INSTITUTO SUPERIOR DE ENGENHARIA DO PORTO
MESTRADO EM ENGENHARIA DE INTELIGÊNCIA ARTIFICIAL




From Brainwaves to Emotions: Emotion Recognition in Media Consumption Using EEG and BCI

JOÃO FILIPE MOREIRA VIEIRA
Setembro de 2025

From Brainwaves to Emotions: Emotion Recognition in Media Consumption Using EEG and BCI

João Filipe Moreira Vieira
Student No.: 1201005

Dissertation for the Master's Degree in Artificial Intelligence Engineering

Supervisor: Maria Goreti Carvalho Marreiros
Co-Supervisor: Miguel Pais-Vieira
Advisor: Alexandre Ulisses

Evaluation Committee:
President:
Luís Manuel Silva Conceição, Assistant Professor, Institute of Engineering, Polytechnic of Porto

Members:
Ana Rita Teixeira, Assistant Professor, Institute of Engineering, Polytechnic of Porto
Maria Goreti Carvalho Marreiros, Full Professor, Institute of Engineering, Polytechnic of Porto

Porto, September 30, 2025

"You can change what you do, but you can't change what you want."

- Tommy Shelby

Abstract

The recognition of emotions from Electroencephalogram (EEG) signals captured by Brain-Computer Interface (BCI) is one of the emerging challenges in the field of Artificial Intelligence (AI) applied to digital media consumption. This study aimed to investigate how complete pipelines of pre-processing, feature extraction, and deep learning model training can be used to identify emotional states during the viewing of audiovisual content, focusing not only on obtaining quantitative results but also on better understanding the methodological factors that condition the models' generalisation capacity.

To this end, datasets from the SEED family were analysed and processed from raw signals, applying filtering, resampling, referencing, Independent Component Analysis (ICA), and baseline removal steps. Feature extraction included Differential Entropy (DE), Power Spectral Density (PSD), and Wavelet Transform (WT) in different frequency band schemes. Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), and CNN-LSTM models were implemented and evaluated in multiple validation scenarios, including Cross-Subject (Single-Session), Cross-Session, and Cross-Subject (Multi-Session).

The results showed that the inclusion of the *delta* band contributes to better generalisation, while high *gamma* did not show clear benefits. The combination of DE and PSD proved to be more robust than the isolated use of each feature. In terms of architectures, CNN proved to be the simplest and most computationally efficient solution, while both LSTM and CNN-LSTM showed greater potential for generalisation. It was also found that inadequate evaluation methodologies, such as classic 80/20 splits, can lead to artificially high metrics but without practical validity, reinforcing the importance of rigorous approaches such as Leave-One-Subject-Out (LOSO) validation.

Beyond the technical aspect, this dissertation gave rise to a scientific contribution, materialised in the publication of an article proposing a technical-ethical guide for the collection and processing of EEG data, in accordance with the General Data Protection Regulation (GDPR) and the AI Act. The work carried out thus constitutes not only an advance in the critical analysis of emotion recognition pipelines, but also a solid basis for future research, including the continuity ensured within the European DataPACT project.

Keywords: EEG, BCI, Emotion Recognition, Deep Learning, Feature Extraction

Resumo

O reconhecimento de emoções a partir de dados de *Electroencephalogram* (EEG) captados por *Brain-Computer Interfaces* (BCIs) constitui um dos desafios emergentes na área da inteligência artificial aplicada ao consumo de media digital. Este trabalho teve como objetivo investigar de que forma pipelines completos de pré-processamento, extração de features e treino de modelos de deep learning podem ser utilizados para identificar estados emocionais durante a visualização de conteúdos audiovisuais, privilegiando não apenas a obtenção de resultados quantitativos, mas também a compreensão crítica dos fatores metodológicos que condicionam a capacidade de generalização dos modelos.

Para tal, foram analisados e processados datasets da família SEED, a partir dos sinais brutos, aplicando etapas de filtragem, resampling, referenciação, ICA e remoção de baseline. A extração de features incluiu *Differential Entropy* (DE), *Power Spectral Density* (PSD) e *Wavelet Transform* (WT), em diferentes esquemas de bandas de frequência. Foram implementados e avaliados modelos *Convolutional Neural Network* (CNN), *Long Short-Term Memory* (LSTM) e CNN-LSTM em múltiplos cenários de validação, incluindo *Cross-Subject (Single-Session)*, *Cross-Session* e *Cross-Subject (Multi-Session)*.

Os resultados demonstraram que a inclusão da banda *delta* contribui para uma melhor generalização, enquanto a *gamma* alta não apresentou benefícios claros. A combinação de DE e PSD revelou-se mais robusta do que a utilização isolada de cada feature. Em termos de arquiteturas, a CNN mostrou-se a solução mais simples e eficiente em termos computacionais, enquanto que tanto a LSTM como a CNN-LSTM apresentaram maior potencial de generalização. Verificou-se ainda que metodologias de avaliação inadequadas, como divisões clássicas 80/20, podem conduzir a métricas artificialmente elevadas mas sem validade prática, reforçando a importância de abordagens rigorosas como a validação *Leave-One-Subject-Out* (LOSO).

Para além da vertente técnica, esta dissertação deu origem a um contributo científico autónomo, materializado na publicação de um artigo que propõe um guia técnico-ético para a recolha e tratamento de dados EEG, em conformidade com o Regulamento Geral sobre a Proteção de Dados (RGPD) e o AI Act. O trabalho desenvolvido constitui assim não apenas um avanço na análise crítica de pipelines para reconhecimento de emoções, mas também uma base sólida para investigações futuras, incluindo a continuidade assegurada no âmbito do projeto europeu DataPACT.

Palavras-chave: EEG, BCI, Emotion Recognition, Deep Learning, Feature Extraction

Acknowledgement

Completing this dissertation represented a demanding and significant challenge in my academic and personal journey, which would not have been possible without the contribution and support of various individuals and institutions. It is therefore with deep gratitude that I express my thanks.

First and foremost, I would like to express my sincere gratitude to my supervisor, Doctor Goreti Marreiros, for her time and suggestions throughout the development of this dissertation.

I would like to thank my co-supervisor, Professor Miguel Pais-Vieira, for his guidance, for always being available to answer not only theoretical but also practical questions, and for giving me the opportunity to visit his laboratory in Aveiro, allowing me to familiarise myself with a real-time EEG data collection system.

To Engineer Alexandre Ulisses, I express my gratitude for the trust placed in me, for integrating me into the EVS team and for the opportunity to carry out this work in a business environment, as well as for the ideas and contributions shared throughout the project.

To all my colleagues on the EVS team, I would also like to express my gratitude for your collaboration and support, which in one way or another contributed to the completion of this work.

To my parents, I offer my deepest gratitude for their unconditional support and for providing me with all the necessary resources, not only for the completion of this dissertation, but also for the whole of my academic career.

Finally, to my girlfriend Joana, I thank you for always being by my side, for your infinite patience during the most demanding moments, for the ideas and advice that helped this work so much and, above all, for reminding me that, even in the middle of a formatting crisis in LaTeX, there is life beyond the dissertation.

Contents

List of Figures

List of Tables

List of Acronyms

ADHD	Attention Deficit Hyperactivity Disorder.
AI	Artificial Intelligence.
ANN	Artificial Neural Network.
BCI	Brain-Computer Interface.
BDAE	Bimodal Deep Auto Encoder.
BSS	Blind Source Separation.
BVP	Blood Volume Pulse.
CAR	Common Average Reference.
CNN	Convolutional Neural Network.
CSP	Common Spatial Patterns.
DCCA	Deep Canonical Correlation Analysis.
DCT	Discrete Cosine Transform.
DE	Differential Entropy.
DFT	Discrete Fourier Transform.
DL	Deep Learning.
ECG	Electrocardiography.
EEG	Electroencephalogram.
EOG	Electrooculography.
ERP	Event-Related Potentials.
FACS	Facial Action Coding System.
FF	Feed-Forward.
FFT	Fast Fourier Transform.
GDPR	General Data Protection Regulation.
GPU	Graphic Processing Unit.
GRU	Gated Recurrent Units.
GSR	Galvanic Skin Response.
HCI	Human-Computer Interfaces.
HRV	Heart Rate Variability.
ICA	Independent Component Analysis.
KNN	K-Nearest Neighbors.
LOSO	Leave-One-Subject-Out.

LSTM	Long Short-Term Memory.
ML	Machine Learning.
OCC	Ortony, Clore and Collins.
PAD	Pleasure-Arousal-Dominance.
PCA	Principal Component Analysis.
PE	Permutation Entropy.
PLV	Phase Locking Value.
PSD	Power Spectral Density.
RNN	Recurrent Neural Network.
RQ	Research Question.
RSP	Respiration.
SE	Sample Entropy.
SNR	Signal-to-Noise Ratio.
STFT	Short-Time Fourier Transform.
SVM	Support Vector Machine.
WT	Wavelet Transform.

Chapter 1

Introduction

This chapter presents the project carried out during the semester for the curricular unit Project/Dissertation/Internship (PROJIA), of the 2nd year of the Master's Degree in Artificial Intelligence Engineering (MEIA), of the Department of Computer Engineering (DEI), of the Porto Polytechnic School of Engineering (ISEP). The project was developed as part of a professional internship at EVS Broadcast Equipment and constitutes the core technical component of DataPACT, a European-level research and innovation project. Within this context, the dissertation focuses on recognising emotions using Electroencephalogram (EEG) signals captured by a Brain-Computer Interface (BCI) while the user is consuming audiovisual content.

1.1 Contextualization

In recent years, the consumption of audiovisual content, such as films and series, has become a central part of the entertainment industry, fuelling interest in technologies capable of understanding the emotion of the audience (J. Z. Wang et al. 2023). At the same time, the interaction between people and technology has evolved significantly, driven by advances in BCIs. These interfaces, especially when based on EEG signals, have emerged as promising tools for monitoring emotional states in real time, allowing direct analysis of brain activity and the emotional impact of different stimuli. BCIs allow direct communication between the human brain and external devices, using brain signals captured by non-invasive methods. EEG signals, in particular, stand out due to their ability to measure the brain's electrical activity with high temporal resolution via sensors (electrodes) placed on the head (Berger 1929; Bhise et al. 2020). This non-invasive method has opened the door to a wide range of practical applications, including the development of emotion recognition systems based on brain signals (Al-Nafjan et al. 2017).

Emotion recognition or sentiment analysis plays a crucial role in understanding people's responses to external stimuli in fields such as mental health, marketing, gaming, and entertainment. Traditional methods, such as subjective questionnaires and facial expression analysis, have been widely used to identify emotional states. However, these approaches present significant limitations, including dependence on participant's self-awareness and susceptibility to falsification or biased and incorrect interpretations (Houssein et al. 2022). On the other hand, the use of EEG signals offers a more objective and real-time approach to identifying emotions by capturing brain activity directly associated with external stimuli (Huang et al. 2019), such as audiovisual content (films, series, etc.).

In the context of audiovisual content consumption, the ability to interpret users' emotional states can provide valuable insights into the viewing experience. This approach paves the way for systems that are capable of analysing the audience's emotional responses, offering a deeper understanding of their reactions during the visualisation of films or videos (Houssein et al. 2022). EEG signals contain critical information distributed across various frequency bands (*delta, theta, alpha, beta,* and *gamma*), each associated with different cognitive and emotional activities (Kalaganis et al. 2017). For example, high levels of activity in the *gamma* band are often associated with states of attention and focus, while the *alpha* band can reflect relaxation (Abhang et al. 2016b).

Despite technological advances, the interpretation of EEG signals in practical contexts, such as media consumption, remains a challenge. Individual variability (including participant's unique and own interpretations), data noise and difficulties in generalising the applied Artificial Intelligence (AI) models are obstacles that still need to be overcome (Saha et al. 2021). Furthermore, the inherent complexity of people's emotional states requires robust and innovative approaches, combining signal processing techniques, Machine Learning (ML) and Deep Learning (DL) techniques to achieve reliable results (Hamzah and Abdalla 2024). These issues reinforce the importance of investigating not only the relevant parameters of EEG signals but also the most effective methods for their analysis, validation in real-world scenarios and analysing the associated ethical and technical challenges.

This dissertation not only fits within the academic context of the Master's in Artificial Intelligence Engineering, but is also integrated into DataPACT which is a European research project (Grant Agreement ID 101189771, 2025-2027) carried out in partnership with EVS Broadcast Equipment. With a total budget of approximately €11.18 million (of which nearly €10 million is EU-funded), DataPACT gathers a balanced consortium of 18 partners from 16 countries, working together to embed compliance by design into data/AI operations and pipelines. To this end, the project delivers a compliance toolbox, a compliance framework, and specialized pipeline management tools. The work carried out here constitutes the central technical component of DataPACT in the media domain, focusing on EEG-based emotion recognition, a role that highlights both its scientific relevance and its potential real-world impact in the broadcasting and entertainment industry, where it will serve as a use case among the project's seven validation tracks.

1.2 Problem Statement

With the growth of the digital media industry and the expansion of streaming platforms, understanding users' emotions has become a strategic goal for companies in the field, such as EVS Broadcast Equipment, the company that prompted this dissertation. This growth, combined with the technological evolution of wearable devices and the popularisation of BCIs, opens up new possibilities for understanding and adapting user experiences based on their emotional states, making it possible, for example, to personalise streaming experiences by adapting content to each user's emotional preferences. Another promising scenario for the application of emotion analysis systems is in film testing sessions. These sessions allow directors and producers to assess, in real time, whether the intended emotions are actually being felt by the audience. However, traditional methods of emotion analysis, such as questionnaires and facial recognition, reveal significant limitations, particularly in terms of objectivity, reliability and the ability to capture emotions in real time.

Despite the potential, there are several challenges that must be addressed. These include, for example, the variability of EEG signals between users, the noise of these signals and the inherent complexity of emotional states, which are often combinations of multiple emotions. In addition, ethical considerations related to the use of brain signal data must also be addressed.

In this context, this dissertation aims to develop an innovative system to explore the capabilities of BCIs to interpret EEG signals and identify users' emotional states during the consumption of audiovisual content, allowing this system to be used in various specific areas, such as content personalisation and film testing sessions.

1.3 Research Questions and Objectives

To ensure a structured approach to this dissertation, a set of Research Questions (RQ) was defined to guide the investigation and its methodological framework. Three main RQs were established, complemented by specific sub-questions that address key aspects of the problem in greater detail. The RQs are as follows:

- **RQ1**: How can EEG signals captured by BCIs be used to identify human emotions and develop reliable models for emotion recognition during audiovisual content consumption?
  - **RQ1.1**: How can EEG-based systems be tested and evaluated in a practical and objective way to guarantee their effectiveness in identifying emotions in the media context?
- **RQ2**: What are the existing approaches and challenges in emotion recognition from EEG signals captured by BCIs?
  - **RQ2.1**: What AI models are used for emotion recognition using EEG signals?
  - **RQ2.2**: Which features of EEG signals can be most effective in identifying/recognising emotions?
  - **RQ2.3**: How can the use of EEG signals overcome the limitations of traditional emotion recognition methods, such as facial recognition or questionnaires?
- **RQ3**: What are the main technical and ethical challenges associated with using EEGs for emotion recognition in the consumption of audiovisual content?

The objectives of this dissertation were carefully formulated to define the scope of the research and provide clear milestones for its development. These objectives provide structure to the investigation and ensure that the most important steps are addressed. The formulated objectives are as follows:

- **O1**: Compare the approach based on BCIs and EEG signals with traditional methods of analysing emotions, such as subjective questionnaires and facial recognition, assessing how BCIs can overcome the limitations of these techniques.
- **O2**: Analyse the most relevant approaches for signal pre-processing, feature extraction, and deep learning models in EEG-based emotion recognition.
- **O3**: To explore how EEG signals, captured by BCIs, can be used to identify emotional states and to investigate the most relevant characteristics and features to the study.

- **O4**: Identify the main technical and ethical challenges related to the use of BCIs for emotion recognition, including privacy, data security, individual variability and generalisation of AI models.
- **O5**: Analysing and pre-processing the datasets relevant to the study, including SEED datasets, in order to understand and prepare the data for emotion recognition.
- **O6**: Evaluate the impact of different methodological decisions (types of evaluation, normalisation strategies, combinations of features and model architectures) on the robustness and generalisation capacity of the models developed.
- **O7**: Develop a functional emotion recognition system based on EEG signals, applying AI, signal processing, deep learning techniques to identify emotional states.
- **O8** Test and validate the system developed in order to evaluate it in terms of generalisation and accuracy in detecting emotions.

1.4 Scientific Contributions

In addition to the technical work carried out throughout this dissertation, an independent scientific contribution was produced that has a direct impact on the academic and professional community. This contribution took the form of an article entitled "Building Trust in EEG-Based Emotion Recognition: Legal, Ethical, and Technical Foundations", presented as a result of this research work and has been submited at the journal "IEEE Transactions on Affective Computing".

The article focuses on analysing the ethical, legal and social challenges associated with the use of artificial intelligence in emotion recognition and the use of EEG signals, framing the technology in the context of the General Data Protection Regulation (GDPR) and the AI Act, recently approved by the European Union. Alongside the discussion of the regulatory framework, the article presents a concrete proposal: a technical-ethical guide for the collection and processing of EEG data, which summarises practical recommendations to ensure legal compliance and the protection of participants' privacy. Among these recommendations are the need for informed consent, the application of anonymisation or pseudonymisation techniques, the clear definition of data use purposes, and the minimisation of risks associated with the processing of neural information.

This contribution is particularly relevant because it is not only a conceptual reflection but also a practical tool that can guide future EEG data collection. By combining ethical and regulatory analysis with technical expertise gained in processing and modelling SEED datasets, the article provides a bridge between legal principles and research practices.

It is also important to note that the technical and ethical guidelines developed in this article, together with the reflection on legal and ethical challenges, will constitute an essential reference for future work within the scope of DataPACT. In addition, the results and findings achieved throughout this dissertation, namely the pre-processing methodologies, comparisons of features and frequency bands, and critical analysis of evaluation strategies, will also be used as a basis for further research, including a doctoral project already outlined, reinforcing the scientific, practical and strategic relevance of this work.

1.5 Document Organization

This document is structured into five main chapters:

Chapter 1 - Introduction provides the context for the topic, defines the research problem, formulates the research questions and objectives, identifies the main scientific contributions during the course of this project, and finally, the document organisation.

Chapter 2 - State of the Art presents the main concepts and methodologies related to emotion recognition from EEG signals and brain-computer interfaces. The most relevant emotion models, the most commonly used measurement methods, advances in signal processing techniques and deep learning, as well as existing scientific and commercial solutions are discussed.

Chapter 3 - Methods, Tools and Experimentation describes the tools used, data collection and preparation procedures, and methodological details of the experimental tests performed. It also includes an analysis of ethical and legal issues related to the use of EEG data.

Chapter 4 - Implementation, Analysis and Discussion of Results presents the analysis and processing of data, feature extraction methods, experimental setup and tests performed. The results obtained are discussed, different methodological approaches are compared and the main challenges encountered are identified.

Finally, **Chapter 5** - Conclusion summarises the main contributions of the work, discussing its scientific and practical implications, and identifies lines of future research that can be explored based on the results obtained.

Chapter 2

State of the art

This chapter will present the main concepts and methodologies related to emotion recognition based on Electroencephalogram (EEG) signals and Brain-Computer Interface (BCI) technologies, as well as their main inherent challenges. In addition, the main existing scientific and commercial solutions will be explored. Although this work does not follow a systematic review, a literature search was conducted with the aim of identifying relevant articles in the areas of affective computing, emotion recognition and BCI applications.

Recognised databases such as IEEE, ScienceDirect, PubMed, ACM and Web of Science were used, with specific keywords related to the topics in question as the research progressed. The selection of articles was based on their relevance to the topics covered, the credibility of the sources, their scientific impact (measured by the number of citations and year of publication) and whether they were peer reviewed.

2.1 Affective Computing

Affective computing serves an umbrella term used to describe most techniques that integrate emotion and Artificial Intelligence (AI) (Assunção et al. 2022). Affective computing is a multidisciplinary field of research dedicated to developing computer systems capable of recognising, interpreting, processing and responding to human emotions (Y. Wang et al. 2022). This field, which crosses AI, psychology, neuroscience and engineering, aims to create more natural and intuitive Human-Computer Interfaces (HCI), where the technology is able to detect the user's emotional state and adapt intelligently to their needs and intentions (Houssein et al. 2022). In this section, the definition of what an emotion is from a scientific perspective is presented, along with how emotions are measured and the main existing models of emotion.

2.1.1 Definition and importance of emotions

From a scientific perspective, emotions are understood as individual experiences and reactions, influenced by the conditions of the social context. Emotions are an essential part of every human being, with the power to influence behaviour, reasoning ability, decision-making, resilience, well-being and the way humans communicate (Yadegaridehkordi et al. 2019). They are also recognised as a complex by-product of sentient life, whose objective definition and comprehension remain challenging. In other words, emotions are viewed as individual processes that manifest themselves through both subjective psychological changes and objective neurophysiological alterations, and where emotional experience involves a complex interaction between psychology and physiology, where changes in physiological signals reflect changes in emotions (Assunção et al. 2022). Since emotions affect both the physiological and psychological states of individuals, they play a a very important role in human life where positive emotions help to improve human health, efficiency at work and even relationships with other people while negative emotions can cause health problems and may disrupt social life (Shu et al. 2018a).

2.1.2 Emotion measurement methods

Measuring emotions involves analysing different methods. The measurement of emotions is a crucial field in human-computer interaction research and in various areas of the social and health sciences (Shoumy et al. 2020). The ability to detect and interpret human emotions plays an important role in various applications, ranging from the development of more intuitive interfaces to supporting mental health. Methods for measuring emotions vary in terms of complexity and accuracy, with each approach offering distinct benefits and limitations (Egger et al. 2019). Although emotions are transmitted across all modalities of human communication, their analysis is typically divided into two main areas: the analysis of behavioural signals and the analysis of physiological signals. Behavioural signals include facial expressions, body language, and voice intonation, while physiological signals encompass parameters such as electrodermal activity, heart rate, respiration, and brain activity measured through EEG. Affective measurement channels can be classified into five main categories: visual, textual, vocal, physiological, and multimodal (Gunda et al. 2024):

Visual Channel

This channel mainly includes facial expressions, though it also includes Body Gesture (Y. Wang et al. 2022). Analysing facial expressions is one of the most common methods for recognising emotions, involving the capture and analysis of images or videos of the face (Egger et al. 2019). The Facial Action Coding System (FACS), which example is in Figure 2.1, is a famous tool designed to encode facial movements into action units, allowing emotions to be identified based on these movements (Yadegaridehkordi et al. 2019). This method/channel is often used in human-computer interaction environments and can be implemented using video cameras, without requiring any sensors to be attached to the user (Egger et al. 2019). However, it is important to note that facial expressions can be masked, simulated and even manipulated, which can lead to errors in detecting the correct emotions (Torres et al. 2020).

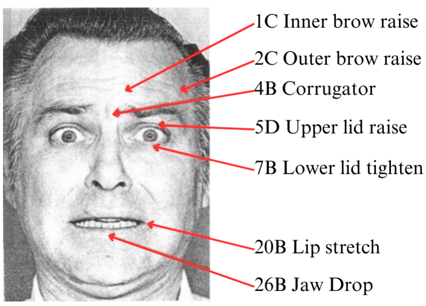

Figure 2.1: Example FACS codes for a prototypical expression of fear (Lainscsek 2006).

Textual Channel

This channel mainly includes questionnaires and text analysis. With the rapid increase in online social networks and e-commerce platforms, where users freely express their ideas, a huge amount of textual data is generated and collected, often already with labels corresponding to the type of emotion they are feeling (e.g., positive or negative in evaluating a product) (Y. Wang et al. 2022). Figure 2.2 is an example of Emotion Recognition from Text.

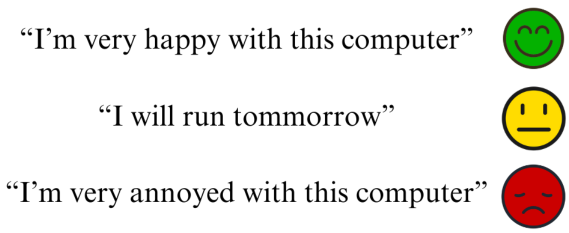

Figure 2.2: Example of Emotion Recognition from Text.

