits that can influence the psychological and physiological responses under study. In total, data from 26 female and 23 male participants were gathered, all ranging in age from 25 to 60 years.

3.3. Equipment Used

Given the specific objectives of this experimental phase focusing on emotion recognition, it is vital to consider capturing as many human signals as feasible to determine which are most indicative of various emotional states. However, this approach introduces substantial challenges: Firstly, gathering a wide range of signals can complicate the experimental design and pose hurdles in the subsequent analysis due to the sheer volume and complexity of the data. Secondly, some signal acquisition methods might not integrate seamlessly with the settings or the methods used, potentially conflicting with the experimental framework designed for VR-based emotional studies.

After reviewing various academic sources and considering the practical applications of previous studies, the most relevant signals for effectively monitoring the emotional responses in our virtual reality setting were identified:

- Brain activity (electroencephalography EEG): A flexible headband equipped with a specialised 12-channel EEG headset that is compatible with the virtual reality equipment was used to monitor brain activity. This setup focuses primarily on capturing

signals from the frontal, central, and posterior regions of the brain, areas known to be involved in emotional processing.

- Galvanic skin response (GSR) and heart rate (HR): Electrodes were positioned on the index and ring fingers of the non-dominant hand to measure both galvanic skin response and heart rate. These metrics are particularly responsive to emotional changes; excitement or stress, for instance, can activate sweat glands, altering the skin's electrical resistance. The same device used to gather GSR data was simultaneously used to monitor heart rate variations, which often fluctuate in response to different emotional states. A low-voltage current passed between these points allows for detection of skin conductance changes as well as pulse rate, providing a comprehensive measure of physiological arousal associated with diverse emotions.
- Facial expressions: They were collected via camera recordings at high resolution (UHD-Ultra High Resolution 8K 7680 × 4320 and 24 fps) from a controlled distance of 1.5 m.

These selected methods are attuned to the needs of this experiment, focusing on capturing clear, interpretable signals that reflect the emotional states elicited by the virtual reality scenarios.

3.4. Experimental Process

The experimental process is designed in stages outlined in a detailed protocol, intended to sequentially achieve the experimental objectives. Each participant undergoes a series of tasks within the virtual reality environment while their physiological responses are monitored. The data collected include both the psychophysiological signals and self-reported emotional assessments after each virtual reality scenario, ensuring comprehensive profiling of emotional responses. Emphasis on ethical considerations is paramount, with GDPR compliance ensured through informed consent forms detailing participant rights and data protection measures. This structured approach to design, execution, and data handling aims to robustly capture the nuances of human emotion in response to virtual stimuli, significantly contributing to the field of emotion recognition research.

The process starts with detailed participant onboarding. Here, each participant is thoroughly informed about the general experiment's goals. They receive detailed information sheets developed according to the specifications in Supplementary Materials, which also include tailored informed consents, ensuring ethical compliance and participants' awareness of their rights, such as data protection and the right to withdraw. This stage is critical not just for ethical reasons but also to ensure that participants are comfortable and fully informed about the experiment.

During preparation, participants are equipped with the necessary devices for collecting psychophysiological data, including an EEG headset optimised for use with VR technologies. This preparation stage is crucial as it involves setting up the participant with the various measurement devices, with special attention to ensure that there are no restrictions that could affect the data quality or participant comfort (see Figure 2).

The second stage is the calibration phase. Here participants engage in predefined activities that elicit states of deep relaxation followed by mild stress. This allows for the precise calibration of the equipment to each individual's unique physiological responses. The calibration ensures that subsequent measurements of EEG, GSR, HR, and facial expressions are accurate and reflective of true emotional responses to the stimuli introduced later in the experiment.

Following calibration, the familiarisation phase helps participants acclimate to the specific activities they will encounter during the main experimentation phase. By introducing them early to the processes and equipment, any potential stress or anxiety relating to technological unfamiliarity or the experimental environment is minimised. This preparatory stage is crucial for setting a controlled baseline from which genuine emotional and physiological responses can be observed and recorded.



Figure 2. Participants wearing the measuring devices during the experimental phase.

The core of our study is the experimentation phase. In this stage, participants are exposed to various controlled stimuli that are specifically designed to invoke emotional responses as outlined in our research objectives. They adhere to instructions that were iteratively refined during the preparation phase. Throughout this critical phase, our research team closely monitors adherence to protocol, ensuring that participant interactions with the stimuli are consistent and that the data integrity is maintained. Data collected here form the primary dataset for our eventual analysis.

After the completion of the experiment, a debriefing session is conducted. Here, participants are assisted in removing the measurement devices. Then an informal interview is held to discuss their experiences and gather feedback, which provides critical subjective insights into the participants' emotional journey and any significant occurrences during the experiment. The data collected, which include EEG readings, heart rate, galvanic skin response, and self-reported emotional states, are processed and analysed to derive conclusions about the psychophysiological impact of VR-induced emotional states.

4. Results

In this section, the results of previous experimentation are presented and delve into the development of two distinct classifiers aimed at detecting basic human emotions by leveraging a combination of psychophysiological signals and facial images. By integrating data from both physiological responses, such as brain signals, heart rate, skin conductance, and facial expressions captured through advanced image processing techniques, these classifiers offer a comprehensive approach to emotion recognition. The fusion of psychophysiological signals and facial features allows for a more nuanced and accurate understanding of emotional states, enabling the classifiers to distinguish between emotions like joy, sadness, anger, fear, disgust, surprise, anticipation, and trust with greater precision and reliability.

4.1. Classifier Based on Psychophysiological Signals

4.1.1. Dataset Analysis

The purpose of this classifier is to establish the relationship between the acquired psychophysiological signal variations and the corresponding emotional state that provoked them. To do so, the designed experimental procedure must trigger the desired emotional spectrum. The participants were asked to rate their emotional states via a standardised 1–5 scale for each emotion both as a general perception of the stimuli and on different stages of them. Table 5 shows the distribution of the self-assessment emotional responses of the participants for the general perception of each stimulus, segmented according to both gender and experimental stimuli, along with the emotional response expected per stimulus.

Table 5. Mean value of participants' emotional assessment responses over the stimuli.

