ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE INGENIERÍA DE SISTEMAS

BCI-BASED EMOTION RECOGNITION

THESIS SUBMITTED AS PART OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF DOCTOR OF PHILOSOPHY IN INFORMATICS

EDGAR PORFIRIO TORRES PROAÑO
edgar.torres@epn.edu.ec

SUPERVISOR:

DR. SANG GUUN YOO
sang.yoo@epn.edu.ec

Quito, september 2022

ESCUELA POLITÉCNICA NACIONAL

THESIS

For the award of the degree of

DOCTOR OF PHILOSOPHY

IN INFORMATICS

Resolution RPC-SO-43-No.501-2014 of the Consejo de Educación Superior

Presented by

EDGAR PORFIRIO

TORRES PROAÑO

Thesis supervised by Dr. **Sang Guun Yoo** of Escuela Politécnica Nacional (EPN).

BCI-BASED EMOTION RECOGNITION

Oral examination by the following committee:

- Doctor Marco Benalcázar, Escuela Politécnica Nacional (EPN)
- Doctora Sandra Sánchez, Escuela Politécnica Nacional (EPN)
- Doctora Tanya Calle, Escuela Politécnica Nacional (EPN)
- Doctor Vinicio Carrera, Universidad de la Fuerzas Armadas (ESPE)
- Doctor Víctor Sánchez, The University of Warwick, United Kingdom

DECLARATION

I hereby declare under oath that I am the author of this work, which has not previously been presented for obtaining any academic degree or professional qualification. I also declare that I have consulted the bibliographic references included in this document.

Through this declaration, I transfer my intellectual property rights corresponding to this thesis, to the Escuela Politécnica Nacional, as established by the Intellectual Property Law of Ecuador, its Regulations and the current institutional norms.

I declare that this work is based on the following articles of my authorship (as main author or co-author) related to the title of this thesis:

Journals

Scopus Q1, SJR Q2:

E. P. Torres P, E. A. Torres H, M. Hernandez-Alvarez, and S. G. Yoo, “EEG-Based BCI Emotion Recognition : A Survey,” Sensors (Switzerland), no. Emotion and Stress Recognition Related Sensors and Machine Learning Technologies, 2020.

Scopus Q1, SJR Q2:

E. P. Torres, E. A. Torres, M. Hernandez-Alvarez, and S. G. Yoo, “Emotion Recognition Related to Stock Trading Using Machine Learning Algorithms With Feature Selection,” *IEEE Access*, vol. 8, pp. 199719–199732, 2020.

Conference Proceedings:

E. P. Torres P, M. Hernández-Álvarez, E. A. Torres Hernández, and S. G. Yoo, “Stock Market Data Prediction Using Machine Learning Techniques,” in *Advances in Intelligent Systems and Computing*, vol. 918, 2019.

E. P. Torres, E. A. Torres, M. Hernández-Álvarez, and S. G. Yoo, “Machine Learning Analysis of EEG Measurements of Stock Trading Performance,” in *Advances in Artificial Intelligence, Software and Systems Engineering*, Springer Nature, 2021, pp. 53–60.

E. P. Torres, E. Torres H., M. Hernández-Álvarez, and S. G. Yoo, “EEG-Based BCI Emotion Recognition Using the Stock-Emotion Dataset,” in *Advances in Emerging Trends and Technologies*, 2021, pp. 226–235.

E. P. Torres, E. A. Torres, M. Hernández-Álvarez, and S. G. Yoo, “Real-Time Emotion Recognition for EEG Signals Recollected from Online Poker Game Participants,” in *Advances in Artificial Intelligence, Software and Systems Engineering*, 2021, pp. 236–241.

I also declare that I have acknowledged the collaboration of third parties, and the contribution made by other published or unpublished material.

Firmado electrónicamente por:
EDGAR PORFIRIO TORRES PROANO

Edgar Porfirio Torres Proaño

CERTIFICATION

I certify that Edgar Porfirio Torres Proaño has carried out his research under my supervision. To the best of my knowledge, the contributions of this work are novel.