Vocal/Auditory Channel

This channel shares some similarities with the textual channel, focusing on the analysis of speech and discourse, by examining spoken words and phrases that have been spoken. It has been significantly boosted by the growth of social media platforms, which gave wings to various sources such as podcasts, audio messages, videos, but also from music, television, radio and more. The key difference is that the vocal channel also includes features such as intensity, loudness and pitch of the speech signal in terms of amplitude and frequency (Shoumy et al. 2020). Although it is possible to interpret and detect emotions from people's voices, much like facial expressions, the voice can also be manipulated and simulated, which can lead to errors in emotion detection (Houssein et al. 2022).

Physiological Channel

This channel includes data such as Respiration (RSP), blood pressure, EEG, Electrooculography (EOG), Electrocardiography (ECG), Blood Volume Pulse (BVP), Heart Rate Variability (HRV), Galvanic Skin Response (GSR) and even finger temperature (Egger et al. 2019). The physiological channel is capable of providing more accurate and objective emotion recognition (Houssein et al. 2022). Although it is more challenging to obtain data due to the need to attach sensors to people, compared to the previous channels, physiological signals are not affected by "social masking" (e.g., an average person cannot consciously control their heart rate) and, as a result, they are significantly harder to manipulate and are considered more objective and reliable than the previous modalities (Y. Wang et al. 2022). Figure 2.3 illustrates some physiological methods.

Multimodal Channel

This channel integrates two or more of the previous channels to obtain a richer and more comprehensive representation of the emotional state combining, for example, physiological signals with visual signals (Yadegaridehkordi et al. 2019). The combination of multiple methods can improve the accuracy in emotion detection, as it has the potential to overcome the limitations of each method when used in isolation (Egger et al. 2019). In the

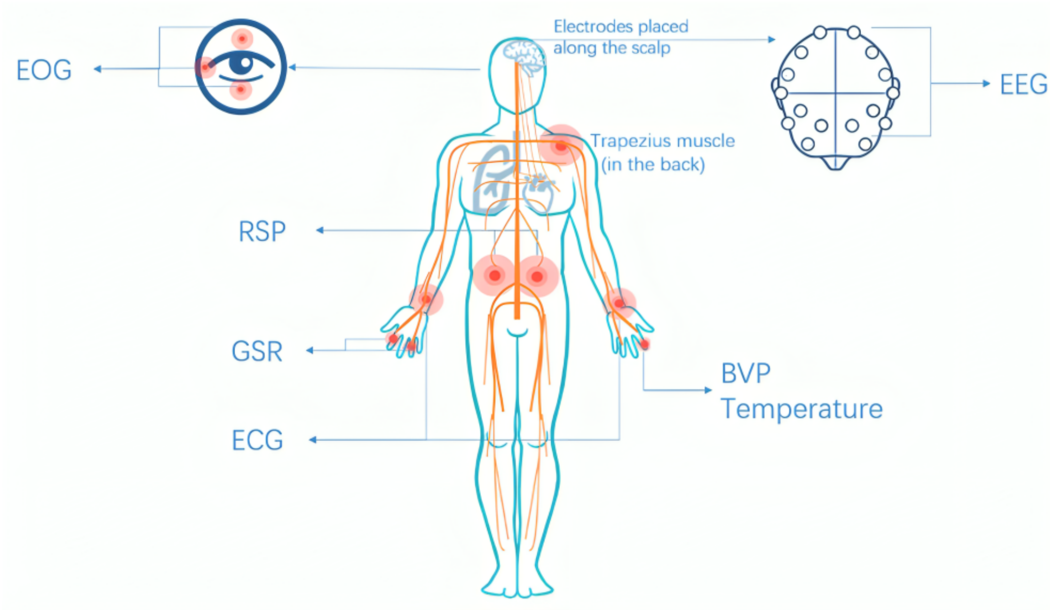

FIGURE 2.3: Example Physiological Sensors (Shu et al. 2018b).

field of affective/sentiment analysis within the big data domain, information such as service and product reviews, social media content, etc., are gradually shifting from single modality (uni-modal) to multimodal in order to capitalise on the advantages of each of the channels (Shoumy et al. 2020).

2.1.3 The main models of emotions

Emotion models are theoretical approaches that aim to describe, classify, and understand the complex nature of human emotions and although psychologists attempt to classify emotions in different ways in the multidisciplinary fields of neuroscience, philosophy, and computer science, there are no unanimously accepted emotion models (Y. Wang et al. 2022). However, there are two general types of emotion models in affective computing namely, the the discrete (or categorical) emotion model and the dimensional (or continuous) emotion model (Erat et al. 2024).

The **Discrete models of emotion** identifies basic, innate, and universal emotions from which all other emotions can be derived (Torres et al. 2020):

Paul Ekman model of six emotions

Ekman describes discrete emotions as a group of universal or in other words as being commonly recognised in different cultures and identifies six basic emotions namely anger, fear, sadness, joy, surprise, and disgust (Erat et al. 2024). Ekman also studied the facial expressions associated with these emotions, concluding that each one has a specific pattern of muscle movements in the face, where, for example, happiness is identified through raising of the mouth corners and tightening of the eyelids. Ekman later expanded this set of emotions, adding twelve new positive and negative emotions, including amusement, contentment, embarrassment, excitement, guilt, pride, relief, satisfaction, pleasure and shame (Shoumy et al. 2020). Nevertheless, his model of six basic emotions, visible in Figure 2.4, remains the most widely used and recognised.

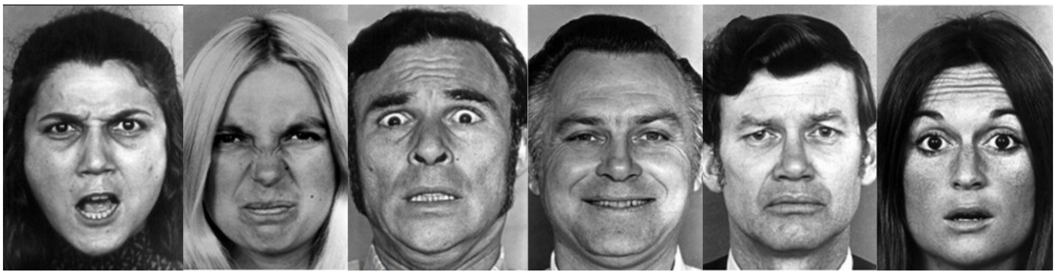

FIGURE 2.4: Six Emotions of Ekman's model (Anger, Disgust, Fear, Happy, Sad, Surprise) (Ekman and Friesen 1976).

Plutchik's wheel of emotions

Plutchik's wheel model involves eight basic emotions: joy, trust, fear, surprise, sadness, anticipation, anger, and disgust and the way of how these are related to one another, visible in Figure 2.5. For example, joy and sadness are opposites, and anticipation can easily develop into vigilance. This wheel model is also referred to as the componential model, where the stronger emotions occupy the centre, while the weaker emotions occupy the extremes, depending on their relative intensity levels (Y. Wang et al. 2022). Despite being a discrete model with 8 separate emotions, the Wheel of Emotions also has dimensional details in how emotions are represented in terms of intensity and relational proximity in which there are common patterns in each emotional state, which is why some authors, such as (Shoumy et al. 2020), consider it a dimensional model.

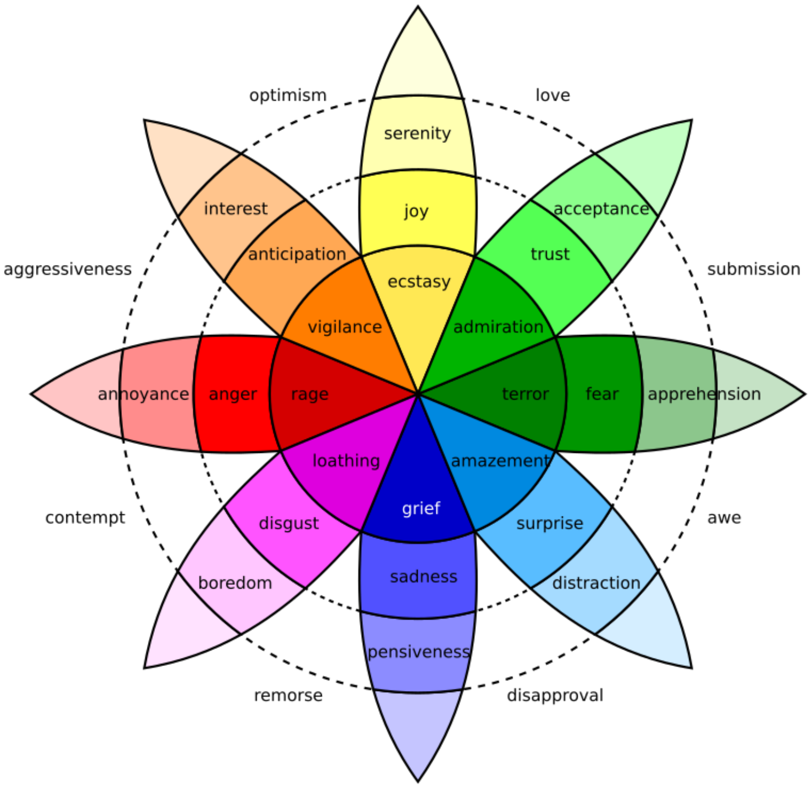

FIGURE 2.5: Robert Plutchik's Wheel of Emotions (Elf 2011).

Ortony, Clore and Collins (OCC)

The OCC model is one of the most widely used categorical models in studies related to affective computing (Yadegaridehkordi et al. 2019). Instead of limiting itself to a restricted number of basic emotions, the model specifies 22 emotions that are based on valence reactions to situations (Bartneck 2002). Emotions are then categorised hierarchically, initially situations are divided into 3 branches, each one is the result of different types of cognitive interpretations (Zhaoxia Wang et al. 2020).

- Consequences of events: These emotions result from the evaluation of whether an event is desirable or undesirable. Examples include joy (when an event has a positive outcome) and sadness (when an event has a negative outcome) and even anticipatory emotions such as hope and fear for future events (Zhaoxia Wang et al. 2020). Emotions in this category can be directed towards the outcomes of oneself or others (Clore and Ortony 2013).
- Actions of Agents: These emotions are related to the approval or disapproval of agents' actions, including the actions themselves in which, for example, one can feel pride or shame (among others) (Clore and Ortony 2013).
- Aspects of Objects: These emotions focus on the characteristics of objects, people or ideas, assessing whether they are appealing or unappealing (Clore and Ortony 2013).

Figure 2.7 shows the original structure of the OCC model, where the origin of the 22 emotions can be found. Later, the authors (Steunebrink et al. 2009) found some ambiguities and discussed them, resulting in a total of 24 emotions in their model (OCC Revisited).

Dimensional models of emotion define emotions on two (Valence-arousal) or three dimensional planes (Valence, arousal, and dominance) (Erat et al. 2024). Unlike discrete models, dimensional models describe emotions as points in a multidimensional space, where each dimension represents a different emotion (Islam et al. 2021).

Russell Circumplex Model

Russell proposed a circumplex model based on Valence-Arousal to represent complex emotions. This model defines a continuous, two-dimensional emotional space with the Valence and Arousal as its axes (Y. Wang et al. 2022). Feelings of pleasant to unpleasant emotions are represented on a valence scale, while feelings of activity and inactivity are represented on an arousal scale (Erat et al. 2024). For example, as shown in figure 2.6, anger has a high level of arousal, meaning it is very active, while it has low valence, making it a unpleasant emotion.

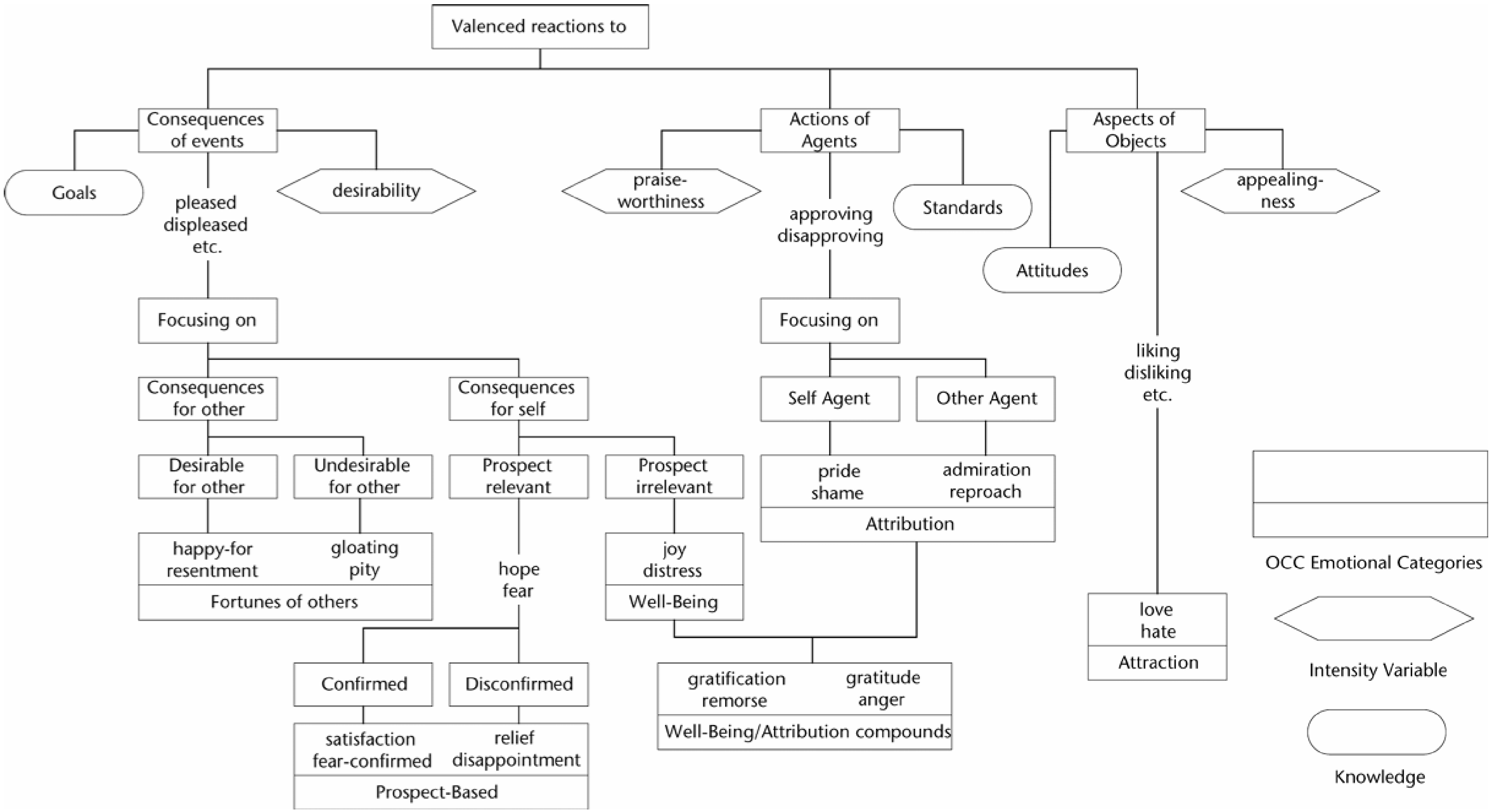

Figure 2.7: The Original OCC Model (Bartneck 2002).

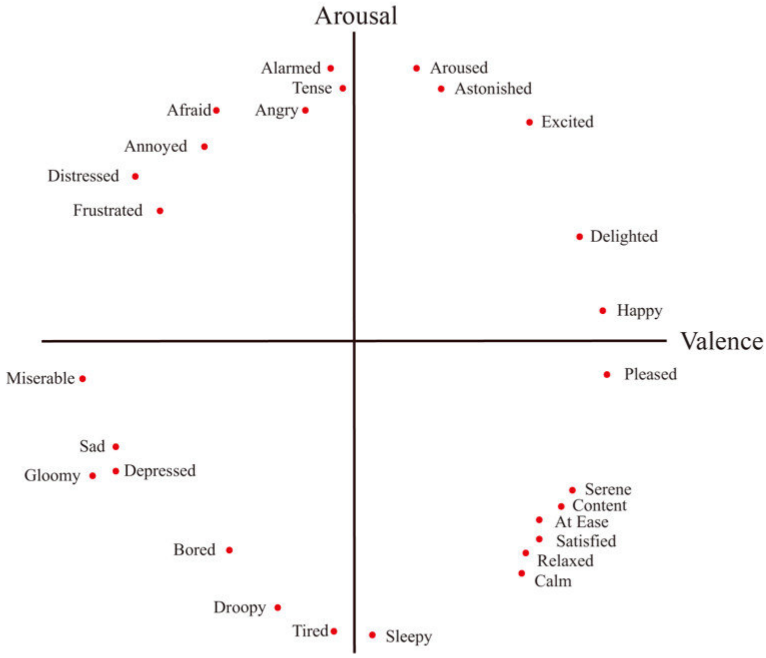

FIGURE 2.6: Russel Circumplex Model (Seo and Huh 2019).

Pleasure-Arousal-Dominance (PAD) Model

One of the three-dimensional emotion models is the PAD Emotional State Model (Zhaoxia Wang et al. 2020). The added dimension axis, dominance, ranges from submissive to dominant and reflects an individual's perceived control in a given emotional state. In this dimension, anger and fear can be easily identified as anger is in the dominant axis while fear is in the submissive axis (Shu et al. 2018a). The PAD model is useful for describing more complex emotional states and subtle, such as exuberance and bored which include all three axes of the PAD scale, and can even predict personality traits (Shoumy et al. 2020). A visual representation of this model is shown in Figure 2.8.

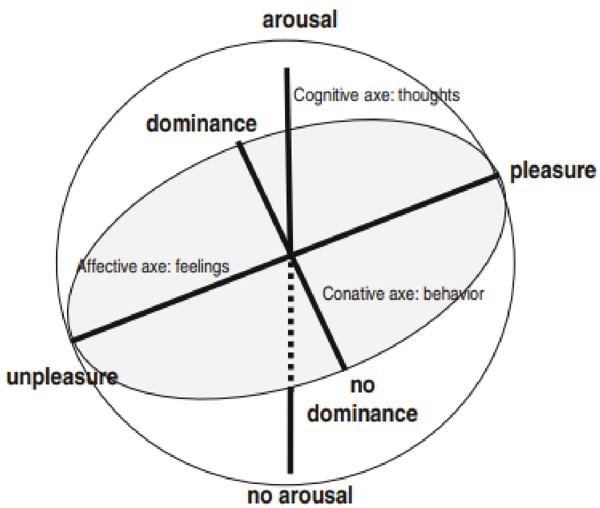

FIGURE 2.8: Three Dimensional Model (Bakker et al. 2014).

2.2 Emotion Recognition

Emotion Recognition is an area of affective computing and is the process of identifying and interpreting an individual's emotional state (Islam et al. 2021). Progress in emotion recognition has significantly advanced over the years, driven by the advancement of BCI systems and with the contribution of many disciplinary fields like psychology, neuroscience, medicine, and even computer science (Abhang et al. 2016d).

The ability to successfully recognise emotions is essential for improving the interaction between people and machines, allowing computer systems to understand not only the context, but also the emotional state of users (Egger et al. 2019). Throughout this section, fundamental topics for emotion recognition will be explored, including EEG and its characteristics, BCIs, other approaches such as analysing facial expressions, and signal processing and Deep Learning (DL) methods.

2.2.1 Electroencephalogram (EEG)

EEG is a non-invasive neurophysiological technique that measures the electrical voltage fluctuations resulting from neuronal activity in the brain (Islam et al. 2021). Instead of recording the activity of a single neuron, EEG reflects the combined activity of millions of neurons distributed over a few centimetres of cortical tissue (Houssein et al. 2022). However, the signals captured are extremely weak and susceptible to noise, which makes it necessary to use multiple electrodes to improve spatial resolution and increase measurement accuracy (Erat et al. 2024). To ensure reproducible setups, standardised systems for positioning electrodes on the skull have been established, the most widely used being the international 10/20 (IS) system (Alarcão and Fonseca 2019).

Before addressing this system in more detail, it is important to understand the organisation of the cerebral cortex, which is divided into four main lobes: frontal, temporal, parietal and occipital, as illustrated in Figure 2.9. The frontal lobe, located in the anterior part of the brain, is involved in higher cognitive functions such as planning, reasoning and emotional regulation, while the parietal lobe, located behind the central region of the brain, integrates sensory information and processes spatial stimuli, including pain, pressure, temperature, taste and touch, as well as playing an important role in logical and mathematical thinking. The temporal lobe, located laterally, is essential for auditory and olfactory processing and is also fundamental in memory and emotional regulation, and the occipital lobe, located in the posterior region of the brain, is responsible for processing visual information and contributes to memory and abstract thinking (Houssein et al. 2022).

Logically, we can relate each of the lobes to its involvement in emotional processing. The frontal lobe, as mentioned before, is central to emotional regulation and recognising emotions. The parietal lobe integrates sensory information that can influence emotional states, while the temporal lobe is associated with emotional memories and the processing of emotional stimuli. Finally, the occipital lobe, although less directly associated with emotions, may be relevant in the processing of emotional visual stimuli. It should be noted that there are two main areas of the brain correlated with emotional activity: the amygdala (located close to the hippocampus, in the frontal portion of the temporal lobe); and the prefrontal cortex (covers part of the frontal lobe) (Alarcão and Fonseca 2019).

With this foundation established, we can now turn to the international 10/20 system, shown in Figure 2.10, which is the most widely used standard for placing electrodes

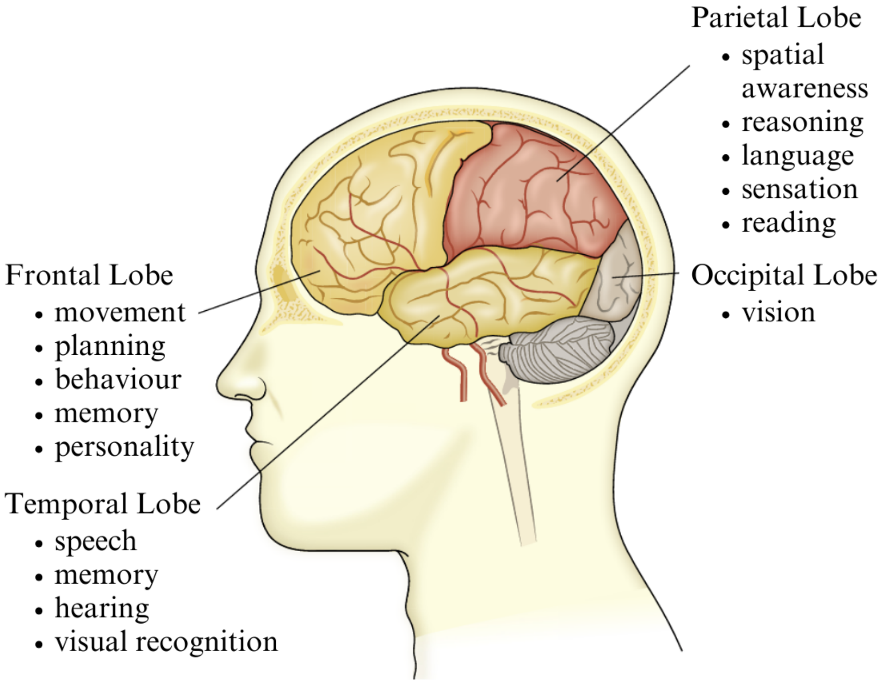

Figure 2.9: The four main lobes of the brain (Abhang et al. 2016a).

on the skull and which was standardised by the International Federation of Societies for Electroencephalography and Clinical Neurophysiology. In this system, the electrode positions are identified by letters and numbers, where each letter corresponds to a cerebral lobe: F for frontal, T for temporal, P for parietal, O for occipital and the letter C is used to designate electrodes in the central region, although there is no central lobe. The numbers indicate the relative position of the electrodes in each hemisphere, with the even ones assigned to the right hemisphere and the odd ones to the left hemisphere, while the letter Z identifies the electrodes positioned along the midline of the skull (Erat et al. 2024).