Stimuli		#1	#2	#3	#4	#5	#6	#7	#8	#9	#10
Anger	Mean	0.096	0.058	0.971	0.058	0.297	0.360	0.096	0.866	0.210	1.113
	Female	0.192	0.115	1.115	0.115	0.115	0.720	0.192	0.385	0.192	1.400
	Male	0.000	0.000	0.826	0.000	0.478	0.000	0.000	1.348	0.227	0.826
Anticipation	Mean	4.033	1.591	1.533	3.468	4.454	1.383	1.584	1.202	1.818	3.293
	Female	4.327	1.442	1.654	3.154	4.365	1.000	1.038	1.231	2.000	2.500
	Male	3.739	1.739	1.413	3.783	4.543	1.767	2.130	1.174	1.636	4.087
Disgust	Mean	0.090	0.731	0.304	0.692	0.567	1.462	0.314	0.239	0.114	0.234
	Female	0.115	0.462	0.346	0.385	0.600	1.750	0.192	0.000	0.000	0.120
	Male	0.065	1.000	0.261	1.000	0.533	1.174	0.435	0.478	0.227	0.348
Fear	Mean	0.051	1.016	0.500	2.059	0.313	0.113	0.144	0.472	0.137	0.791
	Female	0.058	1.250	0.827	1.923	0.627	0.060	0.288	0.769	0.115	0.800
	Male	0.043	0.783	0.174	2.196	0.000	0.167	0.000	0.174	0.159	0.783
Joy	Mean	4.232	0.532	0.148	0.104	0.071	2.353	3.047	0.065	1.163	1.255
	Female	4.115	0.173	0.231	0.077	0.077	1.540	3.615	0.000	0.962	0.880
	Male	4.348	0.891	0.065	0.130	0.065	3.167	2.478	0.130	1.364	1.630
Neutral	Mean	1.373	0.087	0.087	0.000	0.385	0.000	1.894	0.397	0.867	0.239
	Female	1.442	0.000	0.000	0.000	0.769	0.000	1.962	0.577	0.462	0.000
	Male	1.304	0.174	0.174	0.000	0.000	0.000	1.826	0.217	1.273	0.478
Sadness	Mean	0.096	0.083	2.736	0.061	0.000	0.000	0.178	0.340	0.092	0.054
	Female	0.192	0.058	2.712	0.058	0.000	0.000	0.269	0.115	0.115	0.000
	Male	0.000	0.109	2.761	0.065	0.000	0.000	0.087	0.565	0.068	0.109
Surprise	Mean	0.150	0.457	0.388	0.304	0.690	0.420	0.154	0.822	0.338	0.030
	Female	0.038	0.000	0.058	0.346	0.423	0.640	0.308	0.712	0.154	0.060
	Male	0.261	0.913	0.717	0.261	0.957	0.200	0.000	0.932	0.522	0.000
Trust	Mean	0.558	0.602	0.714	0.154	0.362	0.847	0.707	0.424	1.514	0.230
	Female	0.442	0.269	0.212	0.308	0.288	0.360	0.500	0.000	1.346	0.200
	Male	0.674	0.935	1.217	0.000	0.435	1.333	0.913	0.848	1.682	0.261

Grey columns indicate the expected main emotion to be triggered by each experimental stimulus.

This table shows the adequacy of the experimental procedure as all the expected emotional responses are triggered during the experiment. Furthermore, despite the data variation and the inherent subjectivity on the self-assessment task, each emotion reaches its maximum self-assessed values on the expected stimuli, such as fear on stimuli #2 and #4 and sadness in stimuli #3. Table 5 also shows that anticipation was a dominant emotion along the whole process, probably due to the participants' interest and curiosity concerning the experiment itself.

Table 5 also shows a slight disparity in emotional responses among genders, especially regarding the secondary emotional responses. For instance, stimulus #6 provoked a spike of joy among the male participants, even surpassing the expected emotion of disgust, whereas female participants showed surprise and anger responses, (almost) inexistent among the male population. Similarly, stimulus #8 elicited a broader range of emotional responses among male participants, such as anger, disgust, sadness, or even trust, whereas the female participants' reaction varied between indifference and fear.

4.1.2. Classifier Development

The observed emotional asymmetry between genders suggests that developing distinct machine learning models tailored to each gender could be advantageous, providing more precise results. Consequently, the complete emotional assessment dataset was classified according to participants' gender.

In this study, the preprocessing of signals was strategically tailored to align with the specific needs of our research. For EEG, while artefact removal is often crucial, the initial focus was on frequency domain filtering and wavelet-based time-frequency decomposition.

These methods were chosen based on the data quality and the conceptual focus of our investigation, which emphasised spectral features over temporal artefacts. Frequency domain filtering effectively limited the analysis to relevant EEG frequencies, reducing noise while maintaining the data integrity for spectral analysis. Simultaneously, wavelet transformations facilitated a robust time-frequency analysis, providing deeper insights into the dynamics of brain activity across various cognitive and emotional states. For GSR data, Butterworth filtering was employed to ensure a smooth signal conducive to reliable physiological interpretation. Additionally, for the BVP data, preprocessing involved both high-pass and low-pass filtering to remove artefacts caused by slower physiological changes and high-frequency external noise.

This combined approach to data preprocessing was designed to optimise both the clarity and relevance of the signal data, thereby supporting reliable and reproducible outcomes in our study of cognitive and emotional processes. Data from each gender-specific set were then meticulously selected to extract the cleanest possible data for algorithmic training. To achieve this, any record with corrupted and/or discontinuous data was discarded from the analysis. Subsequently, features were extracted from each record, and those with 5% or more outlier features were excluded from further analysis.

To identify the most representative psychophysiological variations for each emotion, the purity of the participant's emotion was taken into account. Thus, based on the standardised 1–5 values of the self-assessed emotions on different stages of the stimuli, the impact of each emotion was calculated as the contribution of its self-assessment score to the total sum of the assessment. The emotional purity threshold was set at 0.8; thus, only records presenting an emotion with a score over the threshold were considered as pure emotional status in contrast with mixed emotional statuses. Finally, data resembling pure emotional status per gender were divided into the usual train and test datasets. These steps prioritised data quality over quantity, as major data loss took place when dividing pure and mixed emotional statuses. At the end, only 39% of the original data were used in the classification training and testing processes. Table 6 shows the data distribution along the process.

Table 6. Psychophysiological dataset data distribution.