Firmado electrónicamente por:
SANG GUUN YOO .

Sang Guun Yoo

ADVISOR

DEDICATION

To God for blessing and helping me in every moment of my life, and allowing me to complete this doctoral research process.

To my parents Porfirio Salomón Torres Acevedo (+) and María Esther Proaño Calderón, whom I love so much, for their great love, help and dedication as their son, educating me, guiding me, and motivating me permanently, thanks to them I have managed to be what I am.

To my beloved sons Edgar Torres Hernández and Andrés Torres Hernández, who with their love have been able to inspire all my actions.

To my dear brothers Galo Torres Proaño and Diana Torres Proaño, with whom I shared all their love since I was a child and who have been able to motivate me to always surpass me.

To my beloved life partner Grace Rosado Jaime, for her love and permanent support, always motivating me positively in all my activities, and in particular in this research work.

To my dear Elizabeth Trocellier Vergara, who at the time knew how to give me all her love and support in my life, and also in this investigative work.

To my grandparents and grannies, uncles and aunts, cousins and cousins, and in general to my whole family, for always trusting me, and encouraging me at all times to surpass myself in life.

ACKNOWLEDGMENTS

To Dr. Sang Guun Yoo, for the wise and valuable research direction. His knowledge, guidance, help, patience, advice, and understanding were always greatly appreciated.

To Dr. Myriam Hernández Alvarez, for her valuable collaboration and supervision in the research development.

To the distinguished members of my Doctoral Thesis board for their kind help with their valuable recommendations, suggestions and observations, which allowed to improve and successfully complete this doctoral research.

To Engineer Edgar Alejandro Torres Hernández, for the valuable contribution of his knowledge within the field of the stock market, his support in the method to provoke emotions, generation of the poker game dataset; and additionally, the use and testing of various algorithms.

To Licentiate (BS) Andrés Antonio Torres Hernández, for his valuable help and participation in the topics of design, multimedia, and data collection tests.

To Mr. Edison Cabrera, student of Systems Engineering at the Escuela Politécnica Nacional (EPN), for his valuable contribution in obtaining the Stock-Emotion Dataset.

To Engineer Jefferson Sanchez, student of the Computation Master at the Escuela Politécnica Nacional (EPN), for his valuable contribution to the Convolutional Network applications.

To Dr. Walter Fuertes, professor at the Universidad de las Fuerzas Armadas (ESPE), for his important help, support, and guidance in the initial phase of the research.

To Master Gustavo Samaniego, professor at the Escuela Politécnica Nacional (EPN), for his permanent and important motivation to my work in this research.

To my professors, directors, and administrative staff of the Doctorado en Informática of the Escuela Politécnica Nacional (EPN), for their valuable knowledge, advice and contribution during my doctoral career.

PROLOGUE

The present research proposed to contribute to the field of emotion recognition, through the use of brain-computer interfaces, which allow to adequately collect EEG signals related to the recognition process, for subsequent analysis and processing. It is worth noting that in the present work an investigation is carried out on emotion recognition, using two sets of data that are generated from the EEG signals collected by us, by carrying out experiments designed to elicit emotions, using direct methods for provoking them, such as activities in the Stock Market and in the online Poker game.

Different BCI devices were used for data recollection, with 8 and 14 electrodes respectively. All this allowed us to have a new perspective of research in this field, focusing on interactive and dynamic methods for emotions elicitation, as opposed to the predominant methods, which are basically mainly of a passive nature.

Adequate systematic review of the literature is presented, focused on the recognition of emotions using EEG signals, combined with the use of machine learning, which allowed us to realize the best focus for this research; additionally, we are sure that this systematic review of the literature will also be very useful as a reference for other researchers in this field.

As it was indicated before, the two databases developed in this research were very useful for the purposes of this work, and we are also sure that they might be used as references for future research of this type.