EEG stands out as a simple, portable and accessible method for identifying emotions in various areas, including entertainment, e-learning, virtual worlds and e-healthcare applications, and its applications range from instant communication and online games to supporting therapists and psychologists in the performance of their duties (Alarcão and Fonseca 2019). In addition, EEG is widely used in the diagnosis of various neurological conditions, such as epilepsy, brain inflammation, traumatic injuries, brain tumours, memory deficits, strokes and sleep disorders, among others (Abhang et al. 2016a).

Brain Waves

As previously mentioned, the EEG captures the electrical activity of the brain through electrodes on the head, revealing neural oscillations that reflect neuronal communication and are linked to emotional and cognitive states, and emotions, far from being mere subjective states, are the product of a complex interaction between various regions of the brain, each with its own pattern of electrical activity.

These neural oscillations manifest themselves in different frequency bands and their study offers a detailed view of the brain processes behind emotions and their analysis makes it possible to identify the brain activity associated with different emotional experiences

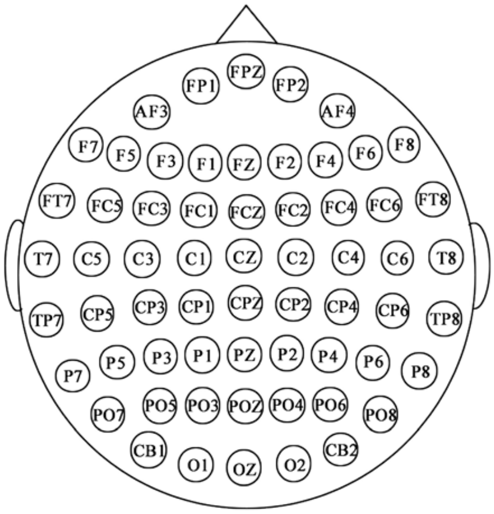

FIGURE 2.10: The international 10/20 system with 62 channels (W.-L. Zheng and Lu 2015).

(Abhang et al. 2016c). The EEG frequency bands are divided into *delta, theta, alpha, beta* and *gamma*, each with different frequency ranges and functions (Torres et al. 2020).

Frequency bands and their relation to emotions:

- ***Delta* band (1-4 Hz):** This is the lowest frequency band and is associated with deep sleep, states of unconsciousness and anaesthesia (Houssein et al. 2022). In terms of emotions, although less relevant for recognising emotions, it can be linked to states of extreme relaxation, recovery and deep states of meditation (Erat et al. 2024). In addition, *delta* is prominently seen in brain injuries, learning problems, inability to think as if sleepy, and severe Attention Deficit Hyperactivity Disorder (ADHD) (Abhang et al. 2016c).

- ***Theta* band (4-8 Hz):** *Theta* waves are observed during sleep, dreams, meditation and cognitive processes such as learning and memory (Erat et al. 2024). In terms of emotions, studies have shown that midline frontal *theta* activity increases with the perception of positive stimuli, such as pleasant music (W.-L. Zheng and Lu 2015). In addition, increases in *theta* power have been reported in response to emotionally intense stimuli in both frontal and parietal regions. These results suggest that the *theta* band is involved in the experience and processing of emotions, especially those related to valence and arousal (Mühl et al. 2014).

- ***Alpha* band (8-12 Hz):** This band is associated with states of relaxation and calm and is considered a marker of emotional regulation (Islam et al. 2021). Increased *alpha* activity can indicate reduced stress and states of moderate concentration, while its decrease can be observed in moments of anxiety and emotional tension (Houssein et al. 2022). This means that *alpha* is often associated with neutral emotions since, when presented with neutral videos, subjects are prone to relaxing and paying less attention, which leads to high *alpha* responses (W.-L. Zheng and Lu 2015).

- ***Beta band* (13-30 Hz):** This band is related to states of alertness, attention and active cognitive processing and tends to increase in moments of excitement or anxiety, and is often associated with the stress response and decision-making processes (Abhang et al. 2016c). In terms of emotions, for positive emotions such as happiness, the energy of *beta* increases, with greater activation. For negative emotions, *beta* has a lower energy (Jiang et al. 2024).

- ***Gamma* band (30-100 Hz):** This is the highest frequency band and is associated with complex cognitive processes such as perception, memory, consciousness and learning (Abhang et al. 2016c). In terms of emotions, similarly to the *beta* band, for positive emotions there is an increase in energy while for negative emotions it has a lower energy and *gamma* is considered the most discriminating band between the different emotions in Paul Ekman's model (Section 2.1.3) (Jiang et al. 2024).

The analysis of frequency bands has been widely used by several studies in EEG-based emotion recognition, since different emotional states are associated with specific patterns of activity in these frequencies. For example, (Jiang et al. 2024) indicates that the happy emotion exhibited greater activation levels in temporal areas in the *beta* and *gamma* bands, the surprise emotion yielded very low DE features in all bands, the neutral emotion produced strong *alpha* responses, the temporal areas of the sad and angry emotions in the *gamma* band were lower than those of other emotions, the occipital area of the disgust emotion was the lowest among those of all emotions, and the frontal area of the fear emotion displayed high activation.

In general, the segmentation of EEG frequencies into bands makes it possible to better understand the mechanisms underlying emotional and cognitive processing, and is an essential element in the research and development of EEG-based systems for emotion recognition and other applications.

Features of EEG signals for emotion recognition

Emotion recognition based on EEG signals depends on the extraction of features that capture relevant information about brain activity. These features transform the raw signals into more compact and discriminative representations, facilitating the classification of emotional states using Machine Learning (ML) and DL techniques. Extraction can be carried out in different domains, including the time domain, frequency domain, time-frequency, spatial information and non-linearity, each contributing different perspectives on the neuronal patterns associated with emotions.

Time Domain Features in the time domain are extracted directly from the EEG signal, without transformations to other domains, reflecting variations in the amplitude and morphology of the signal over time (Hamzah and Abdalla 2024). These features include statistical measures such as mean, variance, kurtosis and skewness (Torres et al. 2020).

Other relevant features include the Hjorth parameters , which consist of three statistical metrics (Islam et al. 2021):

- Activity - Measures signal strength.
- Mobility - Measures the mean frequency of the signal.
- Complexity - Indicates variations in frequency over time.

Furthermore, Event-Related Potentials (ERP) are widely used in cognitive and emotional analysis. These EEG signal variations occur in response to specific stimuli, reflecting neuronal processes underlying emotional processing (Alarcão and Fonseca 2019).

Frequency domain Frequency domain analysis makes it possible to identify which frequency bands (*delta, theta, alpha, beta* and *gamma*) are most active at certain times. Given that different frequencies are associated with different brain functions, as analysed in the previous section, this type of analysis is essential for capturing emotional and cognitive states.

In addition to the frequency bands themselves, two of the main features in this field include:

- Power Spectral Density (PSD) - measures the power distribution of the EEG signal at different frequencies and is one of the most widely used metrics (Islam et al. 2021).
- Differential Entropy (DE) - measures the complexity of the EEG signal within each frequency band (W.-L. Zheng and Lu 2015). Studies such as (Duan et al. 2013) and (W. Zheng et al. 2018) show that DE is highly discriminative for recognising emotions, making it possible to distinguish EEG patterns associated with different emotional states.

Time-Frequency Domain The EEG is a non-stationary signal, which means that its characteristics vary over time. The time-frequency domain combines information from the time and frequency domains, making it possible to capture temporal dynamics of spectral variations (Hamzah and Abdalla 2024).

- Wavelet Transform (WT) - Uses variable windows, allowing more accurate analysis of non-stationary signals. WT offers high temporal accuracy for high frequencies and high spectral accuracy for lower frequencies (Houssein et al. 2022).
- Another example of a feature in this domain is the Short-Time Fourier Transform (STFT), which will be discussed later in this section.

Spatial Information The spatial distribution of EEG electrodes provides information on the origin and propagation of brain oscillations and this domain is essential for identifying spatial patterns in the brain during emotional states. Features such as frequency bands provide some spatial information, since each one is associated with an area of the brain (Mühl et al. 2014). Still, a concrete example of one of the main spatial features is:

- Common Spatial Patterns (CSP) - Identifies discriminative spatial patterns between different brain states and is widely used in BCI's applications, projecting multichannel signals into a subspace where differences between classes are maximised and similarities minimised (Hramov et al. 2021).

Features Based on Connectivity Emotions significantly influence neural connections, making the analysis of functional connectivity between brain regions a relevant approach for emotion recognition (Torres et al. 2020). A common example is:

- Phase Locking Value (PLV) - Measures the phase synchronisation between two EEG signals, assessing whether two brain regions maintain a constant phase relationship over time. PLV makes it possible to separate the phase component from the amplitude of the EEG signal, making it easier to analyse neuronal interactions (Zhongmin Wang et al. 2019).

Non-Linear Features The EEG is highly non-linear in nature, which means that traditional methods may not capture all the relevant information. Non-linear features are useful for identifying complex patterns in brain activity (Hamzah and Abdalla 2024). Examples of features include:

- Sample Entropy (SE) - Measures the randomness or irregularity of the signal over time, assessing the predictability of subsequences (Rahman et al. 2022).
- Permutation Entropy (PE) - Represents the degree of irregularity of the EEG signal, and is suitable for analysing non-linear and non-stationary signals (Rahman et al. 2022).
- Higuchi's Fractal Dimension - Measures the complexity and irregularity of temporal signals, and is a relevant metric for EEG (Rahman et al. 2022).

Feature Extraction Methods Extraction methods transform EEG signals into usable representations for analysis. Some of the most commonly used include:

- Discrete Fourier Transform (DFT) - Converts discrete signals from the time domain to the frequency domain (Islam et al. 2021).
- Fast Fourier Transform (FFT) - An efficient algorithm for calculating the DFT, significantly reducing processing time (Islam et al. 2021). The FFT is widely used to calculate the PSD of frequency bands but has limitations, such as low-frequency resolution and high spectral loss of information (Torres et al. 2020).
- Short-Time Fourier Transform (STFT) - Analyses spectral variation over time, resolving a limitation of the FFT by allowing time-frequency analysis. The STFT divides the signal into short windows and applies the FFT to each one, creating a two-dimensional representation (time x frequency) (Hamzah and Abdalla 2024).
- Discrete Cosine Transform (DCT) - Converts time domain signals into basic frequency components (Islam et al. 2021).

2.2.2 Brain-Computer Interface (BCI)

BCIs are systems that establish a direct communication link between the brain and an external device This connection allows brain signals to be translated into commands that make it possible to control machines, recover motor functions or facilitate communication (Shih et al. 2012). Although traditionally associated with device control and motor rehabilitation, BCIs have been increasingly exploited to recognise emotions and cognitive states (Torres et al. 2020).

Essentially, BCIs capture and interpret EEG signals, converting them into instructions for an output device, without relying on the traditional pathways of peripheral nerves and muscles (Shih et al. 2012). In this context, the concept of affective BCIs emerged, which focuses on detecting neuronal patterns associated with different emotions, enabling advances in emotional recognition in applications such as entertainment and human-machine interactions (Mühl et al. 2014).

The functioning of a BCI can be broken down into a few essential phases: signal acquisition, pre-processing, feature extraction, classification and evaluation. The EEG signal is acquired using electrodes positioned on the scalp, which capture variations in the electrical potentials generated by neuronal activity (Hamzah and Abdalla 2024). EEG signals are often noisy and contaminated by artefacts, such as eye movements or muscle activity, so it is necessary to carry out pre-processing, which involves applying filters and other techniques to remove noise and artefacts, which will be explored further in section 2.2.4, improving the quality of the signals (Hramov et al. 2021). Relevant features are then extracted from the signal, identifying the most informative characteristics of the EEG signals that are related to emotional states, as previously explained in section 2.2.1. Finally, the features are then used to train deep learning models to identify patterns of brain activity associated with emotional states and subsequently evaluated (Erat et al. 2024).

Furthermore, the functioning of BCIs can be categorised into three distinct categories. Active BCIs require the user to consciously alter their brain activity to generate commands, such as imagining a movement. Reactive BCIs detect and classify the brain response to external stimulation for controlling commands, for example they can be used to control a virtual keyboard where each key flashes at a different frequency and when the user looks at the desired key, the BCI detects the corresponding frequency of the brain response, allowing typing without the need for physical movement. Passive BCIs, on the other hand, monitor the user's mental state in real time, without requiring conscious control, and are used in the entertainment industry, computer games, neuromarketing, as well as emotional recognition (Hramov et al. 2021).

BCI Applications in Audiovisual Content Consumption

The use of BCIs in the field of entertainment and multimedia has grown significantly, with applications ranging from adapting content to monitoring user engagement. In the context of audiovisual content consumption, BCIs can be used to measure viewers' emotional response in real time, allowing for a better understanding of the impact of films, advertising and video games, as previously mentioned.

Recent studies show that BCIs can be used to personalise audiovisual content based on the user's emotional state, such as:

- Affective media tagging: BCIs can monitor emotional responses to music, videos or films, allowing content to be labelled with the induced affective state. This can be used to create emotional metadata to organise and adjust the intensity of a film or video game's narrative based on the user's level of arousal and involvement (Mühl et al. 2014).
- Automatic media recommendation: Based on the affective states that content brings out in the user, BCIs can selectively offer or play media. For example, a system can detect sadness and automatically recommend happy music (Mühl et al. 2014).
- Communicating emotional preferences: Users can use BCIs to communicate emotional responses to certain objects or content, offering real-time feedback and enabling dynamic adjustments to the presentation of content and to optimise marketing strategies (Mühl et al. 2014).

Impact of Equipment Quality on EEG Signal Capture

The reliability of EEG-based emotional recognition depends heavily on the quality of the equipment used to capture the signals. Commercial devices, such as low-cost EEG headbands, have a reduced number of channels and use dry electrodes, which, although convenient, tend to capture lower quality signals due to higher impedance and susceptibility to noise. On the other hand, laboratory systems with wet electrodes and a greater number of channels distributed across the scalp allow for a more detailed capture of brain activity and, consequently, better discrimination between emotional states (Hramov et al. 2021).

The number of electrodes, the sampling rate and the quality of the electrodes are crucial factors since systems with a greater number of electrodes provide more detailed information about brain activity (Alarcão and Fonseca 2019). In addition, high sampling rates make it possible to capture signals with greater temporal precision (Hramov et al. 2021). High quality electrodes ensure better conductivity and lower noise (Shih et al. 2012). Low-cost EEG devices are becoming more common, but it is essential to consider their limitations in terms of signal quality and number of channels (Alarcão and Fonseca 2019).

Challenges and opportunities ahead

Despite significant advances in the field of BCIs applied to emotion recognition, technical and methodological challenges persist that limit their implementation in real-world scenarios. The inter- and intra-individual variability of EEG signals represents a significant obstacle, since the neuronal patterns associated with emotions can differ substantially between individuals and at different times (Hamzah and Abdalla 2024). Furthermore, the presence of physiological and environmental artefacts introduces an additional level of complexity to data analysis, which can compromise the accuracy of systems (Mühl et al. 2014).

However, the development of more robust deep learning algorithms (Gunda et al. 2024) and the improvement of EEG signal capture technology with the constant evolution of BCIs open up new opportunities to make these interfaces more accurate and accessible. The integration of BCIs with other physiological sensor modalities, such as heart rate variability and facial expressions or voice, can contribute to more reliable multimodal systems for detecting emotional states (brainwave book). In addition, the evolution of wireless BCIs and wearable devices could boost the adoption of this technology in entertainment contexts and the personalisation of audiovisual experiences (Hramov et al. 2021).

2.2.3 Traditional Methods vs EEG/BCIs

Emotion recognition traditionally focuses on analysing observable and expressive features such as facial expressions, tone of voice, body language and physiological signals such as heart rate (Huang et al. 2019). In contrast to non-physiological methods, the use of EEG signals and BCIs to collect these signals offers an alternative way of detecting emotional states through the direct measurement of brain activity, as analysed in more depth in sections 2.2.1 and 2.2.2.

Methods such as analysing facial expressions, tone and pattern of voice and body language are commonly used and can indicate emotions, however, these methods are highly influenced by context, culture and can even be deliberately manipulated making it difficult to obtain representative and authentic signals of emotion (Hamzah and Abdalla 2024). In addition, there are various physiological methods such as respiration, heartbeat and even skin temperature. Although these physiological methods are less susceptible to deliberate manipulation, their relationship with emotions is not direct or clear and can be influenced by external factors such as ambient temperature and level of physical activity (Egger et al. 2019).

EEG is therefore a promising alternative for recognising emotions, allowing direct analysis of the brain activity associated with emotional states as seen in section 2.2.1. Unlike non-physiological methods, EEG is not subject to context, culture and language and is hardly deliberately manipulated by people since it is "automatically" generated by the brain, just like other physiological methods (Islam et al. 2021). Therefore, it can be concluded that EEG has a number of advantages over traditional methods, such as:

- Not dependent on context and culture (Hamzah and Abdalla 2024).
- More objective and reliable since they are spontaneous and difficult to manipulate deliberately (Erat et al. 2024).
- It can be used to recognise emotions in people with physical or communication limitations where traditional methods fail severely, including people with disorders of consciousness (Huang et al. 2019).
- It is a method with good temporal resolution and acceptable spatial resolution which makes EEG an ideal technology for studying the precise timing of cognitive and emotional processing underlying behaviour (Erat et al. 2024).
- The high temporal resolution of EEG signals enables continuous, real-time monitoring of emotional changes (Huang et al. 2019).

However, the use of EEG also presents significant challenges. EEG signals have a low Signal-to-Noise Ratio (SNR) and are often mixed with much external noise when collected (W.-L. Zheng and Lu 2015). They are sensitive to interference from the environment and are usually mixed with other artefacts such as eye blinks or muscle movements, which requires advanced processing methods that increase computational complexity and in turn increase hardware requirements (Houssein et al. 2022). Furthermore, although BCI systems are evolving in terms of wearability, price, portability and ease-of-use, the need for specialised equipment with well-positioned electrode systems for emotion recognition using EEG remains a challenge (Alarcão and Fonseca 2019). This makes the installation and maintenance of the equipment much more complex than traditional methods, which in turn means that much of this equipment is still "exclusive" to the laboratory (Egger et al. 2019). Another critical challenge is the variation between subjects in which each person's brain exhibits distinct characteristics and patterns and the intra-individual variation in which the emotional response of the same individual can vary significantly depending on the context and even external factors, such as fatigue or different levels of attention between different tests, which makes the creation of generalisable models for emotion recognition quite complex. In addition, the use of different equipment with different calibration settings or even inconsistencies in the placement of the electrodes on the head can also cause variations which can jeopardise the effectiveness of model training (Hamzah and Abdalla 2024).

Table 2.1 shows a brief comparison between EEG and traditional methods.

Challenges inherent to recognising emotions regardless of the method

Despite the differences between traditional and EEG-based methods, emotion recognition faces transversal challenges that affect the accuracy and generalisability of the models developed. One of the main challenges is the subjectivity of emotions, since they are internal

Table 2.1: Comparison of traditional methods and EEG.

Criteria	Traditional Methods	EEG
Impact of context and cultural factors	High	Low
Possibility of deliberate manipulation	High	Low
Dependence on physical expressions	Yes	No
Equipment complexity	Low	High
Temporal resolution	Dependent on user reaction time	High, with real-time, continuous visualisation

and personal experiences and can be expressed and felt differently between individuals (Houssein et al. 2022).

Furthermore, the definition and categorisation of emotions remains an ongoing challenge (Assunção et al. 2022). Although there are well-established models, such as Ekman's model, based on discrete emotions, and Russell's circumplex model, the exact delimitation of emotions and their representation in a continuous or discrete space are still debated, and there is no scientific agreement on the definition of emotions (Erat et al. 2024). This theoretical uncertainty is also reflected in the difficulty of labelling training data in assays and interpreting the results obtained (Martinez and Valstar 2016). The subjectivity of emotions plus the challenges of each method make it difficult to create universal models capable of recognising emotions consistently in different users (Hamzah and Abdalla 2024).

Finally, there are also challenges related to ethics and privacy. The recognition of emotions, especially with neurophysiological signals, raises concerns about the inappropriate use of data and the protection of users' identities. Issues such as informed consent, secure data storage and transparency in the use of models are important to ensure the responsible adoption of these technologies (Saha et al. 2021).

2.2.4 Signal Processing and Deep Learning

Recognising emotions based on EEG requires a set of techniques ranging from signal processing to the application of Machine Learning or Deep Learning methods. This section presents the main processing steps for EEG signals as well as the deep learning architectures most commonly used for emotion recognition and the potential of transfer learning in this area.

Signal Processing for EEG

As mentioned in the previous section, EEG signals are prone to interference from various sources, including muscle activity, eye movements and environmental noise. The aim of signal processing is to "clean up" the data to improve signal quality, remove noise and artefacts and to prepare the signals for the extraction of relevant features for later analysis, which makes this stage critical (Torres et al. 2020).

The most common EEG signal processing methods include:

- **Downsampling** - Raw EEG data can come with very high frequencies such as 1000 Hz or 2000 Hz and it can be advantageous to downsample to reduce computational complexity without significant loss of information, as several studies have done, such as the various SEED and DEAP datasets.
- **Filtering** - The application of filters is essential to remove unwanted frequencies. A notch filter is used to remove the frequency of electrical networks, which normally varies between 50 and 60 Hz based on the frequency of the standard electrical signal in the particular country (Houssein et al. 2022). In addition, bandpass filters help to isolate the frequencies of interest for analysis using high-pass and low-pass filters that remove frequencies below or above a certain threshold, respectively (Hramov et al. 2021).
- **Principal Component Analysis (PCA)** - PCA is a statistical method and a data dimensionality reduction algorithm widely used in various fields. Mainly applied in pattern recognition, PCA can be used to reduce the dimensionality of a high-dimensional feature space and reduce the redundant information contained in the data (Yu and M. Wang 2022).
- **Independent Component Analysis (ICA)** - ICA is often used to remove artefacts. The raw EEG signal is a composite signal and, by applying ICA, distinct artefacts (e.g. related to eye movements, blinks, heart and muscle) can be isolated and removed to increase SNR (Hamzah and Abdalla 2024).
- **Common Average Reference (CAR)** - CAR is a technique used to generate a reference for each channel, in which the algorithm obtains an average of all the recordings on each electrode and then uses it as a reference, which in turn results in an improvement in the quality of the Signal to Noise Ratio (Torres et al. 2020).
- **Blind Source Separation (BSS)** - BSS is a technique used to eliminate internal and external artefacts and extract accurate brain signals from EEG signals (Hamzah and Abdalla 2024). This technique includes ICA and other algorithms (Yu and M. Wang 2022).

Deep Learning Methods

DL is an area of AI and ML involving computational models made up of several processing layers to learn data representations with various levels of abstraction and which enables computers to learn and understand the environment through a hierarchy of concepts (Hramov et al. 2021).

Over the years, many classic machine learning algorithms have been used to recognise emotions and have achieved satisfactory results (Yu and M. Wang 2022). However, DL has some advantages when compared to traditional ML methods, namely:

- Deep learning is more capable than traditional methods of handling high-dimensional and non-linear EEG data (Hamzah and Abdalla 2024).
- DL is capable of learning from time-frequency spectra and from raw EEG signals without requiring prior knowledge in this field (Hamzah and Abdalla 2024). Traditional machine learning algorithms, on the other hand, rely mostly on human expertise in feature selection and design, which significantly limits their accuracy (Gunda et al. 2024).

- DL can take advantage of distributed systems and parallel computing power, such as cloud platforms and Graphics Processing Units (GPUs), to process large-scale EEG signals and train complex models (Hamzah and Abdalla 2024). The development of increasingly fast GPUs is also a point in favour of DL in this regard (Gunda et al. 2024).
- DL models require large volumes of data to perform well, which can be a limitation (Hamzah and Abdalla 2024). However, with the proliferation of large-scale EEG datasets, this turns out to be another advantage for DL (Gunda et al. 2024).

The main types of Deep Learning models include Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU), where LSTMs and GRUs are specific types of RNNs. There are also hybrid models, such as CNN-LSTM.

CNNs are an architecture that excels at processing data with a spatial structure, such as images (Hramov et al. 2021). A Feed-Forward (FF) neural network is an Artificial Neural Network (ANN) where the connections between the nodes do not form a cycle, that is, the information only moves in one direction (Hramov et al. 2021). Unlike a conventional fully connected network, in which one neuron is connected to all the neurons in the layer that precedes it, CNNs extend and add two network structures, the convolution layer and the pooling layer, between the input layer and the fully connected layer (Yu and M. Wang 2022).