Experimental Data					
2773					
100%					
Correct					Corrupted
2496					277
90%					10%
Male			Female		
972			1524		
39%			61%		
Standard		Outliers	Standard		Outliers
917		55	1442		82
94%		6%	95%		5%
Pure emotions		Mixed emotions	Pure emotions		Mixed emotions
430		487	655		787
47%		53%	45%		55%
Train	Test		Train	Test	
344	86		526	129	
80%	20%		80%	20%	

The uncorrupted psychophysiological signals were used to extract different features for the ML process. According to the nature of each signal, the following were calculated:

- Blood volume pressure (BVP): Using 5 s signals chunks, the participants' heart rate (HR) and heart rate variation (HRV) were calculated.

- Galvanic skin response (GSR): Based on 5 s signal chunks, the signal was split into tonic (continuous) and phasic (spontaneous) components. Later, the amplitude, rise time, and recovery time of the phasic signal along with its mean, standard deviation, energy, and root mean square values were calculated. These last four components for the gradient function of the phasic component were also calculated.
- Electroencephalography (EEG): The brain activity variation is much faster than the skin response; thus, in this case, 2 s signal chunks were used in the calculation process. For each of the 12 electrodes involved, we performed a time domain analysis comprising mean value of the signal, the Hjorth parameters—activity, mobility and complexity—root mean square, signal range, signal energy, and signal's Higuchi fractal dimension. Moreover, a spectral analysis was carried out to obtain each electrode's theta, alpha, beta, and gamma bands. For each of the bands, the band power spectrum, its contribution to the total brain activity energy, and bands' entropy were calculated. This led to a total of 26 features per electrode every 2 s.

This provides two features from the BVP (blood volume pulse) signal, 11 from the GSR (galvanic skin response) signal, and 312 from the EEG (electroencephalography) signal (26 for each of the 12 electrodes involved), leading to a grand total of 325 features for each of the data records. As the number of features is high, especially compared to the number of records available for training (344 records for the male model and 526 for the female) and considering the nature of the features (continuous numerical values) and the response (categorical emotion label), the ANOVA f-value score was used to identify and select the best features for the model, thus reducing the dimensionality.

Regarding the model itself, different approaches were implemented with different success rates:

- First, the problem was modelled as a multiclass classifier (MCC). Each of the training records was labelled according to the most predominant emotion, resulting in nine possible classes: anger, anticipation, disgust, fear, joy, sadness, surprise, trust, and neutrality/indifference. Utilising a random forest classifier, each of the test records was then assigned to one of these emotion classes based on the learned patterns from the training data.
- Then, the one-versus-one (OvO) approach was used to model the classification problem. In this approach, the problem is decomposed into pairwise comparisons between each pair of emotional classes. During testing, the model makes multiple binary decisions, and the final emotional class is determined by a voting mechanism.
- Next, the problem was modelled using the one-versus-rest (OvR) strategy. In this case, the multiclass problem is framed as a series of binary tasks, treating each emotional class as positive while considering the rest as negative. Then, the model assigns a record to the class with the highest probability.
- Considering that the continuous emotional modelling approach might provide some insight, the reference valence–arousal–dominance (VAD) values for each of the nine basic emotions was established and the problem modelled as a three-dimensional regression model (VAD-R). Once the regression model was trained, the VAD values of each of the test records was estimated and, considering that the emotions to detect may not be 100% pure, the inverse linear superposition problem was resolved via optimisation techniques to identify the combination of emotions that best matches the calculated VAD values. Finally, the dominant emotion was set as a result of the problem. In this particular case, as both features and response are numerical in nature, Pearson's correlation was used to determine the best features for the model.
- Finally, the last model also employs the VAD (valence–arousal–dominance) emotional model, but in this case, instead of using a three-dimensional continuous variable, each of the dimensions were modelled as categorical variables, setting them as high–medium–low values. This way, each emotion is set within a specific space on the VAD spectrum, and the problem is modelled as a triple classification problem (VAD-C).

Once the algorithm is trained, the VAD levels of each test record are calculated, and the predicted emotion is set accordingly.

4.1.3. Obtained Results

Table 7 shows the different accuracy metrics obtained during the training process.

Table 7. Accuracy metrics of the different methods.

Model	Scope	Metrics		
		Accuracy	Cohen's Kappa	R ²
MCC	Male emotional state	0.66	0.52	-
	Female emotional state	0.64	0.54	-
OvO/OvR	Male emotional state	0.62	0.46	-
	Female emotional state	0.47	0.34	-
VAD-R	Male valence	-	-	0.44
	Male arousal	-	-	0.57
	Male dominance	-	-	0.34
	Female valence	-	-	0.26
	Female arousal	-	-	0.29
	Female dominance	-	-	0.27
VAD-C	Male valence	0.93	0.84	-
	Male arousal	0.85	0.66	-
	Male dominance	0.85	0.70	-
	Male emotional state	0.73	0.61	-
	Female valence	0.93	0.89	-
	Female arousal	0.86	0.76	-
	Female dominance	0.96	0.93	-
	Female emotional state	0.73	0.68	-

For the classification problems, both the accuracy score and Cohen's kappa metrics are employed. For the regression approach, the R² correlation was used to measure the adequacy of the proposed method.

The table shows that, overall, the regression approach offered poor results, which may be related with the non-numerical nature of the emotions. Also, the OvO (one-versus-one) approach offered identical results to OvR (one-versus-rest), even if it required additional computing time. Simultaneously, the VAD-C approach outperforms other algorithms, proving to be the most effective way to detect the emotional status for both genders.

Also, Figures 3 and 4 show the classification matrix of test datasets for the VAD-C model. Note that the values are similar to those obtained during the training. The accuracy score for the male emotional model shows the greatest difference between the training and the test results, with all the other metrics maintaining very similar values between both datasets.

These figures show that the models work correctly for most of the emotional states. However, both models fail to detect the fearful state and predict it mostly as 'surprise' or 'disgust'. Also, the number of pure 'trust' emotional states is very low, just a single case for the female model and none for the male model. However, as the training process involved is based upon the VAD values and not directly upon the labelled emotions, the resulting models are able to deal with said emotional state.