For the present work, the process of selecting characteristics was proposed as a mixture of two methods: The Mutual Information Matrix (MIM) and the Chi-squared statistics, which turned out to be very appropriate.

The procedures used in this research, and the results obtained, were, in general, better than those found in the literature review.

The extraction and selection of characteristics, through the use of algorithms that extract the required information from the EEG signals, proved to be better than just using raw data, and allowed to improve the performance of the processes.

TABLE OF CONTENTS

ABSTRACT

This study aims to contribute to the emotion recognition field that uses EEG-based BCIs. The application of this scientific area range is vast, from the medical diagnosis of mental illnesses to the improvement of AI designed to interact with humans, passing through daily use programs either to improve learning environments or to enhance emotional awareness to facilitate decision making in stressful activities that require calm and objectivity to avoid mistakes. Those are some possible uses, among many others.

The present doctoral thesis has researched emotion recognition using two EEG datasets generated by the author. Participants were subjected to two emotion elicitation methods: stock market activities and online poker games. The datasets were developed as part of this study to research results using active emotion elicitation methods. They offered a new perspective where EEG signals are recollected using two different BCI devices of 8 and 14 electrodes, respectively, applied to participants while engaged in active undertakings such as stock trading or online poker games. These activities were chosen because of their essence: they are somewhat unpredictable, ludic, and produce strong emotions in their participants. Besides, these forms of emotion elicitation are more interactive and dynamic than predominant methods using only passive stimulation with images, sounds, or films.

Machine learning systems were also developed. They comprehended the phases of pre-processing, feature extraction, feature selection, classification algorithms, and performance evaluation. These systems recognize emotions using the valence-arousal space. We applied feature extraction techniques such as HOC, statistics, frequency bands, Differential Entropy, Differential Asymmetry, and Rational Asymmetry, i.e., methods in the time, frequency, and spatial domains, to obtain an original feature vector with very informative data.

We propose a feature selection technique combining a filter plus a wrapper algorithm to get an optimized input vector to achieve a more precise classification with less overfitting probabilities and better performance. A mutual information matrix is used as a filter algorithm to evaluate the correlation between pairs of features. Then, after a trial classification, Chi-Square Statistics is used, a wrapper method to eliminate elements that are not significant because they do not influence the output classes. Our proposed method obtains non-redundant features without correlation pair-wise because of the mutual information matrix results; and relevant features that affect the output categorization due to the Chi-Square statistics application.

The machine learning algorithms are chosen using three criteria: (1) because of their better performances compared with other tested, (2) since each represents different types of approaches to cover an ample range of machine learning algorithms: kNN is simple, non-parametric, Random Forest is an ensemble algorithm, MLP is a neuronal network, and 1DCNN is deep learning, and (3) because in the literature review we found that these are the algorithms that performed better for 3 to 4 class categorization. Random forest, multi-layer perceptron, kNN, and one-dimensional CNN are the selected algorithms. They were applied to our datasets and to the DEAP database, which is public access information. The application in a DEAP dataset subset, with 8 and 14 channels, aims to compare the systems' performances that the three datasets present using the same feature extraction and selection methods. These performances, to some extent, could prove the capability of the three datasets' data to give valuable information to be used for emotion recognition.

The obtained results were satisfactory, comparable, or even better than those presented in state-of-the-art.

RESUMEN

Este estudio tiene como meta contribuir al campo de reconocimiento de emociones usando interfaces cerebro-computador para recolección de señales EEG. La aplicación de este campo científico es vasta y va desde el diagnóstico médio de enfermedades mentales a la mejora de IA diseñada para interactuar con humanos, pasando por programas de uso diario para mejorar ambientes de aprendizaje o mejorar la identificación personal de las emociones para facilitar la toma de decisiones en actividades que requieren calma y objetividad para no cometer errores. Los mencionados son algunos posibles usos entre muchos otros.