- Convolution layer - The convolution layer is the core of CNNs and can effectively enhance feature extraction of EEG data and reduce the influence of noise in EEG data (Yu and M. Wang 2022). It applies filters/kernels (small matrices) to detect local patterns in the input data and each filter learns to identify a specific feature, such as edges, textures or shapes (Hramov et al. 2021). Each filter slides over the input data generating a "feature map" (Houssein et al. 2022).
- Pooling Layer - Reduce the dimensionality of the data, minimising the size of the feature maps and improving the robustness of feature extraction (Houssein et al. 2022). The main functions of the pooling layer are feature selection and information filtering (Yu and M. Wang 2022).
- Fully connected layer - Receives all the features that have been created (Houssein et al. 2022). Finally, this layer will classify the output using Softmax functions (Yu and M. Wang 2022). The functions will assign probabilities to the existing classes.

The simplified and summarised structure of CNNs can be seen in figure 2.11.

RNNs, including LSTMs and GRUs, are suitable for processing sequential data as they take into account the previous step to make a decision for the next one, making them useful for capturing temporal dependencies in EEG signals or long texts. RNNs in particular have 3 layers, the input layer, hidden layer and output layer and the "memory" function of RNN is in the hidden layer (Yu and M. Wang 2022). However, RNNs are difficult to train because of exploding and vanishing gradient issues, which can make it difficult for the network to back propagate gradients over extended time intervals. To overcome these limitations, more advanced architectures such as LSTM and GRU have been developed (Houssein et al. 2022). The central idea behind LSTM is to introduce memory cells, which can maintain their state over time, and non-linear gating units, which regulate the information flow into and out of the cell (Hramov et al. 2021). GRUs, on the other

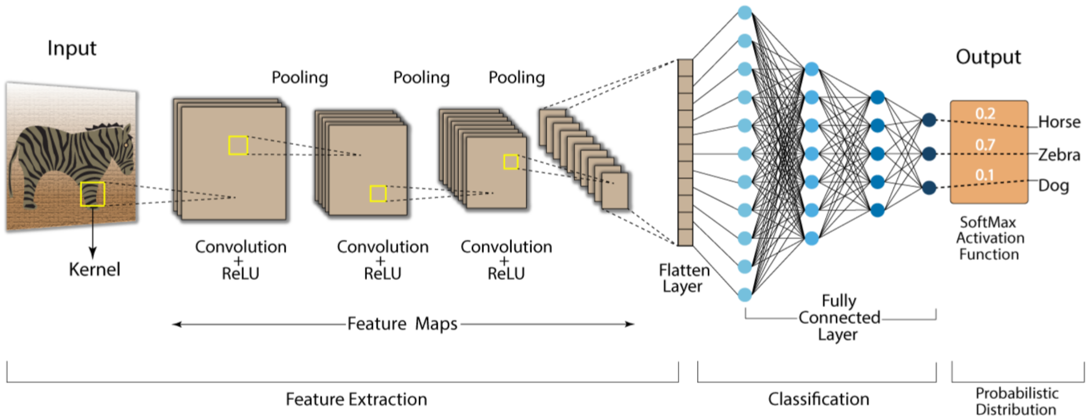

FIGURE 2.11: structure of a Convolutional Neural Network (Swapna 2020).

hand, are a simplified variation of LSTMs, using only two gates: reset gate and update gate. GRUs tend to have similar performance to LSTMs, but with less computational complexity, making them more efficient for various applications (Dey and Salem 2017).

In addition, there are hybrid modelling solutions, such as CNN-LSTM or CNN-RNN, which combine different types of neural networks to get the best out of each one. While CNNs are effective at extracting spatial patterns from data, RNNs specialise in capturing temporal dependencies (Yang et al. 2018). This combination allows hybrid models to be more efficient in analysing complex signals, such as EEG, where it is necessary to identify both spatial and temporal characteristics.

Table 2.2 shows a comparison of the models in terms of advantages and disadvantages.

TABLE 2.2: Comparison of different DL models

Model	Advantages	Disadvantages
CNN	Excellent for handling high-dimensional spatial patterns; Great parallelism	Poor at capturing temporal dependencies
RNN	Well-suited to sequential data and capturing temporal dependencies	Susceptible to exploding/vanishing gradients; difficult to train on long sequences
LSTM	Learns long-term dependencies better than vanilla RNNs; mitigates gradient issues	Lower parallelism; higher computational cost
GRU	Similar benefits to LSTM with fewer parameters; faster to train	Can underperform LSTM on some complex tasks
Hybrid (e.g., CNN-LSTM)	Captures spatial and temporal patterns jointly	Increased model complexity and training time

Transfer Learning

Transfer learning is a ML technique that allows the reuse of knowledge acquired in solving a problem to apply it to other different but related problems, thus avoiding the training of models from scratch as usually happens in traditional model training (Wan et al. 2021).

Emotion recognition based on EEG signals faces several challenges, which have been explored before, including high inter- and intra-subject variability. In other words, Session-to-Session Generalisation for a Same Subject and Subject-to-Subject Generalisation are problems that make it seriously difficult to create generalisable models that work well for all subjects (Jinpeng Li et al. 2020). Transfer learning helps to smooth out these differences by making the model flexibly match different individuals and tasks through adjustments, improving the robustness of the model (Wan et al. 2021).

That said, transfer learning offers several advantages over traditional model training, especially in the context of EEG analysis:

- Less Data Requirements: Collecting and labelling large EEG datasets is difficult and expensive due to the need for specialised equipment and the time-consuming nature of the process, and transfer learning learns the target task according to prior knowledge learned in a similar domain using a small amount of data in the target domain to tune the classifier, which reduces the available data requirements (Wan et al. 2021).
- Generalisation: By using models that have been pre-trained on other people's data, transfer learning improves the generalisation capacity of the models (Jinyu Li et al. 2022).

Furthermore, it is important to bear in mind that the transfer learning methodology can not only utilise the regular techniques and tactics on which traditional learning models are based, but can also transfer useful information resources from the training data in the source domain to the test data in the target domain, across the large domain gap, to benefit classification. Therefore, compared to traditional learning approaches, transfer learning methods are better able to ensure good generalisation while maintaining sufficient discrimination capability (W. Li et al. 2022).

Various approaches have been explored in the context of transfer learning for EEG and, according to the situation of the labelled data, transfer learning can generally be divided into three categories: inductive, transductive and unsupervised transfer learning (W. Li et al. 2022). Most inductive transfer learning studies are carried out in the situation where the labels of the source domain and the target domain are both known, whereas in transductive transfer learning, the labels of the target domain are unknown, and the source domain has a large amount of labelled training data available. In unsupervised transfer learning, the data in the source domain and the target domain are both unlabelled (Wan et al. 2021). Inductive transfer learning and transductive transfer learning are widely used in EEG-based emotion recognition, while unsupervised transfer learning deals with clustering and dimensionality reduction issues. Inductive transfer learning, transductive transfer learning and unsupervised transfer learning methods can be further categorised into modes based on instances, features, parameters and relationships, based on their knowledge transfer mechanisms (W. Li et al. 2022). Instance transfer uses reweighted samples from the source to help train the model on the target, while in feature transfer the transferred knowledge is encoded in the feature representation, transferring features that are informative for new tasks. Parameter Transfer optimises the parameters or hyperparameters of the model and assumes that the target shares parameters with the source in different tasks (Jinpeng Li et al. 2020).

2.3 Commercial and Scientific Solutions

The integration of EEG-based emotion recognition into commercial and scientific (research) applications has made significant progress in recent years. While commercial solutions prioritise accessibility, ease of use and integration with consumer-oriented applications, scientific research has focused more on improving accuracy through advanced automatic and DL techniques. This section presents an overview of both domains, along with the main solutions, methodologies and existing gaps.

2.3.1 Commercial solutions

Commercial EEG devices designed for emotion recognition have gained popularity, particularly in applications related to neurofeedback, meditation, and sleep tracking. Table 2.3 summarizes key commercial EEG solutions available in the market, detailing their technical specifications and areas of application.

Devices such as Muse and NeuroSky are aimed at general consumers and focus on stress management and meditation, while platforms such as OpenBCI and Emotiv offer more flexibility to researchers and developers, allowing them to create their own applications. For instance, OpenBCI has an open page where most of the scientific articles featuring their devices are listed. The site in question is : (OpenBCI 2019).

A notable limitation among these devices is that most systems don't come with deep learning models already inbuilt and their signal processing is basic, so none of these devices are really prepared for emotion recognition tasks.

2.3.2 Scientific Solutions

In contrast to commercial EEG applications, scientific research has taken advantage of advanced ML and DL techniques to improve the accuracy of emotion recognition systems. Studies in this field often use publicly available datasets such as DEAP, SEED and others. Table 2.4 shows examples of studies, among the thousands that exist, as well as some details of these studies.

One of the major gaps in the literature is the use of datasets with more diverse emotions. Most of the studies found use the DEAP (Valence/Arousal) and SEED (Positive, Neutral, Negative) datasets, while there are others (e.g. SEED V, SEED VII) that have a greater diversity of emotion labels.

Table 2.3: Commercial EEG devices and their characteristics.

Device	**Electrodes**	**Emotions**	**Sampling Rate**	**Modalities**	**Applications**	**Electrode Positions**
Muse S Headband	4 + 2 aux	Active, Neutral, Calm	256 Hz	EEG	Meditation routine, Sleep Tracking	TP9, AF7, AF8, TP10
Muse 2	4 + 1 aux	Active, Neutral, Calm	256 Hz	EEG	Sleep Tracking, meditation practice	TP9, AF7, AF8, TP10
OpenBCI	8/16	N/A	250/125 Hz	EEG	Research	Fp1, Fp2, F3, F4, F7, F8, T7, T8, C3, C4, P7, P8, P3, P4, O1, O2
NeuroSky MindWave	1	Attention, Meditation	512 Hz	EEG/ECG	Monitors attention levels	FP1
Emotiv	2/5/14/32	N/A	128/256 Hz	EEG	Research	–
Neuphony	6	Stress, Calm, Mood, Focus	250 Hz	EEG	Monitors focus, calmness, and mood	Fp1, Fp2, F3, F4, Fz, T3, T4, Pz
iMotions	8–64	N/A	256–500 Hz	EEG//Facial Expressions	Neuromarketing, research	–

Table 2.4: Summary of deep learning models for emotion recognition using EEG.

Reference	DL Model	Pre-Processing	Modalities	Emotions	Accuracy	Features	Dataset
(Yang et al. 2018)	CNN–LSTM	Base Mean Removal	EEG	Valence, Arousal	90%	–	DEAP
(Iyer et al. 2023)	CNN, LSTM, Hybrid, Ensemble	Dataset-based	EEG	Positive, Neutral, Negative, Valence, Arousal	89.5%, 89.99%, 93.46%, 97.16%	DE	SEED, DEAP
(W.-L. Zheng and Lu 2015)	DBN	Dataset-based	EEG + Eyes	Positive, Neutral, Negative	86.85%	PSD, DE	SEED
(Tao et al. 2023)	Attention-based CNN–RNN	Dataset-based	EEG	Valence, Arousal	93.85%	–	DEAP, DREAMER

Chapter 3

Tools, Methods and Experimentations

This chapter outlines the tools and methods used throughout the work, as well as the experiments carried out. It begins by looking at the tools available for analysing and processing the data, followed by an explanation of how the data collection was carried out. The experiments carried out are then detailed, including the characteristics of the datasets used and the structures of the respective tests. Finally, the issue of data protection and ethics is addressed, ensuring that its use complies with the principles of security and privacy.

3.1 Tools

Three main tools were explored for analysing Electroencephalogram (EEG) data: MNE, Matlab with Toolboxes (namely EEGLab), and OpenVibe. The choice of the final tool was based on several criteria, including ease of use, support for the required functionalities, compatibility with the technologies used, and cost.

The MNE tool was selected because it is an open-source library developed in Python, with extensive support for analysing EEG signals. It stands out for its comprehensive and detailed documentation, which is rare and extremely valuable in this area of tools, allowing for a smoother learning curve. In addition, MNE features direct compatibility with most of the file formats used for EEG data, such as *.edf, .bdf* and *.cnt*, eliminating the need for prior conversions, which can introduce errors or consume additional time. However, files in the .mat format presented challenges in their handling, as they are not directly supported by MNE.

On the other hand, Matlab and its toolboxes, although widely recognised for their versatility, have significant limitations. Although there is a free online version, this is limited in terms of time and performance, making it difficult to manipulate large volumes of data, which is the case of EEG data, or more intensive processes. In addition, the full version of Matlab is associated with additional costs, which makes this option less attractive.

In addition, OpenVibe, although it has pipelines dedicated to visualising EEG signals, and being possible to process EEG data, it is an old tool with poor documentation, which results in a longer learning curve. Besides, it also has poor compatibility with de usual file formats, which would possibly give problems and delay the process. However, visualising and processing EEG signals in real time is a strong point which may prove this tool very useful even though it is possible to integrate LSL (Lab Streaming Layer) with MNE also allowing it to visualise EEG data in real-time.

In addition to these tools, various Python libraries were used to complement the workflow, such as NumPy, Matplotlib, Scikit-learn, SciPy and Pandas. The working environment used included Jupyter Notebooks, which facilitate documentation and iterative code execution.

3.2 Datasets Acquisition

The datasets used for this project were obtained by contacting Shanghai Jiao Tong University, through a formal request process. These datasets are called "SEED" and were selected based on their relevance to the study of emotions through EEG signals while individuals were watching audiovisual content (in this case, film clips).

Although each dataset has a main objective in common, which is to provide multimodal data (EEG signals and eye movements) for recognition, each one has distinct characteristics and different contexts that are explored in each of their own articles.

The focus of the SEED dataset is on investigating critical frequency bands and critical channels for efficient emotion recognition based on EEG. And to show the superior performance of Deep Learning (DL) models over other models such as K-Nearest Neighbors (KNN), logistic regression and Support Vector Machine (SVM) (W.-L. Zheng and Lu 2015). The results of the experiment also indicate that features such as Differential Entropy (DE) extracted from EEG data have accurate and stable information for recognising emotions (Duan et al. 2013).

The focus of SEED-IV was to integrate brain signals and users' external subconscious behaviours, presenting a multimodal structure called the EmotionMeter to recognise human emotions using six EEG electrodes and eye-tracking glasses, i.e. combining EEG with eye movements to obtain a particular emotion (W. Zheng et al. 2018).

SEED-V and SEED-VII represent further developments, focussing on more complex experimental configurations and expanding the amount of data available.

The SEED-V dataset's main objective is to provide multimodal data (EEG signals and eye movements) for recognising five different emotions: happy, sad, neutral, fear and disgust. It was also done in the context of comparing the performance and robustness of two multimodal DL models for emotion recognition: Deep Canonical Correlation Analysis (DCCA) and Bimodal Deep Auto Encoder (BDAE) (Liu et al. 2021).

The focus of SEED-VII was specifically to offer a multimodal dataset for recognising emotions, exploring six basic emotions (joy, sadness, fear, disgust, surprise and anger) as well as the neutral emotion. It combines EEG signals and eye movements and includes continuous labels, which is a big difference to previous datasets, which indicate the intensity of emotions. One of the aims is to facilitate the investigation of emotional patterns and the development of emotion recognition techniques in subject-dependent and subject-independent conditions (Jiang et al. 2024).

3.3 Experimentation

The main experimentation carried out in this project consisted of using the MNE library to analyse the main characteristics of the datasets. Initially, the EEG signals were loaded and visualised, allowing the general quality of the data to be assessed, mainly to check that everything matched what was said on the official website for each one. During this process, specific features of the library were explored, such as filtering the signals to remove low and high frequency noise, decomposing the signals into different channels, and visualising the signals.

To deal with the *.mat* format files present in the datasets, online basic Matlab was used, which made it possible to inspect the structure of the files and carry out initial analyses. However, due to the large size of some files, which exceeded the capabilities of this tool, it was necessary to switch to Python tools, such as the SciPy library. This proved to be more efficient in processing large volumes of data and extracting structural information.

During experimentation, various EEG data processing techniques were implemented, including noise removal, signal normalisation and the extraction of specific features. These steps not only helped with familiarisation with the tools, but also offered valuable insights for future steps.

3.3.1 Datasets Details

The datasets collected for this project were designed to investigate emotion recognition based on EEG signals, complemented by multimodal data including eye movements and using film clips as a stimulus for evoking emotions.

The **SEED** dataset includes 15 trials per participant, where each trial corresponds to the presentation of a film clip of approximately four minutes, distributed proportionally between positive, neutral and negative emotions. Fifteen individuals took part in this study (seven males and eight females) aged between 20 and 24 years. EEG signals were collected at a sampling rate of 1000 Hz, using 62 channels positioned according to the international 10-20 system (W.-L. Zheng and Lu 2015).

The **SEED-IV** dataset expanded this approach with three sessions per participant, held on separate days, with 24 trials per session. Each trial involved showing film clips of approximately two minutes, designed to induce the emotions happy, sad, neutral and fear. As with SEED, 15 individuals (eight females and seven males) aged between 20 and 24 years took part and EEG signals were collected with 62 channels at 1000 Hz (W. Zheng et al. 2018).

In the case of **SEED-V**, the participants took part in three sessions with a minimum interval of three days between them. Each session included 15 clips, totalling three for each of the five emotions studied: happy, sad, neutral, disgust and fear. The stimuli lasted between two and four minutes, and the EEG was recorded using the same 62-channel protocol and 1000 Hz sampling rate. This study involved 16 participants (6 males and 10 females), aged between 19 and 28 (T.-H. Li et al. 2019).

Finally, the **SEED-VII** dataset has unique characteristics compared to the previous ones. Four sessions were held per participant, with 20 trials per session. Each trial consisted of showing film clips lasting between two and five minutes, totalling 80 clips distributed among seven emotions: happy, sad, neutral, fear, disgust, surprise and anger. In addition, this dataset introduces continuous labels for emotions, making it possible to measure emotional intensity, an innovative approach compared to other datasets. Twenty individuals took part in this study (10 males and 10 females) aged 19 to 26 years, with EEG signals collected at 1000 Hz and 62 channels following the international 10-20 system (Jiang et al. 2024).

As previously mentioned, all datasets include multimodal data, providing both EEG signals and the participants' eye movements. In addition, both raw and pre-processed data are provided. However, it has yet to be decided whether the already processed data or the eye movement data will be used in this project, since the focus is on doing the experiment especially with the EEG data and it may be interesting to do the pre-processing myself in order to work with the previously mentioned tools, explore and get to know the data and rectify the processing of the same while being sure of all the processing actions carried out. Table 3.1 shows a summary and comparison between the different datasets explored.

TABLE 3.1: Comparison of Used Datasets

Attribute	SEED	SEED IV	SEED V	SEED VII
Labels	Positive, Neutral and Negative	Happy, Sad, Neutral, Fear	Happy, Sad, Neutral, Disgust, Fear	Happy, Sad, Neutral, Fear, Disgust, Surprise, Anger
Participants	15	15	16	20
Clip Duration	4 min	2 min	2-4 min	2-5 min
Number of Clips	15	72	15	80
Data Format	.cnt	.mat	.cnt	.cnt
Sampling Rate (Hz)	1000	1000	1000	1000
EEG Data Size	42.4 GB	6.5 GB	37.6 GB	79.8 GB
Total Dataset Size	93 GB	7.3 GB	40.5 GB	170.9 GB

3.3.2 Structure of experimental trials

The experimental protocols adopted to collect data for the datasets differ, but they all have things in common, such as the age range, all participants having the right hand as their dominant hand, normal hearing, normal vision, all Chinese participants and always with a balanced number of men and women. The details of the trials for each dataset are described below.

Trials on the **SEED** dataset consist of 15 trials per session. Each trial begins with a 5-second warning, followed by the presentation of a film clip lasting approximately 4 minutes. After the clip, participants have 45 seconds to complete an emotional self-assessment, followed by a 15-second rest period. The order of the clips was carefully organised to avoid the consecutive presentation of stimuli associated with the same type of emotion. During the self-assessment, participants reported their emotional reactions by filling in a questionnaire immediately after each clip (W.-L. Zheng and Lu 2015).

In **SEED-IV** dataset, each participant completed three sessions held on different days, with 24 trials in each session. Each trial begins with a 5-second pause before the film clip is shown, which lasts approximately 2 minutes. After the stimulus, participants had

45 seconds to self-assess the emotions they felt, followed by a 15-second rest break (W. Zheng et al. 2018).

The participants in the **SEED-V** study underwent three experimental sessions, with a minimum interval of three days between them. Each session consisted of showing 15 emotional clips, 3 clips for each type of emotion. To avoid boredom and guarantee the effectiveness of the stimuli, the materials shown were completely different between sessions, and the total duration of each session was approximately 50 minutes. Before each clip, a 15-second warning was given about the content and the emotion to be induced. After the stimulus, participants had 15 to 30 seconds for self-assessment and rest, depending on the type of emotion (30 seconds for disgust and fear; 15 seconds for joy, neutrality and sadness). The self-assessment included a scale of 0 to 5 points, where 5 indicated the best emotional induction effect and 0 the worst. For neutral stimuli, it was recommended that a score of 5 represented a natural state and 0 represented mood fluctuations (T.-H. Li et al. 2019).

SEED-VII introduced a different experimental protocol. Each participant took part in four experimental sessions, with 20 trials per session. Each trial consisted of two parts: the visualisation of an emotional clip and a subsequent self-assessment. Before and after each clip, a 3-second warning was presented. Only five of the seven possible emotions were induced in each session, minimising sudden changes in emotional valence. At the end of each session, participants reviewed the 20 clips shown and assigned continuous labels (on a scale of 0 to 1) that reflected the intensity of the emotions experienced (Jiang et al. 2024).

In figure 3.1 is an example of trial from the SEED IV dataset, changing the time it can be adapted for the other datasets of the SEED family.

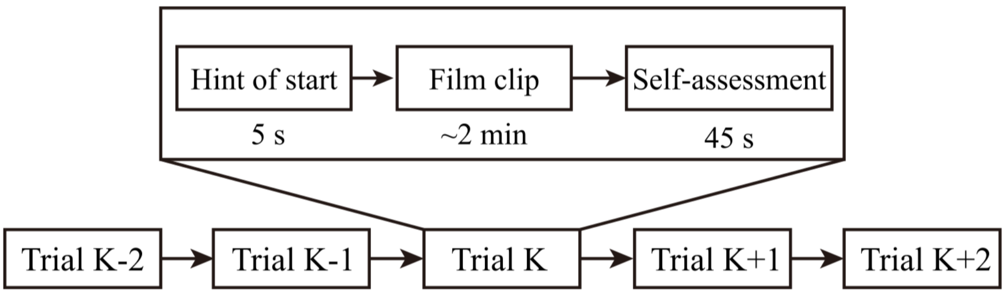

Figure 3.1: Example of Trial from the dataset SEED IV (*SEED-IV Dataset* 2025).

3.4 Data Protection, Security and Ethics

The development and use of emotion recognition systems raise a number of ethical, legal, and social risks that cannot be ignored, especially when considering their potential impact on fundamental rights (Gremsl and Hödl 2022). One of the main concerns relates to the scientific reliability of these technologies as the literature has shown that many approaches are based on simplistic and culturally biased assumptions, which presume that emotions can be universally classified based on observable signals (Hartmann et al. 2024; Katirai

2023). This simplification ignores the psychological and sociocultural complexity of human emotions and, as a result, leads to systems that are susceptible to misinterpretation (Hartmann et al. 2024).

The sensitivity of emotional data is another critical dimension in which, unlike other types of personal information, data relating to an individual's emotional state is inherently intimate and can reveal deep aspects of their identity, mental health or social context (Katirai 2023). This characteristic makes them particularly vulnerable to abuse if strict precautions regarding consent, security, and limitations on processing are not applied as the risk of using these systems in high-impact contexts such as healthcare, education, employment, or policing increases the severity of the problem (Katirai 2023). Decision-making based on emotional inferences can result in discrimination, unfair exclusion, or violation of individual autonomy, compromising essential values such as human dignity and freedom of thought (which includes freedom of expression) (Gremsl and Hödl 2022; Hartmann et al. 2024).