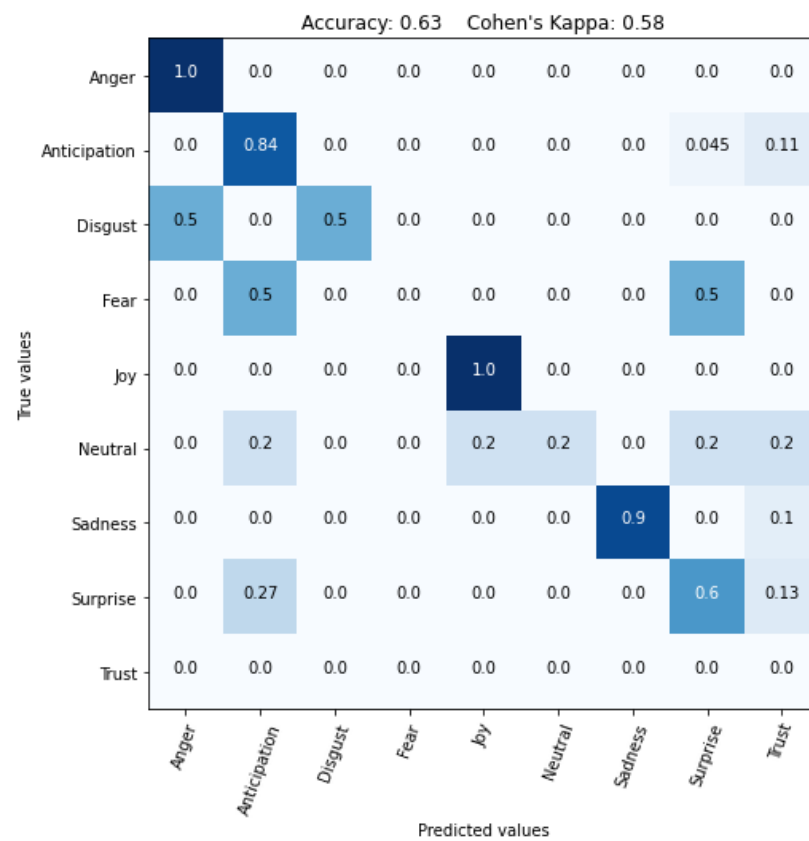


Figure 3. Normalised confusion matrix for male emotions with VAD-C model.

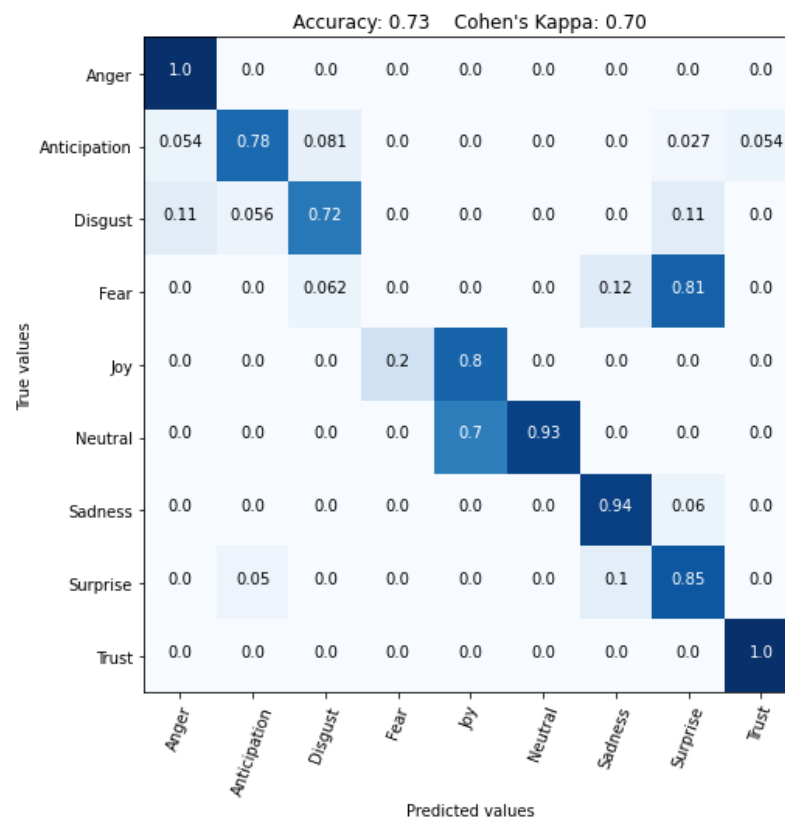


Figure 4. Normalised confusion matrix for female emotions with VAD-C model.

4.2. Classifier Based on Facial Images

4.2.1. Dataset Analysis

The significance of a dataset in training a convolutional neural network (CNN) for image classification cannot be overstated. A robust and diverse dataset serves as the basis for the CNN to learn to recognise patterns, features, and distinctions within images. This is essential to ensure that the algorithm learns to differentiate between different classes accurately. The size of the dataset also plays a crucial role, as a larger dataset often leads to improved generalisation and performance of the CNN, reducing the risk of overfitting.

Another important factor to consider is whether the images are posed, where the subject is instructed to display a specific emotion, or spontaneous, where the expressions occur naturally without prompting. Especially in the latter case, a well-labelled dataset is crucial, as it provides the CNN with accurate ground truth information, enabling it to learn and refine its classification abilities effectively.

To meet these requirements, a total of five commercial datasets were analysed, and two of them were acquired to train the facial emotional recognition: Pixta [34], composed by posed and spontaneous images (3800 Images), labelled by five people and Datatang [35], based on posed expressions of 1350 people, each of them showing all different emotions (see Figure 5).

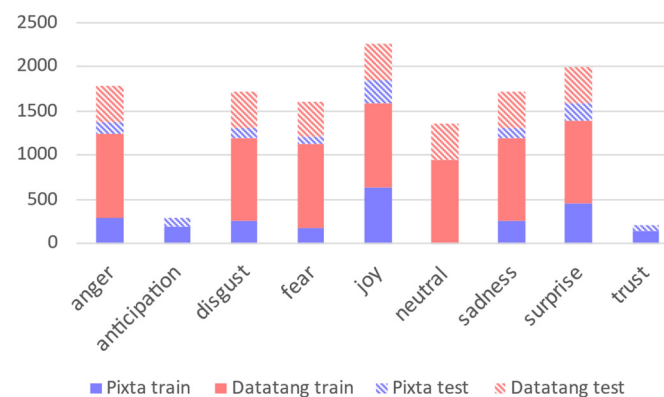


Figure 5. Dataset pictures distribution.

From the original Pixta dataset of 3800 images, 339 images were discarded due to the difficulty of the algorithm to identify the faces they contain because they were rotated or partially occluded (false negative) or because the detection area was not a face (false positive). The available datasets do not cover the full range of target emotions, complementing each other. In any case, upon looking at the images available for each category, it stands out that the emotions “anticipation” and “trust” are poorly represented, so the classifier result will not be relevant for those emotions.