En la presente tesis se realiza una investigación sobre reconocimiento de emociones usando dos datasets de señales EEG generados por nosotros, con participantes sometidos a dos tipos de metodos de elicitación de emociones: actividades en el stock market y el juego de poker en línea. Los datasets fueron generados como parte del estudio para investigar métodos activos para provocar emociones y ofrecen una perspectiva nueva, en la que se recopila señales EEG usando dos diferentes dispositivos con 8 y 14 electrodos respectivamente, aplicados a participantes que realizan actividades de trading y de juego de poker en línea. Estas actividades fueron escogidas porque son en esencia no-predecibles, lúdicas y porque son capaces de producir emociones fuertes en los participantes. Además, esta forma de elicitación de emociones es interactiva y dinámica en contraposición con los métodos predominantes en los que se usa solamente estimulación pasiva con imágenes, sonidos o películas.

Se usan sistemas de aprendizaje automático consistentes en las siguientes fases: pre-procesamiento, extracción de características, selección de características, algoritmos de aprendizaje automático (entrenamiento y prueba) y evaluación de rendimiento. Con ellos se efectúa el reconocimiento de emociones en el plano valencia – arousal. Se usan procesos para extracción de características relevantes de las señales en el dominio del tiempo, tales como HOC; en el dominio de la frecuencia como bandas de frecuencia, Differential Entropy, y con información espacial por las diferencias entre hemisferios: Differential Asymetry and Rational Asymetry.

Se propone un método de selección de características que es una combinación de técnicas tipo filtro y wrapper para obtener un vector de entrada optimizado para una clasificación más precisa con menos posibilidades de overfitting y mejor rendimiento. Usamos una mutual information matrix como algoritmo de filtrado que evalúa la correlación entre pares de features. Luego de una clasificación de prueba se usa Chi-Square statistics,

un método wrapper para descartar características que no son significativas porque no influencian en las clases a la salida de los clasificadores. Es decir que el método propuesto de selección de características produce features no redundantes por el análisis de la correlación entre pares de atributos; y, características relevantes porque afectan la categorizacion de emociones a la salidad debido a la aplicación de la estadística Chi-Square.

Los algoritmos de machine learning se escogieron utilizando tres criterios: (1) que tengan mejor rendimiento comparado con otros algoritmos que se probaron, (2) que representen a un amplio rango de algoritmos de machine learning: kNN es simple, no-paramétrico, Random Forest es un algoritmo tipo ensemble, MLP es una red neuronal y 1DCNN es deep learning, y (3) porque en la revisión de literatura encontramos que estos son los algoritmos que funcionan mejor para clasificación en 3 o 4 clases. Entonces los algoritmos seleccionados son kNN, Random Forest, multi-layer perceptrón y CNN para una dimensión y se los aplica a las dos bases de datos de propia generación. Además, con propósito de comparación también se los aplica en un subset de una base datos pública: DEAP. La aplicación en estos subsets de DEAP con 8 y 14 canales, tienen como objetivo comparar los rendimientos de los sistemas que los tres datasets prsentan usando los mismos métodos de extracción y selección de características de entrada a los clasificadores. Estos rendimientos, en alguna medida, podrían ser prueba de la capacidad de los datos de los tres datasets para dar información útil para ser usada para reconocimiento de emociones.

Los resultados obtenidos con los algoritmos de clasificación usados fueron satisfactorios, comparables o mejores que los del estado del arte.

1. INTRODUCTION

Affective computing is a branch of artificial intelligence that relates to, arises from, or influences emotions [1]. Automatic emotion recognition is an area of study that forms part of affective computing. Research in this area is rapidly evolving thanks to the availability of affordable devices for capturing brain signals, which serve as inputs for systems that decode the relationship between emotions and electroencephalographic (EEG) variations. These devices are called EEG-based brain-computer interfaces (BCIs).