3.4.1 Compliance with GDPR and AI Act

The General Data Protection Regulation (GDPR) is an European Union legislation and is considered to be the strongest privacy and security law in the world. This regulation, approved in 2016 and in force since 2018, defines the rules for collecting, processing and storing personal data, guaranteeing the privacy and rights of European citizens defining the individuals' fundamental rights in the digital age, the obligations of those processing data, methods for ensuring compliance, sanctions for those in breach of the rules (Union 2024). The GDPR applies to any organisation handling the personal data of EU citizens, regardless of where the entity is based so the EU data protection legislation includes safeguards for when transferring data to third countries (Union 2025b). Non-compliance can result in large fines, up to 20 million euros or 4 per cent of global annual turnover, whichever is higher (Consulting 2016). In (Consulting 2016) it is possible to see the official documentation and rules of de GDPR.

The Artificial Intelligence (AI) Act is a European regulation on AI and the first comprehensive regulation on AI by a major regulator in the world (Union 2025d). The regulation follows a risk-based approach, categorising AI applications into four levels: unacceptable risk, high risk, limited risk, and minimal risk (Union 2025c). The unacceptable risks include all AI systems considered a clear threat to the safety, livelihoods and rights of people and are outright banned/ prohibited (Union 2025a). The high risk include AI systems that negatively affect safety or fundamental rights in specific areas such as education or law enforcement, for example (Union 2025c). The limited risks refer to systems with a need for transparency around the use of AI. Finally, in the minimal risks, the AI Act does not introduce rules for which most of AI systems currently used in the EU fall into this category (Union 2025a). Figure 3.2 it is possible to see a summary of the different levels of the AI Act.

As stated in previous studies of the SEED Datasets authors, (W.-L. Zheng and Lu 2015), (W. Zheng et al. 2018), (Liu et al. 2021), (Jiang et al. 2024), the datasets used in this work were collected based on ethical principles approved by the ethics committee of Shanghai Jiao Tong University, which is widely respected, ensuring that participants were not subject to significant risks. Prior to data collection, informed consent was obtained from all participants, who were duly informed about the objectives of the study, the non-invasive methods of data collection and the possibility of withdrawing at any time.

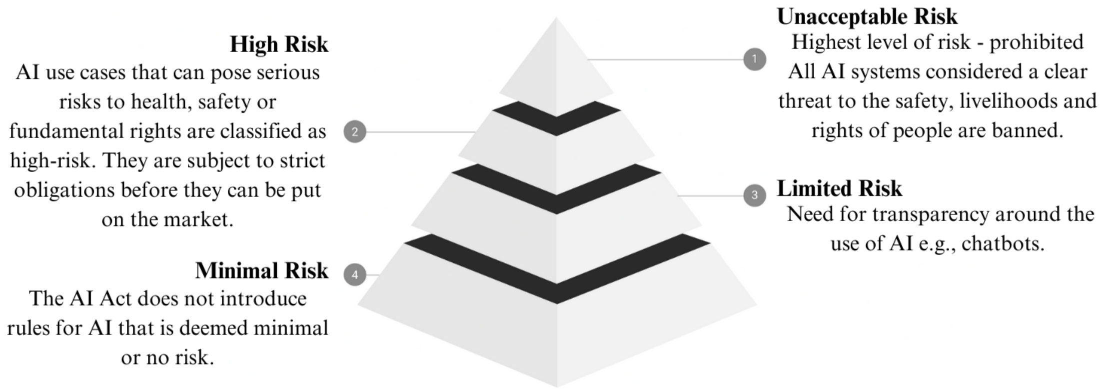

FIGURE 3.2: Overview of AI risk levels as defined by the AI Act Union 2025a.

The privacy of the participants was guaranteed through a process of data anonymisation. There are no elements in the SEED datasets that make it possible to directly or indirectly identify the participants, as was confirmed through the exploration and experimentation carried out in this project (section 3.3). This is particularly relevant in the context of the GDPR, which establishes strict criteria for the protection of personal data in the European Union. Compliance with the principles underlying the GDPR reinforces the commitment to privacy and people's fundamental rights, even considering that the data was collected outside the European area.

Additionally, this project was developed in alignment with the rules of the AI Act, approved by the European Union as the first regulation on AI, which establishes rules and obligations at European level based on risk, categorising AI systems according to their potential impact on health, safety and fundamental rights. Systems considered "high risk", such as those involving the analysis of sensitive data, are subject to strict requirements that ensure their transparency, accountability and ethical alignment.

Although the SEED datasets were collected before the implementation of the AI Act, the adopted measures ensure compliance with the fundamental principles of this legislation, for example, biometric identification and categorisation of people is not possible. Strict anonymisation and respect for informed consent ensure that the data is used ethically, protecting the privacy of participants and mitigating risks associated with the development of models based on AI that will be carried out during this project.

Building trust in EEG-based AI systems requires more than high accuracy on a benchmark. Trust is earned throughout the entire development cycle, from problem definition and data collection to model development, validation and deployment depending on technical rigour and ethical safeguards as well as data management. In the context of EEG, this means demonstrating that signals were acquired and preprocessed using robust and well-documented procedures; that artefacts (e.g., eye and muscle activity) and subject variability were handled appropriately, ensuring that models generalise well and that performance is accompanied by limitations that actually exist in a transparent manner. Equally important are practices that make systems ethical and reproducible: clear protocols for data acquisition, documentation, and assurance that systems are ethical, fair, and respectful of the sensitivity of the data, complying with GDPR and AI Act standards and prioritising data privacy from the beginning.

Finally, adherence to strict ethical protocols in this project demonstrates a commitment to transparency and accountability, core values of the GDPR and the AI Act. These measures ensure that the development of AI systems for this project is conducted ethically, mitigating risks and protecting people's fundamental rights, in compliance with European regulations.

3.4.2 Ethical and Technical EEG Guideline for Data Collection

The collection of EEG data must begin with valid consent, "freely given, specific, informed and unambiguous indication of wishes" (Article 4 of GDPR), with the possibility of withdrawal at any time and without prejudice to the participant (European Data Protection Board 2020; *GDPR* 2016).

This consent must specify the purposes, legal basis, expected risks, security measures, data category, intended audience, and retention period, ensuring transparency appropriate to the context and literacy level of the participants (European Data Protection Board 2020; *GDPR* 2016) (EDPB, pp. 14-19; GDPR Art. 13-14 and Art. 5(1)(a)).

In the case of research (involving human subjects), the protocol must be submitted in advance and approved by an independent ethics committee, in which information to participants, voluntary participation, and the right to withdraw consent are central requirements (World Medical Association 2013) (§§ 23-27).

Given the sensitive nature of EEG signals, which, as mentioned in sections above, are capable of revealing aspects of physical/mental health or serving as a basis for inferences about personal states and traits, this data should be treated as "special categories" whenever it allows inferences about health or is used for biometric identification (*GDPR* 2016) (GDPR Art. 9(1), Art. 4(14) and (15)). Furthermore, the collection must comply with the data minimisation principle (only collecting and processing the minimum necessary data), implementing data protection by design and ensuring data protection by default, and the data should be anonymised (irreversible) or pseudonymised. Pseudonymisation is defined in Art. 4(5) GDPR as "the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person." (European Data Protection Board 2025) There are guidelines produced by the European Union on how to carry out pseudoanonymisation correctly, such as the report made by the European Union Agency for Cybersecurity (ENISA), "Pseudonymisation techniques and best practices" (ENISA 2019).

When collection is intended for products/services (outside an exclusively scientific context), it should be noted that AI intended for emotion recognition is subject to specific restrictions in the EU, including prohibitions in work and educational environments (AI Act - see the European Commission's official summary of prohibited practices, it's very important since the non-compliance can result in very heavy fines) (*EU AI Act* 2024) (Art. 5).

On a technical level, the quality of the acquisition begins with the preparation of the environment (e.g., controlled electrical noise), calibration and recording of impedance, electrode positions (e.g., the international 10-20 scheme) and amplifier parameters, since these factors directly affect the SNR and reproducibility. In addition, documentation of the experimental protocol, such as stimuli, sampling rates, number of trials, sessions, triggers, and exact timing of events, is essential for replication, ethical/technical auditing, and data processing (M. S. Pais-Vieira and C. F. d. S. Pais-Vieira 2018; Pernet et al. 2020).

Finally, data must be stored with security proportional to the risk, with controlled access, encryption, and separation between identification data and raw/processed signals, always respecting minimisation and limitation of personal data conservation as well as prohibited practices (*GDPR* 2016) (GDPR Art. 5(1)(c),(e) e Art. 32).

Chapter 4

Implementation, Analysis and Results discussion

This chapter presents the practical implementation of the work developed, covering the exploratory analysis of the data, the pre-processing pipeline applied to the EEG signals, and the feature extraction methods used. The experimental design and the deep learning architectures tested are also described, as well as the different evaluation scenarios considered.

In addition to presenting the quantitative results, this chapter critically discusses the impact of methodological choices, from the definition of frequency bands to normalisation and evaluation strategies, on the generalisation capacity of the models, seeking to establish the link between experimental practice and the outlined research objectives.

4.1 Exploratory Data Analysis

Before proceeding to pre-processing and feature extraction, an exploratory analysis of the EEG signals from all datasets included in this study was conducted to assess the quality of the recordings and identify any structural problems that could affect the subsequent stages. This initial step proved essential to confirm the need for certain filters and artefact removal techniques, to construct a plan for the processing steps, and to establish a comparative basis between different datasets. One of the first steps was to visually analyse the Power Spectral Density (PSD), calculated for all videos of all subjects in every dataset. This analysis allowed to verify the presence of consistent patterns in different frequency bands (*delta, theta, alpha, beta* and *gamma*) but also revealed significant problems. In all datasets, it was possible to clearly identify noise from the electrical network, reflected in sharp peaks in the 50 Hz region, even though it is clear that even the raw data has already been filtered using a notch filter (probably at the same time as the data was collected), as well as recurring noise in specific channels, suggesting local interference. These observations support the need to apply a mandatory 50 Hz notch filter, as well as to use source separation techniques, such as Independent Component Analysis (ICA), to mitigate persistent artefacts. In Figure 4.1, it is possible to observe the noise present at 50 Hz (line noise), as well as artefacts such as channel noise.

In addition to the frequency domain, the signal was also analysed in the time domain, with the purpose of visually confirming the noise identified in the PSD and establishing a baseline for the "raw" signals, before any cleaning. This temporal inspection was not used primarily to detect specific artefacts, but rather to obtain a clear idea of the amplitude and variability of the signal, serving as a direct comparison with the signals after applying

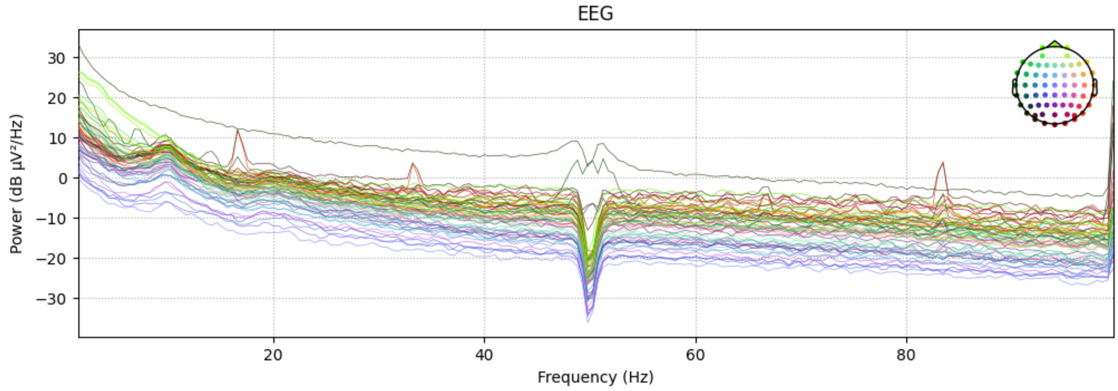

FIGURE 4.1: PSD before any alterations.

the filters and ICA. Figure 4.2 shows how the signal is in the time domain before any processing steps, highlighting the need for such processing.

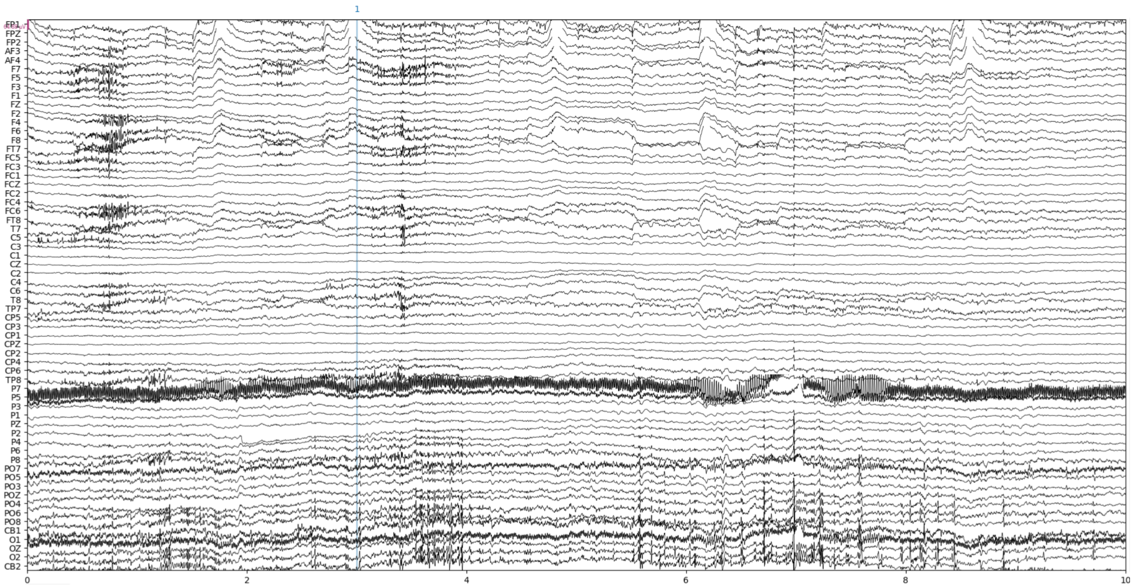

FIGURE 4.2: Signal in Time Domain before any alterations.

Another crucial point in this exploratory analysis was the evaluation of the impact of artefact removal using ICA. Although ICA is indispensable for separating neural components from artefacts, it was observed that overly aggressive use can compromise signal quality, leading to the elimination not only of noise but also of relevant neural information. In extreme cases, this overcorrection resulted in almost flat signals, revealing a significant loss of variability that completely distorts actual brain activity. Figure 4.3 shows how the signal in the time domain looks after a very aggressive ICA.

The analysis also revealed structural differences between the datasets. In SEED I and SEED IV, the data provided was already segmented and partially processed, which made direct access to raw files impossible. In the case of SEED I, there were original files in CNT format, but these were corrupted and could not be used with MNE. In SEED IV, only processed versions were available, with no access to raw records. This limitation means that, for these two datasets, it was not possible to apply a complete pre-processing

FIGURE 4.3: Signal in Time Domain after agressive ICA.

pipeline or use baselines provided by the authors (for baseline removal), forcing to adopt alternatives to ensure comparability between subjects and sessions, as revealed in the next section.

In contrast, in SEED V and SEED VII, raw files were made available, which allowed for a more detailed analysis, not only of the original signal but also of the effects derived from the different processing stages (even so, in SEED V, there was one corrupted file from one person). This availability made it possible, for example, to compare results by performing segmentation on videos before pre-processing and segmentation after pre-processing, revealing relevant differences in the preservation of signal characteristics.

Overall, this preliminary data analysis confirmed the existence of structural noise in all datasets, validated the need for specific filters (notch and bandpass), highlighted the importance of carefully calibrating the ICA, and emphasised the variability between datasets resulting from the availability, or absence, of raw files. This step reinforces methodological transparency and establishes a solid framework for the choices made in the subsequent pre-processing stages.

4.2 Pre-processing

After exploratory analysis of the signals, a pre-processing pipeline was developed with the aim of improving the quality of EEG data, reducing artefacts, and normalising data between subjects and sessions, ensuring the comparability of results obtained in different datasets. This step is critical, given that EEG signals are inherently noisy and vulnerable to multiple sources of interference, requiring a set of carefully balanced processes to preserve relevant neural information. Pre-processing was implemented in MNE-Python, taking advantage of its modules for filtering, resampling, referencing, and ICA. ICLabel was used to identify artefacts in the independent components, with changes made directly to the library to expose the probabilities by class (e.g., brain, eye movement, muscle, heart, line noise, channel noise, other) and not just the class with the highest probability. This probabilistic access allowed to define conservative exclusion policies (e.g., excluding only ICs with a high and consistent probability of being non-neural or even keeping artefacts that have a small probability of containing brain signals even if "brain" is not the dominant class), reducing the risk of excessive removal. It is important to note that ICLabel imposes methodological constraints that directly influence some of the steps in the pipeline.

The pipeline adopted in this work included the following main steps:

1. **Notch Filter (50 Hz)**

   Since all datasets presented noise from the electrical network, clearly identified in the spectral analysis, a notch filter was applied around 50 Hz. This filtering proved essential to remove noise spikes that could otherwise distort feature extraction. Figure 4.4 shows a decrease in noise at 50 Hz (line noise).

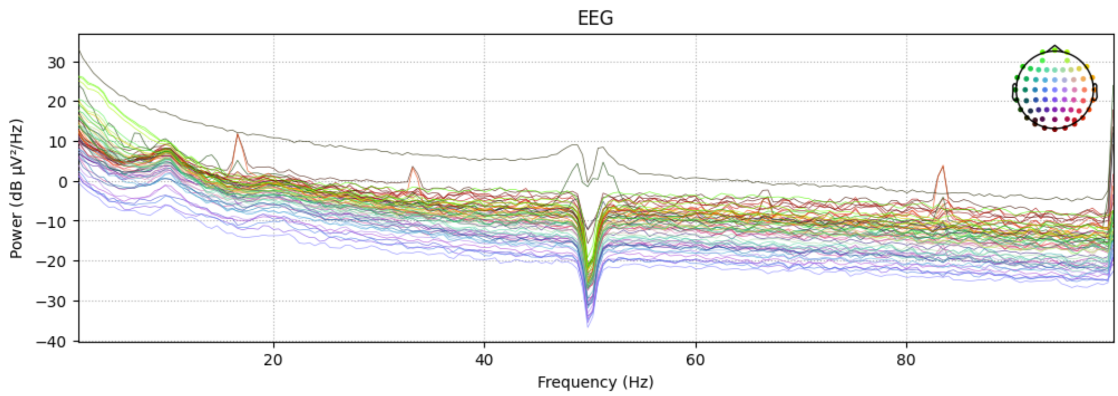

FIGURE 4.4: PSD after aditional notch filter.

2. **Bandpass Filter**

   The bandpass filter was applied to restrict the signal to the relevant frequencies and to eliminate unnecessary noise. In SEED V and VII, where raw files were available, the 1-100 Hz band was used to comply with the requirements of ICLabel, which was trained in this frequency range. On the other hand, in SEED I and IV, the signals had already been provided with prior filtering between 0.1 and 70 Hz, which limited the possibility of directly applying the same criteria.

3. **Resample**

   All signals were resampled to 200 Hz. This sampling rate allowed to adequately preserve the bands of interest for emotion recognition, while significantly reducing the computational cost of processing and training the models, as well as the space occupied by the files.

4. **Common Average Reference (CAR)**

   To mitigate global artefacts and improve the signal-to-noise ratio between channels, the CAR technique was applied. This step had a visible impact on the spectral distribution of the channels, stabilising the relative power between bands. Figure 4.5 shows the impact of CAR, where it can be seen that the powers are closer together.

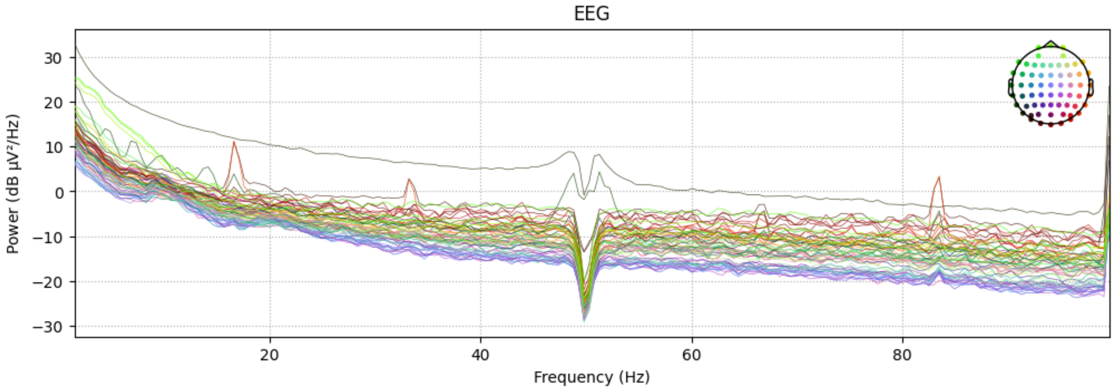

FIGURE 4.5: PSD after CAR.

5. **Independent Component Analysis (ICA)**

   The ICA was applied with the aim of separating independent sources and removing non-neural components, such as eye movements, muscle activity or noise from specific channels. However, its use was carefully balanced, as overly aggressive application can result in the elimination of relevant components of brain activity. To mitigate this risk, integration with ICLabel was used, which allows probabilities to be assigned to each component in different categories, as mentioned above. ICA was adjusted with the Infomax method as required by ICLabel. In addition, the number of components extracted was the maximum, that is, one component per channel. Since not all components are artefacts considered “noise”, ICLabel was used to classify each component. As previously mentioned and shown in Figure 4.3, excessive removal is detrimental to the signal, so a conservative removal was chosen, in which only ICs with a high probability of being artefacts were excluded. Figure 4.6 shows five example components, that were labeled "channel noise", "eye artefact", "eye artefact", "line noise" and "heart beat", respectively.

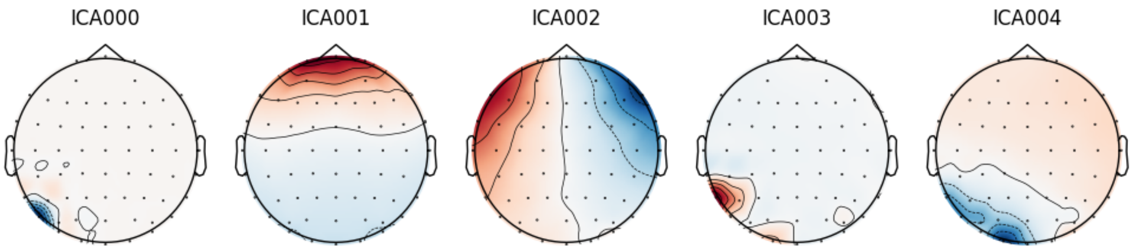

FIGURE 4.6: Example of ICA components.

6. **Baseline Removal**

   The baseline removal strategy was specific to each dataset. In SEED V and VII, raw files, which include the baseline and evaluation time, were available, making it possible to combine the baseline time of each video and perform direct removal by subtracting the average value per channel. On the other hand, in SEED I and IV, there was no explicit baseline, so a baseline approximation per session was implemented, concatenating the first 3 seconds of each neutral video and calculating the average value per channel. This average value was then used as the session baseline. This solution mitigates the lack of an explicit baseline and proved to be stable in practice.

After every step of the pipeline the signal shows clearly an improvement in terms of noise reduction. Figure 4.7 and 4.8 show the signal in the time domain and PSD signal after all the steps of pre-processing.