4.2.2. Classifier Development

The development of the classifier consists of two distinct parts: dataset preparation and classifier algorithm (see Figure 6).

Dataset preparation consists of preprocessing dataset images, distributing them into different groups for training/testing, and covering the eye area to simulate wearing a VR headset.

- **Preprocessing dataset images:** This process begins by employing specialised algorithms to accurately detect and extract faces in each image, then cropping them to isolate them from background noise. The algorithm used for face detection and extraction is based on a dlib’s pretrained model for a face detector through a function called ‘get_frontal_face_detector’. This module allows for localising a face within the image and detecting the key facial structures’ landmarks, such as the position of the

eyes, eyebrows, nose, etc. With this information, the algorithm crops the facial area and resizes it to a uniform size. Resizing ensures consistency across the entire dataset, allowing machine learning models to focus on facial features without being influenced by variations in image sizes.

- **Dataset distribution:** The dataset is split into training and testing subsets. The splitting is done randomly but considers the emotional stratification to ensure a representative distribution of data across both sets and to avoid biases in the model's learning process. In both datasets (Datatang and Pixta), the ratio used to split the images used for train and test has been the same: 70% of the images have been used for train, and the remaining 30% have been reserved for test. The Datatang set offered an image of each emotion analysed per subject, so the process randomly selects 70% of subjects to be included in training, leaving the remaining 30% for testing. As a result, all emotions were perfectly represented in each group, with no contamination between the two groups. In the case of the Pixta dataset, the algorithm first processes all the images, discards those that do not meet the quality criteria (for instance, due to a wrong facial identification), and classifies them according to the labelled emotion before proceeding to the train–test split. This ensures that each emotion maintains the 70–30% ratio between the train and test groups.
- **Eye covering:** Due to the lack of a dataset for people wearing virtual reality headsets, during the preprocessing stage the eye area of the subjects was covered to simulate that they were wearing a VR headset. The process recognises the pose of the face and, based on that data, adjusts the blocked area to the rotation and size of the face, based on the algorithm presented by [23]. The facial features are thus occluded, as the VR headset was not used during the training process.

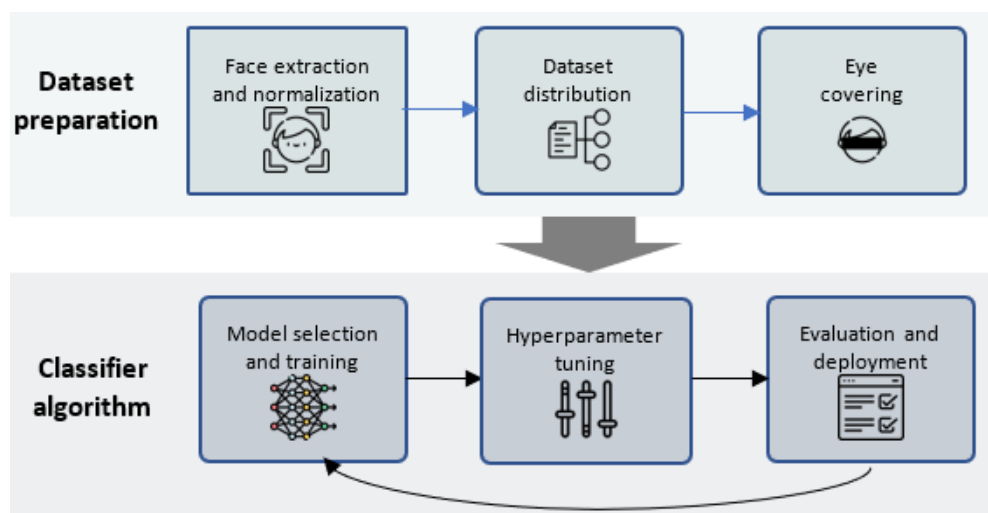


Figure 6. Classifier approach for facial expressions.

The refinement and enhancement of the classifier algorithm was achieved through the systematic execution of the subsequent tasks:

- **Model selection and training:** Based on the problem's nature and dataset characteristics, a convolutional neural network (CNN) was chosen to perform image-based classification. The phase also includes the training of the chosen model on the prepared dataset.
- To improve the capacity of the model, a transfer learning strategy was chosen, based on the RESNET50 model. RESNET50 is a pretrained deep convolutional neural network renowned for its depth and performance in image classification tasks. Leveraging transfer learning with RESNET50 involves using its learned features and architecture as a foundation for a new, specific task, such as image classification, without starting

the training process from scratch. By reusing the already learned representations from RESNET50 and fine-tuning its layers or adding new ones suited to the particular problem, the model can rapidly adapt to new data and tasks. This strategy typically results in quicker convergence, better generalisation, and improved accuracy compared to training a model from the ground up.

- Three new final layers were added for the model to classify the images into the nine target emotions. Furthermore, based on the RESNET50 architecture, some layers have been frozen to maintain their original learning, while others have been left as trainable to adapt to the new classification task.
- Hyperparameter tuning: Fine-tuning the model's hyperparameters to enhance its performance, like epochs, batch size, dropout rate, learning rate, etc.
- Evaluation and deployment: The model's performance is evaluated using various metrics, such as accuracy, loss, or kappa, on a test dataset. Once satisfied with its performance, the classifier algorithm is deployed to make predictions on new, unseen data.

4.2.3. Obtained Results

Figure 7 shows the confusion matrix on the test set, where the highest success rate is achieved by the 'joy' emotion, which has a high number of samples in the dataset and whose main characteristics for identification are focused on the mouth area. In contrast, emotions such as 'anticipation', 'disgust' or 'trust' have a low success rate due to their characteristics and/or their presence in the dataset.