Affective states play an essential role in decision-making. Such states can facilitate or hinder problem-solving. Emotion recognition takes advantage of positive affective states, enhances emotional intelligence, and improves professional and personal success [2]. Moreover, emotional self-awareness can help people manage their mental health and optimize their work performance. Automatic systems can increase our understanding of emotions and promote effective communication among individuals and human-to-machine information exchanges. Automatic EEG-based emotion recognition could also help enrich people's relationships with their environment. Besides, automatic emotion recognition will play an essential role in artificial intelligence designed for human interaction [3].

According to Gartner's 2019 Hype Cycle report on trending research topics, affective computing is at the innovation trigger stage, evidenced by the field's copious publications. However, there are still no defined standards for the different systems components that recognize emotions using EEG signals, and it is still challenging to detect and classify emotions reliably. Thus, a survey that updates the information in the emotion recognition field, focusing on new computational developments, is worthwhile. At the beginning of this research, a literature review was carried out. The results presented in chapter 3 of this document allowed us to detect the gap in the development of the present work because of the lack of public datasets with easy access for emotion recognition using EEG signals. Moreover, the few public datasets available were generated using traditional passive emotion elicitation methods, becoming clear that new different active ways should be investigated. Another question that arises is if the number of electrodes has to be 32 or even 64 to give sufficient EEG information to the emotion recognition systems or if a fewer number of channels would be enough for emotion identification.

Additionally, the results highlight the need to use a feature vector with relevant information for processing them with lower computational costs and better accuracies because results obtained from raw data could be lower than those using feature vectors, and the processing time is lengthy. Also, different results from different techniques to extract and select features exist, but this field needs further research to find the best possible combination for EEG signals.

Based on the literature's detected gaps, our hypothesis is that it is possible to obtain emotion recognition systems using machine learning algorithms with evaluation performances comparable to or even higher than state-of-the-art standards if we:

- Obtain datasets with data corresponding to commercial BCI devices with eight and 14 channels using new active emotion elicitation ways.
- Apply feature extraction with algorithms in time, frequency, and spatial domains, to produce an original feature vector that would be enough to obtain satisfactory classification results and shorter execution times versus those obtained with raw data.
- Present a new combination of feature selection methods to get as good emotion recognition results as those obtained with the original features.

At the end of the present study, we expect to confirm this hypothesis and contribute with two datasets that use active emotion elicitation methods with commercial BCI with fewer channels but with enough information about participants' emotions so they can be detected and classified using a machine learning system. Also, we propose a schema for feature extraction and selection that allows better than state-of-the-art standards for classifiers' performance for four class categorization and better processing times than if raw data is used. The time consideration is vital for systems that can be used in real-time.

The content of the present work is organized as follows: Chapter 2 presents an introduction to the topic. Chapter 3 shows the structure of EEG-based BCI systems for emotion recognition, and their principal processes are revised, i.e. (1) Signal acquisition, (2) preprocessing, (3) feature extraction, (4) feature selection, (5) classification, and (6) performance evaluation. Then, Chapter 4 presents a literature review that analyses the emotion recognition advances using EEG signals and BCI to (1) identify trends in algorithms and technology in this area, (2) detect potential errors that must be overcome for better results, and (3) identify possible knowledge gaps in the field. The aim is to distinguish what has already been done in systems implementations and catch a glimpse of what could lie ahead. Later, Chapter 5 treats the datasets acquisition for our work, and Chapter 6 presents the performed experimentation and its results using different combinations of features and data. Finally, Chapter 7 presents the conclusions and future work.

2. BACKGROUND

The present section introduces EEG-based BCI emotion recognition, representations, and band frequency association with emotions.

2.1. Emotion Recognition using EEG-Based BCI

Many studies suggest that emotional states are associated with electrical activity produced in the central nervous system [1]. Brain activity can be detected through its electrical signals by sensing its variations, locations, and functional interactions [4], and it can be done using EEG devices. EEG signals have an excellent temporal resolution and are a direct measurement of neuronal activity. These signals cannot be manipulated or simulated to fake an emotional state, providing thus reliable information. The challenge is to decode this information and map it to specific emotions.