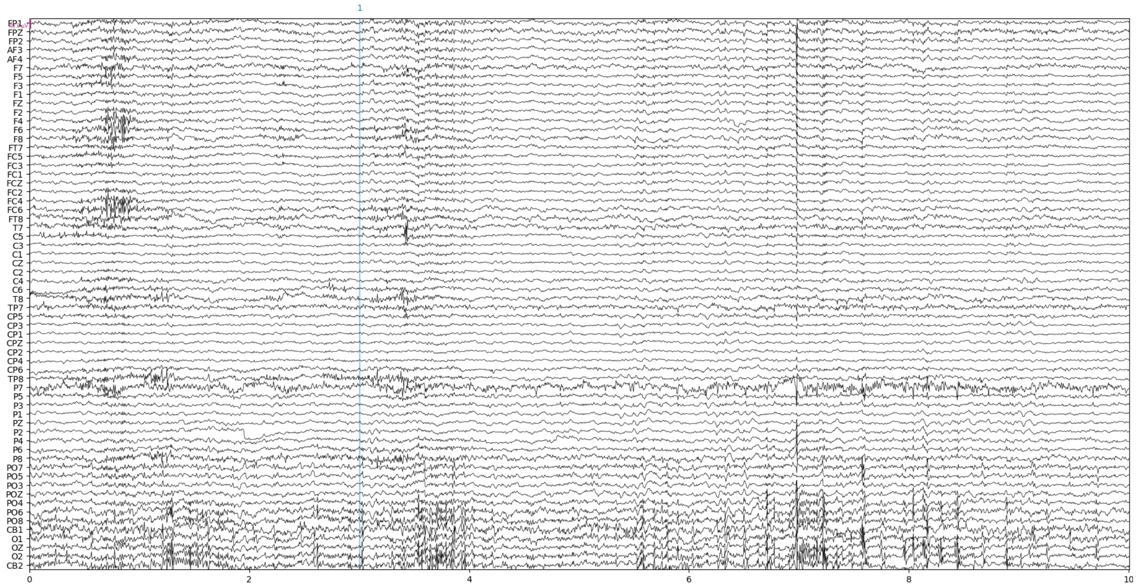

Figure 4.7: Signal in the Time Domain after pre-processing.

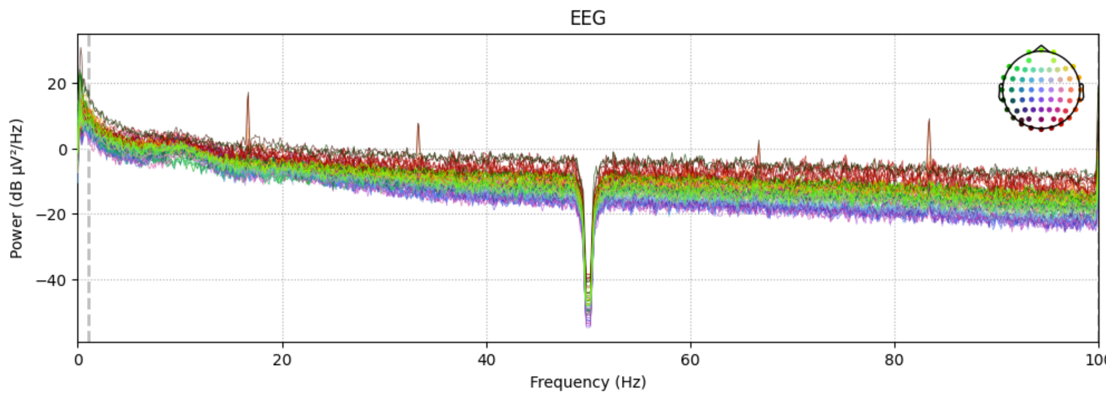

Figure 4.8: PSD after pre-processing.

An additional practical aspect is processing time, which varied substantially across workflows. SEED IV (pre-segmented) required on average 98.3 minutes per subject. By contrast, SEED VII was faster across all variants: 49.1 minutes per subject when the data were already segmented into per-video files, 42.9 minutes per subject when segmented first, and 47.4 minutes per subject when sessions were processed as continuous recordings. The comparatively slower runtime for SEED IV likely reflects overhead from repeatedly applying the full preprocessing chain to many short segments and possibly running processing "on top of" prior steps.

Furthermore, after pre-processing was completed, the EEG signals were organised into files ready for feature extraction and model training. The final space occupied by the data was approximately 7.05 GB in SEED I, 7.01 GB in SEED IV, 10.1 GB in SEED V, and 26 GB in SEED VII, which, compared to the original size in Table 3.1, shows a significant decrease in the space occupied (which, if the processed data were used directly to train the models, would mean a significant reduction in computational weight).

Overall, this pre-processing pipeline ensured the reduction of structural noise (namely from the electrical network), the reduction of physiological artefacts, and the reduction of space occupied and computational cost for the next steps. The differences between datasets, especially with regard to the availability of raw files and baselines, proved to be decisive in defining pre-processing strategies, reinforcing the need for high-quality data regardless of the purpose of the dataset.

4.3 Feature Extraction

After pre-processing, the features that served as the basis for training the deep learning models were extracted. Three main features were chosen: PSD, Differential Entropy (DE) and Wavelet Transform (WT), selected for their widespread use in EEG emotion recognition studies and for the possibility they offer to compare approaches in the frequency and time-frequency domains. It should be noted that, in line with the philosophy adopted in this work, in addition to not using the features extracted by the authors, the features used were extracted directly from the pre-processed signals (Section 4.2), and not from the data previously processed by the authors of the datasets. This point is particularly relevant, as it ensures greater methodological control and consistency between the different datasets, avoiding potential biases and ensuring comparability between experiments.

4.3.1 Strategies for feature extraction

A central aspect of feature extraction was the definition of the band schemes used. In addition to the classic division of the EEG (*delta, theta, alpha, beta* and *gamma*), three variations were considered:

- Four bands (4-45 Hz): *theta, alpha, beta* and low *gamma.*
- Five bands (1-45 Hz): inclusion of the *delta* band.
- Five bands (4-45 Hz + 55-70 Hz): inclusion of high *gamma*, but excluding the 45-55 Hz range, in order to avoid any residual contamination from the electrical network, even after applying the notch filter. This decision, although conservative, ensures greater robustness, preventing electrical noise energy from being confused with neuronal activity.

This scheme allows to evaluate the relevance of the *delta* and high *gamma* bands in the generalisation capacity of the models.

In addition, for each video, the data was segmented into 4 second windows, with a 2 second overlap between consecutive windows. This strategy ensures a greater number of samples for training, while reducing the risk of losing relevant temporal information.

Regardless of the type of feature considered, the final result was always organised in a vector with the format `(n_channels, n_windows, n_bands)`, ensuring structural consistency between different extraction methods and facilitating integration into deep learning models.

After feature extraction, the resulting data became significantly more compact and organised for use in model training. The final space occupied was approximately 357 MB in SEED I, 352 MB in SEED IV, 205 MB in SEED V, and 656 MB in SEED VII, showing a substantial reduction in size compared to the original EEG signals and even the EEG signals after processing.

4.3.2 PSD

PSD was calculated in each window and channel according to the defined band schemes. The Welch method was used for estimation, using the `psd_array_welch` function from the MNE-Python library. The parameters were defined with `n_fft = 256` and average in the segments, ensuring a stable estimate of the power spectral density. For each window-channel combination, the average power in each frequency band was obtained, resulting in vectors representative of the spectral energy distribution.

4.3.3 DE

DE was calculated from the same set of windows and bands, using the variance of the signal in each channel as the base measure. The differential entropy of a continuous Gaussian variable X with variance σ^2 is given by:

$$h(X) = \frac{1}{2} \log \left(2\pi e \, \sigma^2\right)$$

So, for each window and channel, the variance of the filtered signal was calculated, applied in the above formula, resulting in vectors of DE values organised by frequency band.

4.3.4 WT

To complement the representations in the frequency domain, features were also extracted in the time-frequency domain using the Wavelet Transform. For each channel, the Daubechies-4 wavelet (db4) was used, with four levels of decomposition, in order to align the frequency intervals obtained with the bands considered in this work and exclude the 50-100 Hz interval. The decomposition coefficients were converted into relative energies and normalised, resulting in vectors comparable to the PSD and DE representations.

4.4 Experimental Setup and Results

After pre-processing and feature extraction, several experiments were conducted to evaluate the impact of different extracted features, banding schemes, normalisation strategies, and evaluation methods on the generalisation ability of the models. All experiments were performed in Python, using TensorFlow/Keras to implement the architectures, using Leave-One-Subject-Out (LOSO) validation or the variants defined in each case, always avoiding data leakage situations. The evaluation was performed using the Accuracy, Precision, Recall, and F1 Score metrics, reported as mean and standard deviation per subject, ensuring maximum rigour and robustness in the analysis of the results.

4.4.1 Cross-Subject (SEED I e SEED IV)

The first set of experiments evaluated the ability to generalise across subjects, using only one session per participant. For this scenario, several combinations of features and band schemes were tested:

- CNN with DE (4 bands).
- CNN with DE (5 bands with *delta* included).
- CNN with DE (5 bands with high *gamma* included).
- CNN with PSD (5 bands with *delta* included).
- CNN with DE + PSD (5 bands with *delta* included).
- CNN with WT.

In addition, a reference test with CNN + DE (4 bands) was included on the signals processed by the authors (without applying the pipeline developed in this work), in order to quantify the difference in results and highlight the impact of the processing itself. Tables 4.1 and 4.2 show the values of the evaluation metrics for all the tests mentioned above for SEED I and SEED IV, respectively.

Table 4.1: Results for SEED I (mean ± std) by feature type, in percentage.

Feature and bands	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	Max Acc (%)
DE (4 bands-authors)	69.9 ± 13.2	71.7 ± 13.6	69.8 ± 13.2	69.6 ± 13.1	87.0
DE (4 bands)	69.9 ± 13.1	71.7 ± 13.6	69.9 ± 13.1	69.7 ± 13.1	87.0
DE (w/*delta*)	73.0 ± 12.1	71.4 ± 22.5	72.9 ± 12.2	72.8 ± 12.2	87.0
DE (w/high *gamma*)	70.6 ± 11.9	72.3 ± 12.4	70.6 ± 11.9	70.2 ± 12.3	87.0
PSD (w/*delta*)	71.1 ± 13.3	73.5 ± 13.0	71.0 ± 13.3	70.3 ± 13.4	93.3
DE + PSD (w/*delta*)	73.9 ± 13.8	77.3 ± 13.2	73.9 ± 13.8	72.9 ± 14.2	93.3
WT	60.6 ± 12.3	62.9 ± 15.4	60.4 ± 12.4	57.2 ± 12.1	80.0

Table 4.2: Results for SEED IV (mean ± std) by feature type, in percentage.

Feature and bands	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	Max Acc (%)
DE (4 bands-authors)	58.9 ± 8.7	58.5 ± 12.9	58.7 ± 8.6	55.0 ± 9.1	79.2
DE (4 bands)	64.0 ± 12.7	61.5 ± 16.3	63.8 ± 12.8	60.3 ± 14.2	96.0
DE (w/*delta*)	66.5 ± 11.6	64.0 ± 14.7	66.3 ± 11.6	62.5 ± 12.6	96.0
DE (w/high *gamma*)	60.0 ± 11.8	60.7 ± 14.6	59.8 ± 11.8	57.3 ± 12.2	79.2
PSD (w/*delta*)	64.0 ± 12.6	60.5 ± 15.2	63.7 ± 12.6	60.1 ± 14.0	96.0
DE + PSD (w/*delta*)	65.9 ± 11.9	65.3 ± 14.9	65.7 ± 12.0	63.5 ± 13.3	96.0
WT	57.8 ± 13.8	59.0 ± 15.2	57.7 ± 13.8	55.1 ± 14.3	83.3

Analysis of the impact of bands and the choice of the most robust features

Based on the results obtained, a specific analysis was carried out on the impact of including the *delta* band and high *gamma*:

- *Delta* band: its inclusion revealed a significant impact on the generalisation capacity of the models.
- High *gamma* band: it did not present clear benefits and, in some scenarios, reduced performance.

In addition, it was also possible to analyse that the most robust combination was the simultaneous use of DE and PSD, surpassing the other alternatives. This analysis is particularly important for the future tests carried out in the section.

4.4.2 Different types of evaluation

The effect of the type of evaluation used was also analysed, using DE+PSD with the inclusion of the *delta* band, as this had previously shown the best results. To this end, three different approaches were compared in the SEED I and SEED IV datasets:

- Cross-Subject (Single-Session): training with multiple subjects, testing on unseen subjects. In this case, training with 14 subjects and testing with 1 subject.
- Cross-Session: training with two sessions for each subject, testing with unseen sessions for each subject.
- Cross-Subject (Multi-Session): training with all sessions of multiple subjects, testing in all sessions of an excluded subject.

In the most challenging evaluation scenario (Cross-Subject with multiple sessions), as well as in cross session, the impact of the normalisation strategy was also tested:

- Normalisation by subject.
- Normalisation by session.

In the tables of section 4.4.1, all results refer to the Cross Subject (Single-Session) assessment type. Table 4.3 shows the Cross Subject (Multi-Session) assessment and the impact that different types of normalisation have. On the other hand, table 4.4 shows the results for cross Session.

Table 4.3: Cross Subject (Multi-Session): results (mean ± std) per dataset and normalisation strategy, in percentage.

Dataset	Normalization	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	Max Acc (%)
SEED I	Subject-wise	66.9 ± 8.4	68.1 ± 7.5	66.7 ± 8.3	66.2 ± 8.4	84.4
SEED I	Session-wise	71.7 ± 8.1	72.1 ± 8.5	71.7 ± 8.1	70.8 ± 8.3	89.0
SEED IV	Subject-wise	50.0 ± 9.9	54.3 ± 8.8	49.9 ± 9.8	48.4 ± 11.0	65.3
SEED IV	Session-wise	59.1 ± 7.8	60.5 ± 8.6	58.9 ± 7.8	58.2 ± 8.1	81.0

Table 4.4: Cross-Session: results (mean ± std) per dataset and normalisation strategy, in percentage.

Dataset	Normalization	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	Max Acc (%)
SEED I	Subject-wise	69.2 ± 3.0	69.3 ± 3.3	69.0 ± 2.9	68.3 ± 3.4	73.3
SEED I	Session-wise	69.8 ± 1.9	70.7 ± 1.7	69.7 ± 1.7	69.3 ± 2.9	72.4
SEED IV	Subject-wise	56.7 ± 2.7	58.0 ± 3.7	56.7 ± 2.6	55.3 ± 2.5	59.2
SEED IV	Session-wise	56.9 ± 3.3	58.0 ± 2.2	56.7 ± 3.4	56.7 ± 3.3	60.0

4.4.3 Comparison of Model Architectures

Beyond the base CNN architecture, deep learning alternatives were evaluated to understand which type of model has the greatest potential for emotion recognition from EEG signals when used in base configurations and without any fine tuning. For these experiments, a combination of DE+PSD features was used, including the *delta* band, as this proved to be the most robust approach in previous tests.

In order to reduce computational complexity and training time, the evaluation was conducted only in the Cross-Subject (Single-Session) scenario. This choice ensured an adequate compromise between generalisation requirements and practicability, allowing the architectures to be compared under controlled conditions.

In short, the architectures tested were as follows:

- CNN, taking advantage of spatial correlations between channels and the distribution of energy between bands.
- LSTM, taking advantage of its ability to model temporal dependencies between successive windows of the EEG signal.
- CNN-LSTM, which combines spatial convolutions with a LSTM layer, integrating the advantages of the two previous approaches and constituting, theoretically, the most balanced solution.

Model Details

All implemented architectures share the same final classification structure. This consists of a dense layer with 64 units, ReLU activation function and L2 regularisation with a value of 1e-4, followed by a dropout layer with a rate of 0.5, culminating in a softmax layer with the existing number of classes. The difference between the models lies in how they process the data before reaching this final stage.

In the case of CNN, the model starts with a convolution of 16 1×1 filters with ReLU activation and a Layer Normalisation layer, followed by a spatial block consisting of a convolution of 32 3×3 filters also with ReLU and another normalisation layer, complemented by a 2×2 MaxPooling layer to reduce the spatial resolution. A second spatial block applies a convolution of 64 3×3 filters with ReLU activation and normalisation. The convolutional stage ends with a Global Average Pooling layer, responsible for condensing the information before it is passed on to the dense layers.

The pure LSTM receives window sequences, which are initially reformatted into vectors, one for each time step. This sequence is processed by an LSTM layer with 128 units and configured with dropout (0.2), which allows temporal dependencies between windows to be captured.

CNN-LSTM combines the two previous approaches. The convolutional part is identical to CNN, but each layer is wrapped in a TimeDistributed operation, ensuring that each window is processed independently. Instead of Global Average Pooling, a Flatten is applied. The resulting vectors feed a 128-unit LSTM layer, which is responsible for integrating temporal information before the final layers.

The objective of this comparison was not to achieve the absolute best performance, but rather to explore how each type of architecture behaves when faced with this type of data and volume of information. This analysis provides an initial perspective on the overall suitability of the architectures and their limitations when applied to the problem of emotion recognition in EEG.

It is also worth noting that CNN stands out for its simplicity, not only in terms of computational requirements, but also in terms of implementation. While CNN allows independent windows to be used directly as training units, memory-based architectures (LSTM and CNN-LSTM) require the construction of window sequences. In this case, sequences of 10 windows with an overlap of 5 were used in order to increase the amount of available data and capture longer temporal dependencies which is why, in CNN-LSTM, the CNN layers were wrapped in TimeDistributed.

The results of this comparison of architectures are summarised in Table 4.5. For reference, the values corresponding to CNN are already shown in the DE + PSD w/*Delta* rows of tables 4.1 and 4.2, presented previously. The results obtained for the LSTM and CNN-LSTM architectures are now displayed, in order to allow for a clear comparative analysis.

Table 4.5: Cross-Subject (1 session): comparison of models (mean ± std), in percentage.

Dataset	Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	Max Acc (%)
SEED I	LSTM	71.7 ± 8.6	75.1 ± 7.3	71.5 ± 8.6	70.9 ± 9.0	87.0
SEED I	CNN-LSTM	72.1 ± 10.9	76.1 ± 9.8	71.9 ± 11.0	71.3 ± 11.4	87.0
SEED IV	LSTM	69.0 ± 11.2	68.5 ± 13.6	68.9 ± 11.2	67.1 ± 11.8	92.0
SEED IV	CNN-LSTM	72.1 ± 10.9	76.1 ± 9.8	71.9 ± 11.0	71.3 ± 11.4	87.0

4.4.4 Evaluation in SEED V and SEED VII

In the SEED V and SEED VII datasets, a separate experiment was conducted, focusing on the segmentation order:

- Segmentation in videos before pre-processing.
- Segmentation in videos after pre-processing (but before baseline).

The objective was to see if the order of these steps affects how well the models preserve information and perform. Table 4.6 shows the results of that experiment.

Table 4.6: Comparison between results with continuous and segmented processing by video (mean ± standard deviation), in percentage.

Processing	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	Max Acc (%)
Continuous	54.5 ± 11.2	55.2 ± 15.7	54.5 ± 11.2	50.9 ± 12.3	90.0
Segmented	52.3 ± 8.3	55.2 ± 14.4	52.3 ± 8.3	48.5 ± 9.7	70.0

4.5 Critical discussion and challenges

The objective of this study was not merely to obtain the best accuracy values, but rather to understand how methodological decisions, including pre-processing, normalisation, feature selection, and evaluation type, influence the performance and generalisation capacity of the models. The results obtained allow to observe relevant trends and identify the factors that most influence the classification of emotional states in EEG.

Overall, it was found that model performance depends heavily on the type of features used, the band scheme selected and, particularly, how the chosen evaluation method reflects generalisation. Experiments such as Cross-Subject and, especially, Cross-Subject (Multi-session) proved to be more challenging, but also more representative of real-world generalisation scenarios. This point highlights the importance of evaluating models under conditions that avoid data leakage, a practice still common in part of the literature and which can lead to artificially high results but without practical validity. For example, part of the literature continues to adopt classic divisions such as 80% for training and 20% for testing, without taking into account the separation between subjects or sessions. Although these procedures often result in high accuracy values, such results are methodologically invalid and lack practical utility, since the models end up training and testing on highly correlated data. In contrast, approaches such as LOSO validation, despite having lower accuracies, provide a much more reliable assessment of the ability to generalise across individuals.

Despite methodological efforts, the results are inevitably limited by the quality of the available data. The SEED family datasets, although widely used in emotion recognition research, have several limitations that should be considered when interpreting the results:

- Label noise: the labels assigned to emotions may not correspond to each subject's actual emotional experience. This problem is common to all emotion datasets, but it is especially relevant in SEED because labels are imposed based on the video shown, assuming that all participants experience the same emotion. Although SEED VII is a step forward in providing the intensity reported by the subject for the dominant emotion in each video, it still fails to capture the intensity of the other emotions. Ideally, the relative intensity of all emotions for each video should be available, even segmented into windows, reflecting emotional variability over time (this is already available in SEED VII but only for the emotion considered dominant by the authors). Such granularity would allow, for example, the identification of specific moments of greater emotional response and the exclusion of neutral or irrelevant segments. Figure A.1 illustrates the extent of label noise, highlighting in red the videos that participants rated with an intensity of 0.5 or lower (on a scale from 0 to 1). Figure reveals that, in some cases, participants almost unanimously disagreed with the assigned label.

- Reduced or absent baseline: in some datasets, such as SEED I and SEED IV, no explicit baseline was provided, or it was too short to be representative. This required the adoption of alternative strategies, such as constructing approximate baselines from the first few seconds of neutral videos. Although effective, this approach does not replace the robustness of a controlled baseline, introducing limitations in comparability between subjects and sessions.

- Variable video length: the variability in the length of audiovisual stimuli poses a methodological challenge. Models that receive windows or sequences of variable length as input may inadvertently learn the length of the video rather than the neural content. This situation was experienced during the development of this work, resulting in a significant loss of time with initial padding-based approaches. This experience highlighted the need to carefully control the bias associated with duration, ensuring that models do not exploit information irrelevant to the classification task.

In addition to these structural limitations, individual variability in emotional perception should also be highlighted. The same video may be considered funny by one participant and annoying by another, which challenges the idea that there is a homogeneous emotional response for all subjects. This problem reinforces the need for personalised and continuous labels that can reflect interindividual differences more reliably.

Taken together, these challenges highlight that the results presented here should be interpreted in light of the limitations of the datasets, but they also demonstrate the value of the approach taken: rather than seeking absolute accuracy values, this work aimed to understand the influence of methodological choices on the robustness and generalisation of the models. This critical perspective constitutes a relevant contribution to the advancement of research in emotion recognition from EEG, drawing attention to the importance of data quality, the rigorous definition of baselines, and the need for more detailed and individualised emotional annotations.

Chapter 5

Conclusion

This project aimed to explore the potential of Electroencephalogram (EEG) signals captured by Brain-Computer Interfaces (BCIs) for emotion recognition in the context of audiovisual content consumption. The research carried out established a solid bridge between the theoretical framework of affective computing and the practical application of signal processing and Deep Learning (DL) methodologies.

Throughout the study, it was possible to demonstrate that methodological decisions, from pre-processing, through feature selection, to normalisation and evaluation strategies, significantly influence the generalisation capacity of models. Rather than maximising isolated performance metrics, this work emphasised the critical analysis of factors that condition the robustness of results, providing relevant contributions to the construction of reliable EEG-based emotion recognition systems.

The results obtained show that the inclusion of the *delta* band and the combination of different types of characteristics, namely Differential Entropy (DE) and Power Spectral Density (PSD), lead to significant generalisation improvements. On the other hand, it was found that approaches often reported in the literature, but methodologically fragile, such as simple training and testing divisions, can lead to artificially high accuracy values with no practical validity. The work carried out reinforces the importance of using rigorous evaluation methodologies, such as Leave-One-Subject-Out (LOSO) validation, especially in contexts of inter- and intra-subject variability.

Despite the advances achieved, the study also exposed structural limitations of the SEED datasets, namely the noise associated with labels, the absence or reduced duration of baselines, and the variability in video duration. These factors condition comparability between subjects and introduce biases that require creative solutions, such as those implemented in this work, but also point to the need for richer and more realistic datasets.

In terms of overall contribution, this work is not limited to the academic scope of the Master's Degree in Artificial Intelligence Engineering. It also constitutes the technical core of the European DataPACT project, developed in partnership with EVS Broadcast Equipment, highlighting the scientific relevance and practical impact of this research. The integration of EEG and BCI in emotion recognition opens the way for future applications in areas as diverse as entertainment, mental health and the personalisation of digital experiences, contributing to the advancement of more intelligent, ethical and user-centred systems.

5.1 Research Questions and Objectives Achieved

The work developed in this dissertation was guided by a set of previously defined research questions (RQs) and objectives (O1-O8), which served as a basis for structuring the research and evaluating its progress. This section reflects on how each question was addressed and the extent to which the objectives were achieved, relating them directly to the scientific and practical contributions obtained.