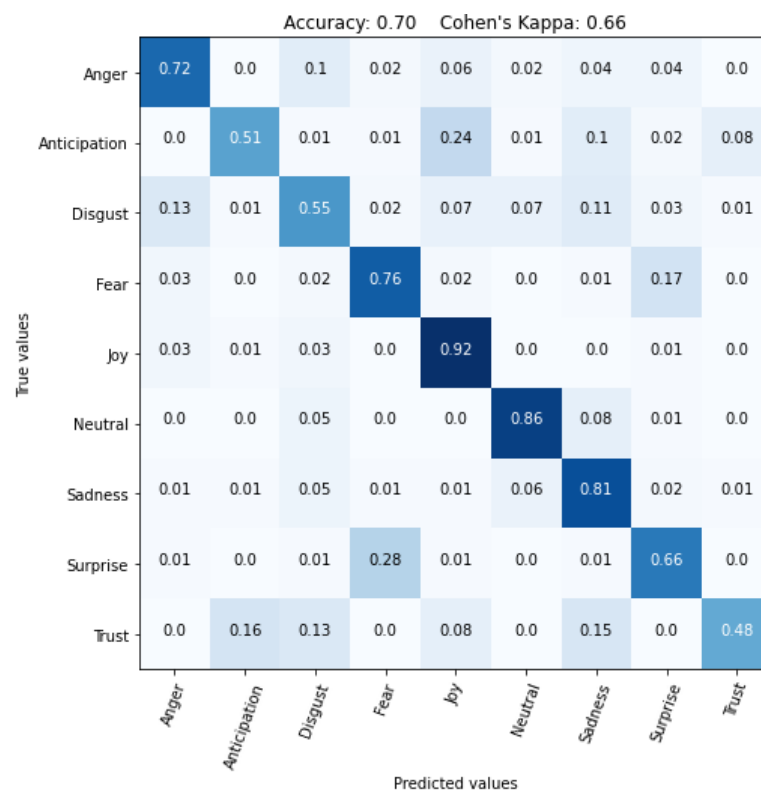


Figure 7. Normalised confusion matrix for facial expression on test set.

However, the results obtained by applying the classifier to real users with the VR headset on during the experiments are far from the results obtained with the test dataset. For example, Figure 8 shows the disparity between the outcome of the real-time classifier and user self-assessment.

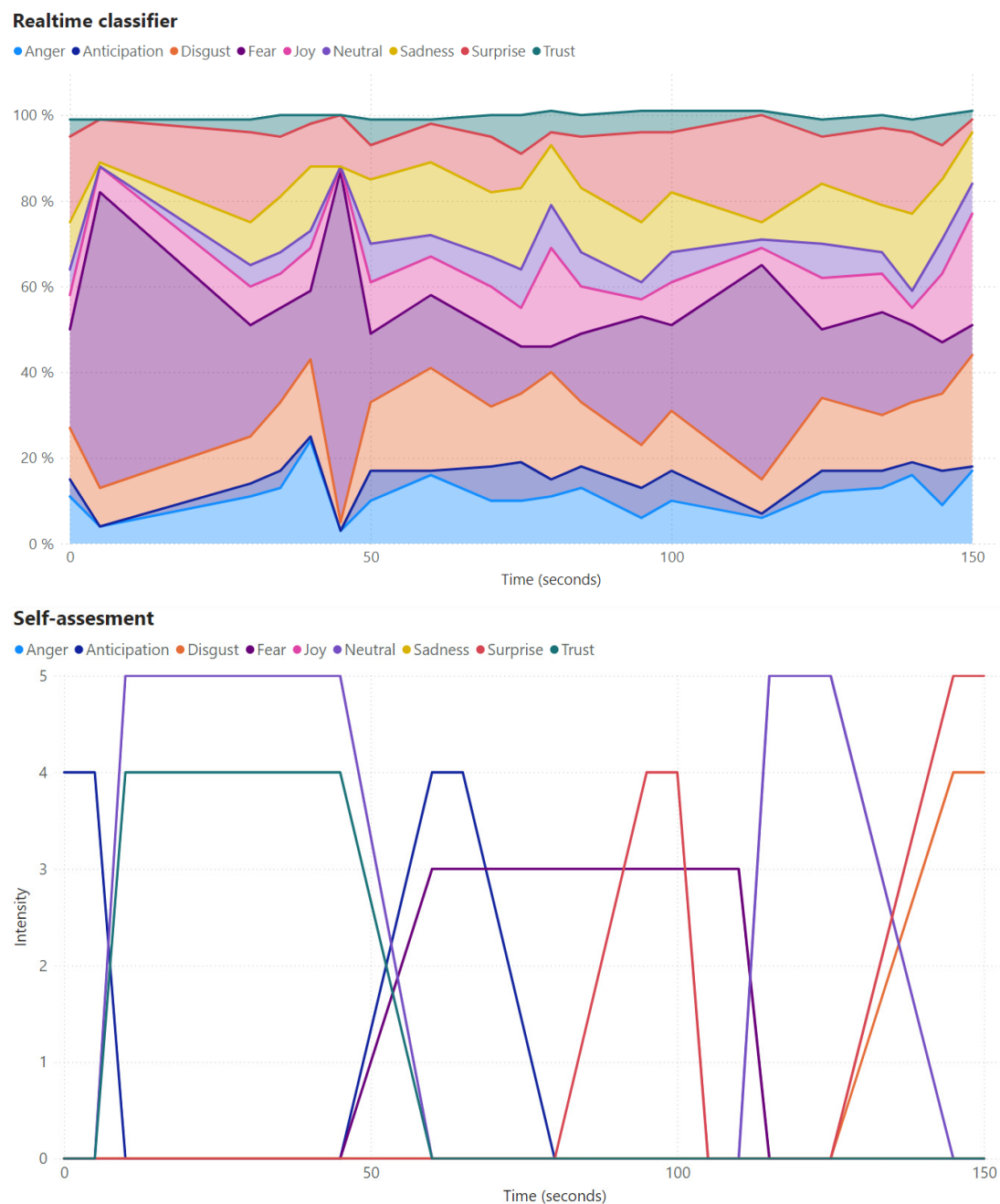


Figure 8. Comparative of real-time classifier and user self-assessment.

The real-time system does not provide a reliable prediction when compared to the emotion stated by users. More restrained expressions, such as 'trust' and 'neutral', are more difficult for the system to interpret. While other more marked expressions, such as 'fear', 'disgust' or 'joy', perform better, they are also more over-represented in the results. This may be due to multiple factors:

- The virtual reality headset used is large so that in addition to covering the eye area, it also hides the nose area, with the consequent loss of information.
- The quality of the images of the participants is poor, mainly due to the lighting conditions, which are a decisive factor for the correct functioning of the system. In addition, the virtual reality headset itself casts a shadow on the small visible part of the face.

- The expressiveness of the users during the experiments is lower than that provided by the models who participated in the creation of the dataset.

5. Discussion

This study undertook a deep dive into emotion detection within virtual reality interactions, specifically targeting three core research hypotheses. The investigation not only broadened our academic understanding but also revealed insights with broad implications across multiple sectors.

Our examination of models for categorising emotions highlighted a rich diversity of frameworks, each attempting to decode the complexity of human emotional experiences. A critical takeaway from our findings is the inadequacy of a one-size-fits-all approach in categorising emotions. This study's adoption of dynamic, spectrum-based models—reflecting recent psychological theories—illustrates that emotions are shaped by various factors, including context, culture, and individual physiological responses. The results from the psychophysiological classifier model challenge the first hypothesis by affirming that the VAD-C model, a mixed spectrum-based and categorical model, provides a more accurate and nuanced representation of emotions, thereby setting the stage for improved detection methodologies.

In addressing the second hypothesis, a thorough exploration of human signals and features was performed. This was crucial for emotional assessment. The combination of psychophysiological signals and facial expressions provided comprehensive insights into individuals' emotional states. This multimodal approach not only harnesses the strengths of each signal type, enhancing detection accuracy, but also aligns with theories of emotional embodiment. This underscores the possibility of creating more empathetic and intuitive human–computer interactions and points towards practical applications that leverage these insights.

Concerning the third hypothesis focused on VR environments, several challenges were identified, notably the accuracy of facial emotion detection hampered by virtual reality headsets. This necessitates a shift towards psychophysiological signals, which are less affected by these limitations. It prompts the field towards refining methodologies and seeking innovative solutions to navigate VR's constraints. This direction of research underlines the need for advancements that ensure emotion detection remains accurate and tailored to user needs within virtual reality contexts.

This study has paved the way towards understanding the complex relationship between human emotions and digital interfaces. By dissecting existing models, identifying key signals for emotion detection, and tackling the unique challenges of virtual reality environments, significant progress towards creating technology capable of genuinely understanding human emotions has been proposed. The findings suggest a promising future for digital tools designed to resonate closely with our emotional states.

Looking forward, it is clear that further research is needed to overcome the technical challenges which have been identified, especially in enhancing a holistic multimodal emotion detection within virtual reality environments. There is also a pressing need to explore the application of these findings in real-world settings, spanning healthcare, education, and beyond, to fully realise the potential of empathetic and intuitive human–computer interactions. This study lays the groundwork for future investigations that can build on our findings to create technologies that bridge the gap between human emotions and digital response capabilities more effectively.

6. Conclusions

In conducting this study, the main aim was to investigate how technology can be better used to understand human emotions, particularly within the context of human–computer interactions. The focus was on three critical areas: emotion categorisation, which human signals are important for detecting emotions, and the role of virtual reality technology in improving emotion detection.

This research shows that emotions are far more complex than traditional models suggest. Instead of fitting emotions into fixed categories, enriching this vision with a spectrum-based approach provides a more accurate reflection of the wide variety of human emotional experiences. This approach is key for developing technologies that can recognise and adapt to the nuances of our emotions more effectively.

The importance of using multiple types of signals, such as physiological cues and facial expressions, to understand emotions better has been emphasised. This multi-signal approach is crucial since it allows for gathering more detailed information about emotional states, leading to more accurate interpretations by machines.

The use of virtual reality technology offers exciting possibilities for creating environments that can naturally elicit emotional responses. However, there are challenges to overcome in order to implement non-intrusive emotional detection solutions, especially with virtual reality headsets that can obstruct full facial expression analysis. Despite these obstacles, virtual reality holds significant promise for advancing emotion detection research, but it requires creative thinking to solve these technical issues.

In conclusion, this study contributes valuable insights into enhancing the interaction between humans and computers through better emotion detection. It points to a future where technology can understand and respond to our emotions in more refined and empathetic ways, which has far-reaching implications for various fields, such as healthcare, education, and entertainment. As we move forward, the key will be to continue innovating and exploring new methods that allow for more nuanced detection and interpretation of human emotions, aiming for a digital world that is more responsive and attuned to our emotional well-being.

Supplementary Materials: The following supporting information are available online at <https://www.mdpi.com/article/10.3390/mti8060047/s1>.

Author Contributions: Conceptualisation, E.L., L.B., A.M., S.S. and E.T.; methodology, E.L.; software, E.L. and E.T.; validation, L.B., A.M. and S.S.; formal analysis, E.L. and E.T.; investigation, L.B., A.M. and S.S.; resources, L.B. and S.S.; data curation, E.L. and E.T.; writing—original draft preparation, L.B., S.S., E.L., E.T. and A.M.; writing—review and editing, L.B., S.S. and A.M.; visualisation, A.M.; supervision, L.B. and S.S.; project administration, L.B.; funding acquisition, L.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received funding from the European Union’s Horizon 2020 Research and Innovation Programme under grant agreement No. 101021714. The results obtained in this work reflect only the authors’ views and not the ones of the European Commission; the Commission is not responsible for any use that may be made of the information they contain.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the LAW-GAME’s Ethics Manager (EM) from VRIJE UNIVERSITEIT BRUSSEL (February 2022).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study, and participation in the study was voluntary.

Data Availability Statement: Data are contained within the article.

Acknowledgments: The authors would like to thank the participants who took part in the study.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Ortony, A.; Clore, G.L.; Collins, A. *The Cognitive Structure of Emotions*; Cambridge University Press: Cambridge, UK, 2022.
- Izard, C. Emotion theory and research: Highlights, unanswered questions, and emerging issues. *Annu. Rev. Psychol.* **2009**, *60*, 1–25. [[CrossRef](#)] [[PubMed](#)] [[PubMed Central](#)]
- McCraty, R.; Atkinson, M.; Tomasino, D.; Bradley, R. The Coherent Heart Heart–Brain Interactions, Psychophysiological Coherence, and the Emergence of System-Wide Order. *Integral Rev.* **2009**, *5*, 10–115.
- Egger, M.; Ley, M.; Hanke, S. Emotion Recognition from Physiological Signal Analysis: A Review. *Electron. Notes Theor. Comput. Sci.* **2019**, *343*, 35–55. [[CrossRef](#)]