The first research question (RQ1) was posed broadly: How can EEG signals captured by BCIs be used to identify human emotions and develop reliable models for emotion recognition during audiovisual content consumption? To answer it, a complete pipeline of pre-processing, feature extraction, and deep learning model training was implemented, demonstrating that, even though the task is complex and limited by data quality, it is possible to achieve consistent performance. This question was covered by achieving objectives O3 ("To explore how EEG signals, captured by BCIs, can be used to identify emotional states and to investigate the most relevant characteristics and features to the study."), O5 ("Analysing and pre-processing the datasets relevant to the study, including SEED datasets, in order to understand and prepare the data for emotion recognition."), and O7 ("Develop a functional emotion recognition system based on EEG signals, applying AI, signal processing, deep learning techniques to identify emotional states.").

Associated with RQ1 was the sub-question RQ1.1: How can EEG-based systems be tested and evaluated in a practical and objective way to guarantee their effectiveness in identifying emotions in the media context? This was addressed by analysing different evaluation strategies, including Cross-Subject (Single-Session), Cross-Session and Cross-Subject (Multi-Session). It was found that classic training and testing divisions, such as the 80/20 division, lead to artificially high metrics, but without practical validity, reinforcing the importance of rigorous methodologies such as LOSO validation. This sub-question was answered with the help of fulfilling objectives O6 ("Evaluate the impact of different methodological decisions such as types of evaluation, normalisation strategies, combinations of features and model architectures on the robustness and generalisation capacity of the models developed.") and O8 ("Test and validate the system developed in order to evaluate it in terms of generalisation and accuracy in detecting emotions.").

The second question (RQ2) related towards the state of the art: What are the existing approaches and challenges in emotion recognition from EEG signals captured by BCIs? The literature review identified widely used pre-processing techniques, features and AI architectures, as well as recurring limitations. Among the transversal challenges, the subjectivity of emotions stands out, which, as internal and individual experiences, can be expressed and perceived differently between subjects. This issue was addressed essentially through the achievement of objective O2 ("Analyse the most relevant approaches for signal pre-processing, feature extraction, and deep learning models in EEG-based emotion recognition.").

To further explore RQ2, three sub-questions were formulated. The first (RQ2.1) asked: What AI models are used for emotion recognition using EEG signals? The literature review identified mainly models based on CNNs, RNNs, LSTMs, GRUs and hybrid models. In addition, during the development of this work, CNNs, LSTMs, and CNN-LSTMs were tested in order to compare them. This sub-question was addressed by achieving objectives O2 ("Analyse the most relevant approaches for signal pre-processing, feature extraction, and deep learning models in EEG-based emotion recognition.") and O7 ("Develop a functional emotion recognition system based on EEG signals, applying AI, signal processing, deep learning techniques to identify emotional states.").

The second sub-question (RQ2.2) asked: Which features of EEG signals can be most effective in identifying/recognising emotions? In the state of the art, several features from different domains were listed and explained, and in the development of the work, DE, PSD, and WT were selected and tested in various frequency band schemes. It was concluded that the inclusion of the *delta* band contributed positively to generalisation, while high *gamma* did not bring significant benefits. This sub-question was answered in conjunction with objectives O2 ("Analyse the most relevant approaches for signal preprocessing, feature extraction, and deep learning models in EEG-based emotion recognition."), O3 ("To explore how EEG signals, captured by BCIs, can be used to identify emotional states and to investigate the most relevant characteristics and features to the study.") and O6 ("Evaluate the impact of different methodological decisions such as types of evaluation, normalisation strategies, combinations of features and model architectures on the robustness and generalisation capacity of the models developed.").

The third sub-question (RQ2.3) asked: How can the use of EEG signals overcome the limitations of traditional emotion recognition methods, such as facial recognition or questionnaires? This question was fully addressed in the state of the art, where, for example, the possibility of traditional methods being deliberately manipulated and depending on context and culture, as well as temporal limitations, was analysed. In contrast, EEG was highlighted as a more objective approach capable of recording brain activity directly associated with emotional responses, thus overcoming restrictions of subjectivity and temporality, and using EEG signals it is also possible to detect emotions in people with physical disabilities. The results in this dissertation only confirm that EEG can indeed be used successfully, but the answer to the question is based mainly on the literature review. This sub-question was addressed through the achievement of objective O1 ("Compare the approach based on BCIs and EEG signals with traditional methods of analysing emotions, such as subjective questionnaires and facial recognition, assessing how BCIs can overcome the limitations of these techniques.").

The third main question (RQ3) focused on technical and ethical challenges: What are the main technical and ethical challenges associated with using EEGs for emotion recognition in the consumption of audiovisual content? This question was explored on two levels. Chapter 3 discussed the ethical and legal framework, including the GDPR and the AI Act, and produced a scientific article focusing on the ethical and legal landscape that presented a technical-ethical guide for the collection and use of EEG data. Chapter 4 analysed technical challenges such as label noise, the absence of consistent baselines, and the variability of video duration. This question was answered by achieving objectives O4 ("Identify the main technical and ethical challenges related to the use of BCIs for emotion recognition, including privacy, data security, individual variability and generalisation of AI models.") and O6 ("Evaluate the impact of different methodological decisions (types of evaluation, normalisation strategies, combinations of features and model architectures) on the robustness and generalisation capacity of the models developed.").

In conclusion, all research questions were addressed and the objectives achieved. The work carried out was not limited to producing quantitative results, but also provided an analysis of the limitations of the datasets, the methodological risks and the ethical implications associated with the use of EEG. The combination of these dimensions is one of the central contributions of this dissertation, reinforcing its scientific, practical and social relevance.

5.2 Future Work

The complexity of recognising emotions from EEG makes it inevitable that several aspects remain unexplored, so this work should be understood as a solid basis for future research, including continuity already ensured through a doctoral project.

One of the most promising directions is the application of transfer learning techniques, which allow knowledge acquired in different contexts or datasets to be reused to improve generalisation in new scenarios as well as to reduce the dependence on large volumes of data and accelerate the achievement of consistent results.

Another very important aspect concerns the quality of the datasets. The experience gained in this work has shown that individual variability in emotional perception cannot be ignored: the same video can be interpreted, for example, as funny by one subject and as annoying by another. To overcome this limitation, it would be desirable for future data collections to include more comprehensive subjective assessments, considering not only the intensity of the dominant emotion, but also intensity distributions across multiple emotions. In addition, a longer and more consistent baseline is essential, as segments of only three seconds or less may not offer sufficient robustness for adequate baseline removal.

From a methodological point of view, an interesting opportunity would be to use signals in the time domain directly as input for models. Although this approach requires significantly higher computational power, it could open the way for models that capture complex patterns without the need for explicit feature extraction.

Another line of future research would be to analyse the impact of feature smoothing techniques. The authors of the SEED datasets apply this type of processing in some variants, and it would be interesting to investigate how feature smoothing influences not only performance but also the generalisation capacity of the models.

Finally, it is particularly important to note that this work will serve as a basis for and will be continued in a doctoral project already planned, in which new EEG data will be collected, taking into account several of the suggestions presented here, including the need for robust baselines, continuous and individualised assessment of emotions, and the creation of datasets that better reflect interindividual variability. The fact that this work will give grounds for doctoral research reinforces its scientific relevance and demonstrates the impact it may have on the evolution of the state of the art in emotion recognition from EEG.

Bibliography

Abhang, Priyanka A., Bharti W. Gawali, and Suresh C. Mehrotra (2016a). "Chapter 1 - Introduction to Emotion, Electroencephalography, and Speech Processing". In: *Introduction to EEG- and Speech-Based Emotion Recognition*. Academic Press, pp. 1–17. ISBN: 978-0-12-804490-2. DOI: https://doi.org/10.1016/B978-0-12-804490-2.00001-4. URL: https://www.sciencedirect.com/science/article/pii/B9780128044902000014.

– (2016b). "Chapter 2 - Technological Basics of EEG Recording and Operation of Apparatus". In: *Introduction to EEG- and Speech-Based Emotion Recognition*. Ed. by Priyanka A. Abhang, Bharti W. Gawali, and Suresh C. Mehrotra. Academic Press, pp. 19–50. ISBN: 978-0-12-804490-2. DOI: https://doi.org/10.1016/B978-0-12-804490-2.00002-6. URL: https://www.sciencedirect.com/science/article/pii/B9780128044902000026.

– (2016c). "Chapter 3 - Technical Aspects of Brain Rhythms and Speech Parameters". In: *Introduction to EEG- and Speech-Based Emotion Recognition*. Academic Press, pp. 51–79. ISBN: 978-0-12-804490-2. DOI: https://doi.org/10.1016/B978-0-12-804490-2.00003-8. URL: https://www.sciencedirect.com/science/article/pii/B9780128044902000038.

– (2016d). "Chapter 5 - Emotion Recognition". In: *Introduction to EEG- and Speech-Based Emotion Recognition*. Ed. by Priyanka A. Abhang, Bharti W. Gawali, and Suresh C. Mehrotra. Academic Press, pp. 97–112. ISBN: 978-0-12-804490-2. DOI: https://doi.org/10.1016/B978-0-12-804490-2.00005-1. URL: https://www.sciencedirect.com/science/article/pii/B9780128044902000051.

Alarcão, Soraia M. and Manuel J. Fonseca (2019). "Emotions Recognition Using EEG Signals: A Survey". In: *IEEE Transactions on Affective Computing* 10.3, pp. 374–393. DOI: 10.1109/TAFFC.2017.2714671.

Assunção, Gustavo, Bruno Patrão, Miguel Castelo-Branco, and Paulo Menezes (2022). "An Overview of Emotion in Artificial Intelligence". In: *IEEE Transactions on Artificial Intelligence* 3.6, pp. 867–886. DOI: 10.1109/TAI.2022.3159614.

Bakker, Iris, Theo van der Voordt, Peter Vink, and Jan de Boon (Apr. 2014). "Pleasure, Arousal, Dominance: Mehrabian and Russell revisited". In: *Current Psychology* 33.3, pp. 405–421. URL: https://doi.org/10.1007/s12144-014-9219-4.

Bartneck, Christoph (2002). "Integrating the occ model of emotions in embodied characters". In.

Berger, Hans (Dec. 1929). "Über das Elektrenkephalogramm des Menschen". In: *Archiv für Psychiatrie und Nervenkrankheiten* 87.1, pp. 527–570. ISSN: 1433-8491. DOI: 10.1007/BF01797193. URL: https://doi.org/10.1007/BF01797193.

Bhise, Pratibha R., Sonali B. Kulkarni, and Talal A. Aldhaheri (2020). "Brain Computer Interface based EEG for Emotion Recognition System: A Systematic Review". In: *2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)*, pp. 327–334. DOI: 10.1109/ICIMIA48430.2020.9074921.

Clore, Gerald L and Andrew Ortony (2013). “Psychological construction in the OCC model of emotion”. In: *Emotion Review* 5.4, pp. 335–343.

Consulting, Intersoft (2016). *General Data Protection Regulation.* Last Accessed: 2025-02-23. URL: `https://gdpr-info.eu/`.

Dey, Rahul and Fathi M. Salem (2017). “Gate-variants of Gated Recurrent Unit (GRU) neural networks”. In: *2017 IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS)*, pp. 1597–1600. DOI: `10.1109/MWSCAS.2017.8053243`.

Duan, Ruo-Nan, Jia-Yi Zhu, and Bao-Liang Lu (2013). “Differential entropy feature for EEG-based emotion classification”. In: *6th International IEEE/EMBS Conference on Neural Engineering (NER).* IEEE, pp. 81–84.

Egger, Maria, Matthias Ley, and Sten Hanke (2019). “Emotion Recognition from Physiological Signal Analysis: A Review”. In: *Electronic Notes in Theoretical Computer Science* 343. The proceedings of AmI, the 2018 European Conference on Ambient Intelligence., pp. 35–55. ISSN: 1571-0661. DOI: `https://doi.org/10.1016/j.entcs.2019.04.009`. URL: `https://www.sciencedirect.com/science/article/pii/S157106611930009X`.

Ekman, Paul and Wallace V. Friesen (1976). *Pictures of Facial Affect.* Consulting Psychologists Press.

Elf, Machine (2011). *Robert Plutchik’s Wheel of Emotions.* accessed 20 Jan 2025. URL: `https://en.wikiversity.org/wiki/Motivation_and_emotion/Book/2014/Plutchik%27s_wheel_of_emotions#/media/File:Plutchik-wheel.svg`.

ENISA (2019). *Pseudonymisation techniques and best practices.* URL: `https://www.enisa.europa.eu/publications/pseudonymisation-techniques-and-best-practices` (visited on 08/27/2025).

Erat, Kübra, Elif Bilge Şahin, Furkan Doğan, Nur Merdanoğlu, Ahmet Akcakaya, and Pınar Onay Durdu (2024). “Emotion recognition with EEG-based brain-computer interfaces: a systematic literature review.” In: *Multimed Tools Appl.* DOI: `https://doi.org/10.1007/s11042-024-18259-z`.

European Data Protection Board (2020). *Guidelines 05/2020 on consent under Regulation 2016/679.* EDPB. URL: `https://edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-052020-consent-under-regulation-2016679_en` (visited on 08/27/2025).

– (2025). *Guidelines 01/2025 on Pseudonymisation.* EDPB. URL: `https://www.edpb.europa.eu/our-work-tools/documents/public-consultations/2025/guidelines-012025-pseudonymisation_en` (visited on 08/28/2025).

Gremsl, Thomas and Elisabeth Hödl (Apr. 2022). “Emotional AI: Legal and ethical challenges”. In: *Information Polity* 27, pp. 1–12. DOI: `10.3233/IP-211529`.

Gunda, Naresh Kumar, Mohammed I. Khalaf, Shaleen Bhatnagar, Aadam Quraishi, Leeladhar Gudala, Ashok Kumar Pamidi Venkata, Faisal Yousef Alghayadh, Shtwai Alsubai, and Vaibhav Bhatnagar (2024). “Lightweight attention mechanisms for EEG emotion recognition for brain computer interface”. In: *Journal of Neuroscience Methods* 410, p. 110223. ISSN: 0165-0270. DOI: `https://doi.org/10.1016/j.jneumeth.2024.110223`. URL: `https://www.sciencedirect.com/science/article/pii/S0165027024001687`.

Hamzah, Hussein Ali and Kasim K. Abdalla (2024). “EEG-based emotion recognition systems; comprehensive study”. In: *Heliyon* 10.10, e31485. ISSN: 2405-8440. DOI: `https://doi.org/10.1016/j.heliyon.2024.e31485`. URL: `https://www.sciencedirect.com/science/article/pii/S2405844024075169`.

Hartmann, Kris Vera, Giovanni Rubeis, and Nadia Primc (Jan. 2024). "Healthy and Happy? An Ethical Investigation of Emotion Recognition and Regulation Technologies (ERR) within Ambient Assisted Living (AAL)". In: *Science and Engineering Ethics* 30.1, p. 2. ISSN: 1471-5546. DOI: 10.1007/s11948-024-00470-8. URL: https://doi.org/10.1007/s11948-024-00470-8.

Houssein, Essam H., Asmaa Hammad, and Abdelmgeid A. Ali (2022). "Human emotion recognition from EEG-based brain–computer interface using machine learning: a comprehensive review." In: *Neural Comput & Applic.* DOI: https://doi.org/10.1007/s00521-022-07292-4.

Hramov, Alexander E., Vladimir A. Maksimenko, and Alexander N. Pisarchik (2021). "Physical principles of brain–computer interfaces and their applications for rehabilitation, robotics and control of human brain states". In: *Physics Reports* 918. Physical principles of brain–computer interfaces and their applications for rehabilitation, robotics and control of human brain states, pp. 1–133. ISSN: 0370-1573. DOI: https://doi.org/10.1016/j.physrep.2021.03.002. URL: https://www.sciencedirect.com/science/article/pii/S0370157321001095.

Huang, Haiyun, Qiuyou Xie, Jiahui Pan, Yanbin He, Zhenfu Wen, Ronghao Yu, and Yuanqing Li (Feb. 2019). "An EEG-Based Brain Computer Interface for Emotion Recognition and Its Application in Patients with Disorder of Consciousness". In: *IEEE Transactions on Affective Computing* PP, pp. 1–1. DOI: 10.1109/TAFFC.2019.2901456.

Islam, Md. Rabiul, Mohammad Ali Moni, Md. Milon Islam, Md. Rashed-Al-Mahfuz, Md. Saiful Islam, Md. Kamrul Hasan, Md. Sabir Hossain, Mohiuddin Ahmad, Shahadat Uddin, Akm Azad, Salem A. Alyami, Md. Atiqur Rahman Ahad, and Pietro Lió (2021). "Emotion Recognition From EEG Signal Focusing on Deep Learning and Shallow Learning Techniques". In: *IEEE Access* 9, pp. 94601–94624. DOI: 10.1109/ACCESS.2021.3091487.

Iyer, Abhishek, Srimit Sritik Das, Reva Teotia, Shishir Maheshwari, and Rishi Raj Sharma (Feb. 2023). "CNN and LSTM based ensemble learning for human emotion recognition using EEG recordings". In: *Multimedia Tools and Applications* 82.4, pp. 4883–4896. DOI: 10.1007/s11042-022-12310-7. URL: https://doi.org/10.1007/s11042-022-12310-7.

Jiang, Wei-Bang, Xuan-Hao Liu, Wei-Long Zheng, and Bao-Liang Lu (2024). "SEED-VII: A Multimodal Dataset of Six Basic Emotions with Continuous Labels for Emotion Recognition". In: *IEEE Transactions on Affective Computing*, pp. 1–16. DOI: 10.1109/TAFFC.2024.3485057.

Kalaganis, Fotis, Dimitrios Adamos, and Nikos Laskaris (Sept. 2017). "Musical NeuroPicks: A consumer-grade BCI for on-demand music streaming services". In: *Neurocomputing* 280, pp. 65–75. DOI: 10.1016/j.neucom.2017.08.073.

Katirai, Amelia (June 2023). "Ethical considerations in emotion recognition technologies: a review of the literature". In: *AI and Ethics* 4, pp. 927–948. DOI: 10.1007/s43681-023-00307-3.

Lainscsek, Claudia (2006). *Automatic Recognition of Facial Actions in Spontaneous Expressions - Scientific Figure on ResearchGate.* accessed 20 Jan 2025. URL: https://www.researchgate.net/figure/Example-FACS-codes-for-a-prototypical-expression-of-fear-Spontaneous-expressions-may_fig8_42803851.

Li, Jinpeng, Shuang Qiu, Yuan-Yuan Shen, Cheng-Lin Liu, and Huiguang He (2020). "Multisource Transfer Learning for Cross-Subject EEG Emotion Recognition". In: *IEEE Transactions on Cybernetics* 50.7, pp. 3281–3293. DOI: 10.1109/TCYB.2019.2904052.

Li, Jinyu, Haoqiang Hua, Zhihui Xu, Lin Shu, Xiangmin Xu, Feng Kuang, and Shibin Wu (2022). “Cross-subject EEG emotion recognition combined with connectivity features and meta-transfer learning”. In: *Computers in Biology and Medicine* 145, p. 105519. ISSN: 0010-4825. DOI: https://doi.org/10.1016/j.compbiomed.2022.105519. URL: https://www.sciencedirect.com/science/article/pii/S0010482522003110.

Li, Tian-Hao, Wei Liu, Wei-Long Zheng, and Bao-Liang Lu (2019). “Classification of Five Emotions from EEG and Eye Movement Signals: Discrimination Ability and Stability over Time”. In: *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*, pp. 607–610. DOI: 10.1109/NER.2019.8716943.

Li, Wei, Wei Huan, Bowen Hou, Ye Tian, Zhen Zhang, and Aiguo Song (2022). “Can Emotion Be Transferred?—A Review on Transfer Learning for EEG-Based Emotion Recognition”. In: *IEEE Transactions on Cognitive and Developmental Systems* 14.3, pp. 833–846. DOI: 10.1109/TCDS.2021.3098842.

Liu, Wei, Jie-Lin Qiu, Wei-Long Zheng, and Bao-Liang Lu (2021). “Comparing Recognition Performance and Robustness of Multimodal Deep Learning Models for Multimodal Emotion Recognition”. In: *IEEE Transactions on Cognitive and Developmental Systems*.

Martinez, Brais and Michel F. Valstar (2016). “Advances, Challenges, and Opportunities in Automatic Facial Expression Recognition”. In: *Advances in Face Detection and Facial Image Analysis*. Ed. by Michal Kawulok, M. Emre Celebi, and Bogdan Smolka. Cham: Springer International Publishing, pp. 63–100. ISBN: 978-3-319-25958-1. DOI: 10.1007/978-3-319-25958-1_4. URL: https://doi.org/10.1007/978-3-319-25958-1_4.

Mühl, Christian, Brendan Allison, Anton Nijholt, and Guillaume Chanel (June 2014). “A survey of affective brain computer interfaces: Principles, state-of-the-art, and challenges”. In: *Brain-Computer Interfaces* 1, pp. 66–84. DOI: 10.1080/2326263X.2014.912881.

Al-Nafjan, Abeer, Manar Hosny, Yousef Al-Ohali, and Areej Al-Wabil (2017). “Review and Classification of Emotion Recognition Based on EEG Brain-Computer Interface System Research: A Systematic Review”. In: *Applied Sciences* 7.12. ISSN: 2076-3417. DOI: 10.3390/app7121239. URL: https://www.mdpi.com/2076-3417/7/12/1239.

OpenBCI (2019). *OpenBCI Research Collection*. URL: https://docs.google.com/spreadsheets/d/1WvolD2-QJ5aUJy5o0Dq5wdFQtLMkMtppZT8s_ihYyA4/edit (visited on 09/22/2025).

Pais-Vieira, Miguel Santos and Carla Ferreira da Silva Pais-Vieira (2018). *Prevenção da Corrupção e Infrações Conexas Associadas à Utilização das Interfaces Cérebro-Máquina*. Lisboa, Portugal: Conselho de Prevenção da Corrupção. ISBN: 978-989-20-9537-0.

Pernet, Cyril, Marta I. Garrido, Alexandre Gramfort, Natasha Maurits, Christoph M. Michel, Elizabeth Pang, Riitta Salmelin, Jan Mathijs Schoffelen, Pedro A. Valdes-Sosa, and Aina Puce (Dec. 2020). “Issues and recommendations from the OHBM COBIDAS MEEG committee for reproducible EEG and MEG research”. In: *Nature Neuroscience* 23.12, pp. 1473–1483. ISSN: 1546-1726. DOI: 10.1038/s41593-020-00709-0. URL: https://doi.org/10.1038/s41593-020-00709-0.

Rahman, Md. Mustafizur, Ajay Krishno Sarkar, Md. Amzad Hossain, and Mohammad Ali Moni (2022). “EEG-based emotion analysis using non-linear features and ensemble learning approaches”. In: *Expert Systems with Applications* 207, p. 118025. ISSN: 0957-4174. DOI: https://doi.org/10.1016/j.eswa.2022.118025. URL: https://www.sciencedirect.com/science/article/pii/S0957417422012404.

GDPR (2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing* *of personal data and on the free movement of such data (General Data Protection Regulation).* OJ L 119, 4.5.2016, p. 1–88. URL: https://eur-lex.europa.eu/eli/reg/2016/679/oj (visited on 08/27/2025).

EU AI Act (2024). *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (AI Act).* OJ L 202, 12.7.2024. Postnote e.g., Art. 5 (prohibited AI practices). URL: https://eur-lex.europa.eu/eli/reg/2024/1689/oj (visited on 08/27/2025).

Saha, Simanto, Khondaker A. Mamun, Khawza Ahmed, Raqibul Mostafa, Ganesh R. Naik, Sam Darvishi, Ahsan H. Khandoker, and Mathias Baumert (2021). "Progress in Brain Computer Interface: Challenges and Opportunities". In: *Frontiers in Systems Neuroscience* 15. ISSN: 1662-5137. DOI: 10.3389/fnsys.2021.578875. URL: https://www.frontiersin.org/journals/systems-neuroscience/articles/10.3389/fnsys.2021.578875.

SEED-IV Dataset (2025). BCMI, Shanghai Jiao Tong University. URL: https://bcmi.sjtu.edu.cn/~seed/seed-iv.html (visited on 09/27/2025).