5. Schöne, B.; Kisker, J.; Lange, L.; Gruber, T.; Sylvester, S.; Osinsky, R. The Reality of Virtual Reality. *Front. Psychol.* **2023**, *14*, 1093014. [CrossRef] [PubMed]
6. Li, R.; Ding, J.; Ning, H. Emotion arousal assessment based on multimodal physiological signals for game users. *IEEE Trans. Affect. Comput.* **2023**, *14*, 2582–2594. [CrossRef]
7. Shu, L. A review of emotion recognition using physiological signals. *Sensors* **2018**, *18*, 2074. [CrossRef]
8. Sreeja, P.; Mahalakshmi, G. Emotion Models: A Review. *Int. J. Control. Theory Appl.* **2017**, *10*, 651–657.
9. Baig, M.; Kavakli, M. A Survey on Psycho-Physiological Analysis & Measurement Methods in Multimodal Systems. *Multimodal Technol. Interact.* **2019**, *3*, 37. [CrossRef]
10. Dzedzickis, A.; Kaklauskas, A.; Bučinskas, V. Human Emotion Recognition: Review of Sensors and Methods. *Sensors* **2020**, *20*, 592. [CrossRef]
11. Ekman, P. Facial expression and emotion. *Am. Psychol.* **1993**, *48*, 384. [CrossRef]
12. Plutchik, R. The Nature of Emotions. *Am. Sci.* **2001**, *89*, 344–350. [CrossRef]
13. Russel, J.; Weiss, A.; Mendelsohn, G. Affect grid: A single-item scale of pleasure and arousal. *J. Pers. Soc. Psychol.* **1989**, *57*, 493–502. [CrossRef]
14. Mehrabian, A. Comparison of the PAD and PANAS as models for describing emotions and for differentiating anxiety from depression. *J. Psychopathol. Behav. Assess.* **1997**, *19*, 331–357. [CrossRef]
15. Verma, G.K.; Tiwary, U. Affect representation and recognition in 3D continuous valence–arousal–dominance space. *Multimed. Tools Appl.* **2017**, *76*, 2159–2183. [CrossRef]
16. Healey, J.; Picard, R. Detecting stress during real-world driving tasks using physiological sensors. *IEEE Trans. Intell. Transp. Syst.* **2005**, *6*, 156–166. [CrossRef]
17. Kreibitz, S. Autonomic nervous system activity in emotion: A review. *Biol. Psychol.* **2010**, *84*, 394–421. [CrossRef]
18. Zheng, W.; Zhu, J.; Lu, B. Identifying Stable Patterns over Time for Emotion Recognition from EEG. *IEEE Trans. Affect. Comput.* **2016**, *10*, 417–429. [CrossRef]
19. Liu, H.; Zhang, Y.; Li, Y.; Kong, X. Review on Emotion Recognition Based on Electroencephalography. *Front. Comput. Neurosci.* **2021**, *15*, 84. [CrossRef] [PubMed]
20. Bota, P.; Wang, C.; Fred, A.; Plácido Da Silva, H. A Review, Current Challenges, and Future Possibilities on Emotion Recognition Using Machine Learning and Physiological Signals. *IEEE Access* **2019**, *7*, 140990–141020. [CrossRef]
21. Hasnul, M.; Abd Aziz, N.; Alelyani, S.; Mohana, M.; Abd Aziz, A. Electrocardiogram-Based Emotion Recognition Systems and Their Applications in Healthcare—A Review. *Sensors* **2021**, *21*, 5015. [CrossRef]
22. Siedlecka, E.; Denson, T. Experimental Methods for Inducing Basic Emotions: A Qualitative Review. *Emot. Rev.* **2019**, *11*, 87–97. [CrossRef]
23. Houshmand, B.; Khan, N. Facial Expression Recognition Under Partial Occlusion from Virtual Reality Headsets based on Transfer Learning. In Proceedings of the IEEE Sixth International Conference on Multimedia Big Data, New Delhi, India, 24–26 September 2020.
24. Dael, N.; Mortillaro, M.; Scherer, K. Emotion expression in body action and posture. *Emotion* **2012**, *12*, 1085–1101. [CrossRef] [PubMed]
25. Chen, J.; Hu, B.; Wang, Y.; Moore, P.; Dai, Y.; Feng, L.; Ding, Z. Subject-independent emotion recognition based on physiological signals: A three-stage decision method. *BMC Med. Inform. Decis. Mak.* **2017**, *17*, 167. [CrossRef]
26. Ali, M.; Al Machot, F.; Haj Mosa, A.; Jdeed, M.; Al Machot, E.; Kyamakya, K. A Globally Generalized Emotion Recognition System Involving Different Physiological Signals. *Sensors* **2018**, *18*, 1905. [CrossRef] [PubMed]
27. Strack, F.; Martin, L.L.; Stepper, S. Inhibiting and facilitating conditions of the human smile: A nonobtrusive test of the facial feedback hypothesis. *J. Personal. Soc. Psychol.* **1998**, *54*, 768–777. [CrossRef] [PubMed]
28. Zuckerman, M.; Hall, J.A.; DeFrank, R.S.; Rosenthal, R. Encoding and decoding of spontaneous and posed facial expressions. *J. Pers. Soc. Psychol.* **1976**, *34*, 966. [CrossRef]
29. Motley, M.T.; Camden, C.T. Facial expression of emotion: A comparison of posed expressions versus spontaneous expressions in an interpersonal communication setting. *West. J. Commun.* **1988**, *52*, 1–22. [CrossRef]
30. Maity, S.; Veer, K. An Approach for Evaluation and Recognition of Facial Emotions Using EMG Signal. *Int. J. Sens. Wirel. Commun. Control* **2024**, *14*, 113–121. [CrossRef]
31. Dong, Z.; Wang, G.; Lu, S.; Li, J.; Yan, W.; Wang, S.J. Spontaneous facial expressions and micro-expressions coding: From brain to face. *Front. Psychol.* **2022**, *12*, 784834. [CrossRef]
32. Banerjee, R.; De, S.; Dey, S. A survey on various deep learning algorithms for an efficient facial expression recognition system. *Int. J. Image Graph.* **2023**, *23*, 2240005. [CrossRef]
33. Jia, S.; Wang, S.; Hu, C.; Webster, P.J.; Li, X. Detection of Genuine and Posed Facial Expressions of Emotion: Databases and Methods. *Front. Psychol.* **2021**, *11*, 580287. [CrossRef] [PubMed]
34. Pixta. Available online: <https://www.pixtastock.com/> (accessed on 13 April 2024).
35. Datatang. Available online: <https://m.datatang.ai/> (accessed on 13 April 2024).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.