Seo, Yeong-Seok and Jun-Ho Huh (2019). *Automatic Emotion-Based Music Classification for Supporting Intelligent IoT Applications.* accessed 20 Jan 2025. URL: https://www.researchgate.net/figure/Russells-circumplex-model-The-circumplex-model-is-developed-by-James-Russell-In-the_fig1_330817411.

Shih, Jerry J., Dean J. Krusienski, and Jonathan R. Wolpaw (2012). "Brain-Computer Interfaces in Medicine". In: *Mayo Clinic Proceedings* 87.3, pp. 268–279. ISSN: 0025-6196. DOI: https://doi.org/10.1016/j.mayocp.2011.12.008. URL: https://www.sciencedirect.com/science/article/pii/S0025619612001231.

Shoumy, Nusrat J., Li-Minn Ang, Kah Phooi Seng, D.M.Motiur Rahaman, and Tanveer Zia (2020). "Multimodal big data affective analytics: A comprehensive survey using text, audio, visual and physiological signals". In: *Journal of Network and Computer Applications* 149, p. 102447. ISSN: 1084-8045. DOI: https://doi.org/10.1016/j.jnca.2019.102447. URL: https://www.sciencedirect.com/science/article/pii/S1084804519303078.

Shu, Lin, Jinyan Xie, Mingyue Yang, Ziyi Li, Zhenqi Li, Dan Liao, Xiangmin Xu, and Xinyi Yang (2018a). "A Review of Emotion Recognition Using Physiological Signals". In: *Sensors* 18.7. ISSN: 1424-8220. DOI: 10.3390/s18072074. URL: https://www.mdpi.com/1424-8220/18/7/2074.

– (2018b). "A Review of Emotion Recognition Using Physiological Signals". In: *Sensors* 18.7. ISSN: 1424-8220. DOI: 10.3390/s18072074. URL: https://www.mdpi.com/1424-8220/18/7/2074.

Steunebrink, Bas R, Mehdi Dastani, and John-Jules Ch Meyer (2009). "The OCC model revisited". In: *Proc. of the 4th Workshop on Emotion and Computing.* Association for the Advancement of Artificial Intelligence Palo Alto, p. 62.

Swapna (Aug. 2020). *Convolutional Neural Network.* accessed 16 Feb 2025. URL: https://developersbreach.com/convolution-neural-network-deep-learning/.

Tao, Wei, Chang Li, Rencheng Song, Juan Cheng, Yu Liu, Feng Wan, and Xun Chen (2023). "EEG-Based Emotion Recognition via Channel-Wise Attention and Self Attention". In: *IEEE Transactions on Affective Computing* 14.1, pp. 382–393. DOI: 10.1109/TAFFC.2020.3025777.

Torres, Edgar P., Edgar A. Torres, Myriam Hernández-Álvarez, and Sang Guun Yoo (2020). "EEG-Based BCI Emotion Recognition: A Survey". In: *Sensors* 20.18. ISSN: 1424-8220. DOI: 10.3390/s20185083. URL: https://www.mdpi.com/1424-8220/20/18/5083.

Union, European (2024). *The general data protection regulation.* Last Accessed: 2025-02-23. URL: https://www.consilium.europa.eu/en/policies/data-protection-regulation/.
– (2025a). *AI Act.* Last Accessed: 2025-03-02. URL: https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai.
– (2025b). *Data protection in the EU.* Last Accessed: 2025-02-23. URL: https://commission.europa.eu/law/law-topic/data-protection_en.
– (2025c). *EU AI Act: first regulation on artificial intelligence.* Last Accessed: 2025-03-02. URL: https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence.
– (2025d). *The EU Artificial Intelligence Act.* Last Accessed: 2025-02-23. URL: https://artificialintelligenceact.eu/.
Wan, Zitong, Rui Yang, Mengjie Huang, Nianyin Zeng, and Xiaohui Liu (2021). "A review on transfer learning in EEG signal analysis". In: *Neurocomputing* 421, pp. 1–14. ISSN: 0925-2312. DOI: https://doi.org/10.1016/j.neucom.2020.09.017. URL: https://www.sciencedirect.com/science/article/pii/S0925231220314223.
Wang, James Z., Sicheng Zhao, Chenyan Wu, Reginald B. Adams, Michelle G. Newman, Tal Shafir, and Rachelle Tsachor (Oct. 2023). "Unlocking the Emotional World of Visual Media: An Overview of the Science, Research, and Impact of Understanding Emotion". In: *Proceedings of the IEEE* 111.10, pp. 1236–1286. ISSN: 1558-2256. DOI: 10.1109/jproc.2023.3273517. URL: http://dx.doi.org/10.1109/JPROC.2023.3273517.
Wang, Yan, Wei Song, Wei Tao, Antonio Liotta, Dawei Yang, Xinlei Li, Shuyong Gao, Yixuan Sun, Weifeng Ge, Wei Zhang, and Wenqiang Zhang (2022). "A systematic review on affective computing: emotion models, databases, and recent advances". In: *Information Fusion* 83-84, pp. 19–52. ISSN: 1566-2535. DOI: https://doi.org/10.1016/j.inffus.2022.03.009. URL: https://www.sciencedirect.com/science/article/pii/S1566253522000367.
Wang, Zhaoxia, Seng-Beng Ho, and Erik Cambria (2020). "A review of emotion sensing: categorization models and algorithms." In: *Multimed Tools Appl.* DOI: https://doi.org/10.1007/s11042-019-08328-z.
Wang, Zhongmin, Yue Tong, and Xia Heng (2019). "Phase-Locking Value Based Graph Convolutional Neural Networks for Emotion Recognition". In: *IEEE Access* 7, pp. 93711–93722. DOI: 10.1109/ACCESS.2019.2927768.
World Medical Association (2013). "World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects". In: *JAMA* 310.20. Postnote e.g., §§ 23–27, pp. 2191–2194. DOI: 10.1001/jama.2013.281053.
Yadegaridehkordi, Elaheh, Nurul Fazmidar Binti Mohd Noor, Mohamad Nizam Bin Ayub, Hannyzzura Binti Affal, and Nornazlita Binti Hussin (2019). "Affective computing in education: A systematic review and future research". In: *Computers & Education* 142, p. 103649. ISSN: 0360-1315. DOI: https://doi.org/10.1016/j.compedu.2019.103649. URL: https://www.sciencedirect.com/science/article/pii/S0360131519302027.
Yang, Yilong, Qingfeng Wu, Ming Qiu, Yingdong Wang, and Xiaowei Chen (2018). "Emotion Recognition from Multi-Channel EEG through Parallel Convolutional Recurrent Neural Network". In: *2018 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–7. DOI: 10.1109/IJCNN.2018.8489331.
Yu, Chaofei and Mei Wang (2022). "Survey of emotion recognition methods using EEG information". In: *Cognitive Robotics* 2, pp. 132–146. ISSN: 2667-2413. DOI: https://doi.org/10.1016/j.cogr.2022.06.001. URL: https://www.sciencedirect.com/science/article/pii/S2667241322000118.

Zheng, W., W. Liu, Y. Lu, B. Lu, and A. Cichocki (2018). "EmotionMeter: A Multimodal Framework for Recognizing Human Emotions". In: *IEEE Transactions on Cybernetics*, pp. 1–13. ISSN: 2168-2267. DOI: `10.1109/TCYB.2018.2797176`.

Zheng, Wei-Long and Bao-Liang Lu (2015). "Investigating Critical Frequency Bands and Channels for EEG-based Emotion Recognition with Deep Neural Networks". In: *IEEE Transactions on Autonomous Mental Development* 7.3, pp. 162–175. DOI: `10.1109/TAMD.2015.2431497`.

Appendix A

Intensity of labels

Video	Label	subject_1	subject_2	subject_3	subject_4	subject_5	subject_6	subject_7	subject_8	subject_9	subject_1	subject_1	subject_1	subject_1	subject_1	subject_1	subject_1	subject_1	subject_1	subject_1	subject_2
1	Happy	0,3	0,2	0,5	0,5	0,4	0,5	0,6	0,2	0,3	0,4	0,2	0,6	0,7	0,2	0,5	0,4	0,1	0,5	0,3	0,6
2	Neutral	0,1	0,7	0,6	0,7	0,7	0,3	0,2	0,6	0,6	0,8	0,9	0,5	0,8	0,5	0,9	0,9	0,8	0,7	0,8	0,9
3	Disgust	0,6	0,7	0,4	0,5	0,9	0,8	0,8	0,8	0,7	0,5	0,8	0,6	0,9	0,7	0,8	0,5	0,7	1	1	0,8
4	Sad	0,1	0,6	0,2	0,5	0,5	0,6	0,7	0,4	0,7	0,8	0,5	0,8	0,6	0,7	0,5	0,7	0,5	0,3	0,5	0,4
5	Anger	0,6	0,8	0,4	0,8	0,8	0,8	0,6	0,7	1	0,7	0,9	0,7	1	0,5	0,6	0,8	0,7	0,9	1	0,9
6	Anger	0,5	0,9	0,7	0,6	0,6	0,6	0,7	0,8	0,8	0,8	0,8	0,6	0,9	0,8	0,6	0,8	0,5	0,7	1	0,5
7	Sad	0,2	0,8	0,4	0,5	0,9	0,6	0,9	0,6	0,7	0,4	0,7	0,9	0,7	0,6	0,4	0,4	0,8	0,5	0,5	1
8	Disgust	0,9	0,9	1	0,6	1	1	1	0,9	0,8	0,8	1	0,8	1	0,7	1	1	1	1	1	1
9	Neutral	0	0,9	0,7	0,7	0,7	0,6	0,1	0,7	0,6	0,8	1	0,6	0,8	0,5	0,8	0,7	0,8	0,5	0,7	0,8
10	Happy	0	0,5	0,4	0,6	0,5	0,6	0,6	0,4	0,6	0,7	1	0,5	0,6	0,6	1	0,6	0,6	0,6	0,6	0,8
11	Happy	0,1	0,9	0,3	0,4	0,7	0,7	0,8	0,7	0,5	0,6	0,8	0,7	0,8	0,6	1	0,9	0,6	0,8	0,5	0,9
12	Neutral	0	1	0,8	0,3	0,5	0,8	0	0,6	0,6	0,4	1	0,6	0,7	0,4	0,8	0,8	0,9	0,4	0,8	0,5
13	Disgust	0,7	1	0,3	0,6	0,9	0,7	0,6	0,5	0,5	0,2	0,9	0,9	0,7	0,6	0,7	0,7	1	1	0,9	1
14	Sad	0,3	1	0,8	0,6	0,5	0,7	0,5	0,5	0,3	0,7	0,8	0,9	0,8	0,6	0,7	0,7	1	0,3	1	1
15	Anger	0,5	0,9	0,4	0,6	0,4	0,6	0,5	0,7	0,2	0,7	0,7	0,4	0,7	0,7	0,6	0,7	0,7	0,4	0,5	0,5
16	Anger	0,9	1	0,9	0,8	0,9	0,8	0,8	0,6	0,3	0,5	1	0,7	1	0,8	1	1	1	0,9	0,9	0,6
17	Sad	0,9	1	0,7	0,7	1	0,8	0,8	0,5	0,7	0,2	1	1	0,7	0,6	0,7	1	1	0,6	1	0,6
18	Disgust	1	1	1	0,8	1	0,8	0,5	0,6	1	0,6	0,9	0,5	0,9	0,8	1	1	1	1	1	1
19	Neutral	0,1	1	0,5	0,6	0,8	0,5	0,2	0,6	0,5	0,8	1	0,6	0,8	0,4	0,8	0,6	0,9	0,3	0,9	0,8
20	Happy	0,6	1	0,6	0,6	0,4	0,6	0,5	0,4	0,5	0,6	0,8	0,7	0,8	0,5	0,9	0,5	0,7	0,2	0,7	0,5
21	Anger	0,5	1	0,5	0,6	0,8	0,6	0,6	0,5	0,2	0,2	1	0,7	0,7	0,5	0,6	0,8	0,7	0,4	0,6	0,9
22	Sad	0,1	0,8	0,3	0,4	0,6	0,6	0,4	0,3	0,2	0,3	0,9	0,1	0,7	0,7	0,5	0,5	0,7	0,6	0,4	0,2
23	Fear	0,5	0,6	0,6	0,6	1	0,7	0,6	0,6	0,7	0,3	0,3	0,4	0,7	0,8	0,6	1	0,6	1	0,8	0,4
24	Neutral	0	1	0,8	0,3	0,6	0,6	0	0,7	0,7	0,7	1	0,6	0,8	0,5	0,8	0,5	1	0,3	0,8	1
25	Surprise	0,3	0,9	0,5	0,7	0,6	0,7	0,7	0,7	0,7	0,5	0,8	0,6	0,8	0,6	0,9	0,7	0,7	0,5	0,6	0,8
26	Surprise	0,7	1	0,6	0,8	1	0,7	0,7	0,7	0,6	0,8	1	0,6	0,8	0,7	1	0,7	1	0,6	0,8	0,9
27	Neutral	0	1	0,8	0,4	0,6	0,4	0,1	0,7	0,7	0,4	1	0,5	0,9	0,7	0,8	0,5	0,9	0,2	0,9	0,8
28	Fear	0,7	0,8	1	0,8	1	0,9	0,9	0,8	0,9	0,9	0,5	0,6	0,9	0,8	0,8	1	0,5	1	1	1
29	Sad	0,4	1	0,3	0,7	0,4	0,7	0,6	0,6	0,5	0,7	0,8	0,7	0,8	0,6	0,6	0,7	1	0,5	0,6	0,4
30	Anger	0,7	1	0,6	0,6	0,8	0,6	0,5	0,8	0,3	0,7	0,8	0,4	0,8	0,6	0,8	0,7	0,6	0,6	0,6	0,4
31	Anger	0,8	1	0,8	0,9	0,7	0,7	0,7	0,6	0,6	1	1	0,7	0,8	0,8	1	0,9	1	0,7	1	1
32	Sad	0,3	0,7	0,4	0,6	0,5	0,6	0,5	0,7	0,5	0,5	0,7	0,9	0,9	0,7	0,4	0,8	0,7	0,7	0,9	0,8
33	Fear	0,6	1	0,7	0,6	1	0,8	0,8	0,7	1	0,6	0,6	0,9	0,9	0,7	0,6	1	0,6	1	1	1
34	Neutral	0,2	1	0,8	0,6	0,8	0,6	0	0,7	0,7	0,5	1	0,6	0,9	0,6	0,8	0,7	0,5	0,3	1	0,7
35	Surprise	0,6	1	0,7	0,7	0,2	0,8	0,9	0,6	0,7	0,9	1	0,5	0,8	0,6	0,9	0,6	1	0,7	0,9	1
36	Surprise	0,6	1	0,5	0,7	0,3	0,6	0,9	0,7	0,8	0,8	1	0,6	0,8	0,6	1	0,7	1	0,6	1	0,9
37	Neutral	0,2	1	0,7	0,6	1	0,7	0,1	0,7	0,7	0,7	1	0,5	0,8	0,7	0,8	0,6	1	0,2	0,8	0,5
38	Fear	0,6	0,7	0,7	0,6	0,9	0,7	0,6	0,7	1	0,3	0,5	0,6	0,7	0,5	0,6	0,7	0,6	1	0,8	0,7
39	Sad	0,9	1	0,6	0,5	0,6	0,7	0,8	0,6	0,8	0,8	1	0,8	0,8	0,5	0,6	0,8	0,7	1	0,5	0,4
40	Anger	0,4	1	0,5	0,6	0,3	0,8	0,5	0,7	0,3	0,1	1	0,3	0,8	0,6	0,5	0,8	1	0,7	0,8	0,2
41	Happy	0,2	0,7	0,6	0,6	0,5	0,8	0,6	0,6	0,3	0,6	0,5	0,5	0,7	0,5	0,6	0,3	0,3	0,5	0,8	0,4
42	Surprise	0,7	0,8	0,3	0,6	0,7	0,7	0,5	0,6	0,6	0,4	0,6	0,5	0,7	0,7	0,8	0,4	0,8	0,5	0,9	0,6
43	Disgust	1	1	0,6	0,7	1	1	0,7	0,8	0,9	0,6	0,7	0,9	0,8	0,8	1	0,6	1	1	1	0,7
44	Fear	0,6	0,7	0,4	0,6	1	0,7	0,7	0,7	0,5	0,2	0,4	0,7	0,8	0,8	0,6	0,5	0,9	1	0,9	0,4
45	Anger	1	1	0,8	0,8	0,7	0,7	0,7	0,7	0,9	1	1	0,9	0,9	0,8	0,8	1	1	0,8	1	1
46	Anger	0,8	1	0,7	0,7	0,7	0,6	0,5	0,6	0,6	0,9	1	0,6	0,9	0,7	1	0,9	1	0,7	0,8	0,7
47	Fear	0,6	0,6	0,6	0,6	0,9	0,7	0,6	0,5	0,5	0,9	0,6	0,8	0,7	0,8	0,8	0,9	0,8	1	1	0,7
48	Disgust	0,9	1	0,5	0,7	0,7	0,8	0,6	0,6	1	0,5	0,7	0,5	0,7	0,8	0,9	0,6	1	0,7	1	1
49	Surprise	0,6	1	0,5	0,7	0,9	0,6	0,5	0,6	0,6	0,6	0,7	0,6	0,8	0,7	0,6	0,3	0,8	0,5	0,8	0,4
50	Happy	0,6	0,4	0,6	0,7	0,2	0,5	0,7	0,7	0,5	0,7	0,8	0,7	0,7	0,6	0,5	0,5	0,5	0,4	0,7	0,7
51	Happy	0,2	0,4	0,5	0,5	0,4	0,5	0,6	0,7	0,3	0,6	0,7	0,4	0,7	0,5	0,6	0,8	0,8	0,3	0,7	0,6
52	Surprise	0,8	1	0,6	0,6	1	0,8	0,8	0,7	0,7	0,8	1	0,6	0,7	0,8	1	1	1	0,8	1	0,9
53	Disgust	0,6	0,5	0,6	0,6	0,5	0,6	0,4	0,5	0,4	0,3	0,3	0,5	0,7	0,7	0,6	0,5	0,8	0,7	0,8	0,5
54	Fear	0,8	0,8	0,9	0,4	1	0,9	0,9	0,7	0,9	0,7	0,4	0,8	0,9	0,7	0,9	1	0,7	1	1	0,8
55	Anger	0,7	0,2	0,6	0,5	0,2	0,5	0,7	0,4	0,5	0,7	0,8	0,6	0,7	0,7	0,5	0,8	0,8	0,6	1	0,8
56	Anger	0,8	0,8	0,6	0,6	0,9	0,7	0,6	0,6	0,4	0,8	1	0,6	0,9	0,5	0,9	0,8	1	0,7	0,8	0,8
57	Fear	0,8	0,7	0,7	0,6	1	0,8	0,7	0,7	0,8	0,7	0,4	0,7	0,7	0,7	0,6	0,7	0,6	1	0,8	0,6
58	Disgust	0,3	0,5	0,4	0,6	0,3	0,7	0,5	0,4	0,7	0,6	0,7	0,1	0,7	0,7	0,6	0,5	0,8	0,5	0,8	0,3
59	Surprise	0,8	1	0,7	0,8	0,4	0,7	0,7	0,5	0,9	0,9	1	0,7	0,7	0,8	1	1	1	0,6	0,9	1
60	Happy	0,7	1	0,7	0,6	0,5	0,7	0,6	0,7	0,5	0,6	0,8	0,7	0,7	0,6	0,9	0,7	0,7	0,3	0,9	0,3
61	Disgust	1	1	0,8	0,7	1	0,9	0,6	0,6	1	1	0,7	0,9	0,9	0,7	1	0,8	1	1	0,9	0,5
62	Sad	0,6	0,5	0,5	0,7	0,5	0,6	0,7	0,6	0,5	0,5	0,4	0,5	0,8	0,7	0,5	0,6	0,7	0,3	0,8	0,5
63	Fear	0,6	0,5	0,9	0,6	1	0,7	0,7	0,7	0,8	0,6	0,4	0,6	0,7	0,4	0,5	0,9	0,6	1	1	0,5
64	Surprise	0,6	0,8	0,5	0,7	0,5	0,7	0,6	0,6	0,5	0,6	0,7	0,6	0,7	0,6	0,9	0,5	1	0,5	0,9	0,8
65	Happy	0,7	0,3	0,6	0,5	0,7	0,6	0,7	0,6	0,5	0,6	0,7	0,6	0,7	0,6	0,8	0,5	0,7	0,5	0,8	0,6
66	Happy	0,6	0,6	0,7	0,7	0,5	0,7	0,8	0,7	0,5	0,7	0,8	0,9	0,8	0,7	0,7	0,8	0,7	0,5	0,8	0,7
67	Surprise	0,6	0,8	0,7	0,7	0,8	0,7	0,7	0,7	0,7	0,8	1	0,5	0,8	0,5	0,8	0,5	0,9	0,6	0,9	0,8
68	Fear	0,4	0,7	0,4	0,5	0,7	0,7	0,5	0,5	0,4	0,2	0,2	0,5	0,6	0,4	0,6	0,8	0,9	0,6	0,9	0,5
69	Sad	0,6	0,7	0,5	0,6	0,3	0,5	0,6	0,5	0,2	0,6	0,7	0,1	0,7	0,5	0,5	0,4	0,7	0,5	0,7	0,2
70	Disgust	0,7	0,6	0,4	0,6	0,4	0,6	0,3	0,5	0,6	0,4	0,5	0,4	0,7	0,7	0,7	0,3	0,6	0,5	0,7	0,3
71	Disgust	0,7	0,8	0,6	0,6	0,7	0,8	0,7	0,6	0,6	0,7	0,5	0,8	0,9	0,8	1	0,5	0,9	0,5	0,8	0,5
72	Sad	0,6	0,7	0,5	0,6	0,8	0,8	0,8	0,4	0,4	0,5	0,8	0,5	0,7	0,5	0,6	0,5	0,8	0,3	0,8	0,6
73	Fear	0,7	0,8	0,7	0,7	1	0,7	0,7	0,6	0,8	0,7	0,5	0,7	0,8	0,6	0,7	0,6	0,8	1	0,9	0,8
74	Surprise	0,9	0,9	0,8	0,7	0,9	0,7	0,7	0,6	0,7	0,9	1	0,6	0,9	0,7	0,8	1	1	0,6	0,6	0,8
75	Happy	0,7	0,8	0,6	0,6	0,6	0,8	0,6	0,6	0,4	0,7	1	0,5	0,7	0,5	0,8	0,6	1	0,7	0,6	0,8
76	Happy	0,3	0,8	0,6	0,7	0,4	0,7	0,6	0,7	0,4	0,7	1	0,8	0,7	0,6	0,8	0,9	0,7	0,4	0,6	0,8
77	Surprise	0,8	0,8	0,7	0,7	0,8	0,8	0,5	0,5	0,6	0,9	0,8	0,7	0,7	0,7	0,8	0,9	1	0,7	0,8	0,6
78	Fear	0,5	0,7	0,5	0,6	0,4	0,5	0,5	0,6	0,7	0,2	0,3	0,4	0,7	0,4	0,6	0,6	0,8	1	0,7	0,5
79	Sad	0,6	0,7	0,6	0,6	0,5	0,5	0,4	0,6	0,7	0,8	0,6	0,3	0,7	0,6	0,6	0,5	0,8	0,5	0,6	0,5
80	Disgust	0,8	0,7	0,7	0,6	0,5	0,8	0,6	0,5	0,5	0,2	0,9	0,3	0,7	0,8	0,5	0,5	0,6	0,5	0,8	0,3

Figure A.1: Intensity of labels per person and video in SEED VII.

DECLARAÇÃO DE INTEGRIDADE

DECLARAÇÃO DE INTEGRIDADE

Declaro ter conduzido este trabalho académico com integridade. Não plagiei ou apliquei qualquer forma de uso indevido de informações ou falsificação de resultados ao longo do processo que levou à sua elaboração.

Declaro que o trabalho apresentado neste documento é original e de minha autoria, não tendo sido utilizado anteriormente para nenhum outro fim.

Declaro ainda que tenho pleno conhecimento do Código de Conduta Ética do P.PORTO.

João Vieira

ISEP, Porto, (dia) de (mês) de (ano)