One affordable and convenient way to detect EEG signals is through EEG-based BCI devices that are non-invasive, low cost, and even wearable, such as helmets and headbands. The development of these tools has facilitated the emergence of much research in the emotion recognition field [5].

As some scientists predict, the usability of EEG-based BCI devices will soon improve. Therefore, shortly, they could be used on an everyday basis for emotion detection for several purposes, such as emotion monitoring in health care facilities, gaming and entertainment, teaching-learning scenarios, and for optimizing workplace performance, among other applications [6].

2.2. Emotion Representations

Emotions can be represented using different general models. The most used are the discrete and the dimensional models [5]. The discrete model identifies basic, innate, and universal emotions from which all other emotions can be derived. Some authors state that these primary emotions are happiness, sadness, anger, surprise, disgust, and fear [7]. Some researchers consider this model has limitations in representing specific emotions in a broader range of affective states.

Alternatively, dimensional models can express complex emotions in a two-dimensional continuous space: Valence, arousal (VA), or in three dimensions: Valence, arousal, and dominance (VAD) [8]. The VA model has valence and arousal as axes. Valence rates positive and negative emotions and ranges from happy to unhappy (or sad). Arousal measures emotions from calm to stimulated (or excited). Three-dimensional models add a dominance axis to evaluate from submissive (powerless) to empowered feelings. This representation distinguishes emotions that are jointly represented in the VA model. For instance, fear and anger have similar valence-arousal presentations on the VA plane [8]. Thus, three-dimensional models improve "emotional resolution" through the dominance dimension. In this example, fear is a submissive feeling, but anger requires power [9]. Hence, the dominance dimension improves the differentiation between these two emotions.

Figure 1 shows a VA plane with the representation of basic emotions. The horizontal axis corresponds to valence dimensions, from positive to negative emotions. Likewise, the vertical axis corresponds to arousal. These two variables can be considered emotional state components [4]. Figure 2 presents the VAD space representing the same basic emotions.

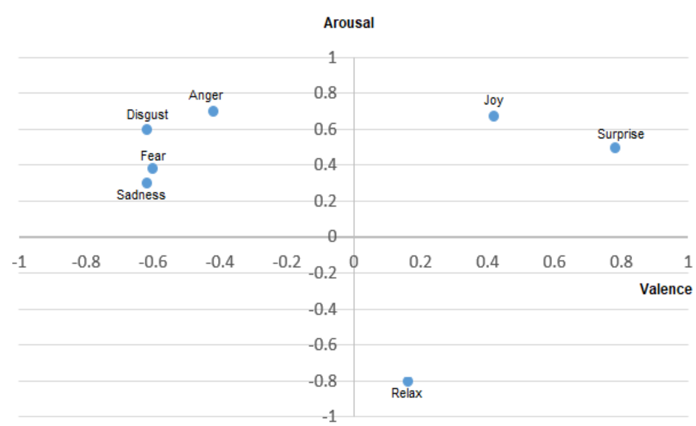


Figure 1. Emotional states in the Valence-Arousal space [10].

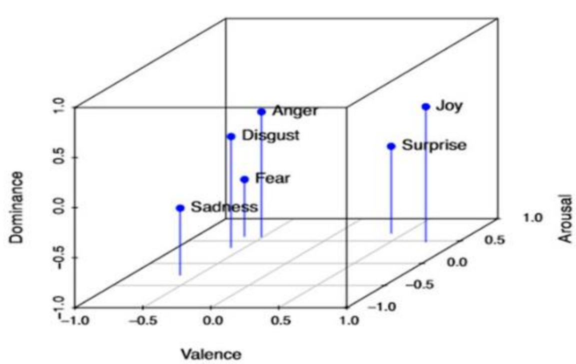


Figure 2. Emotional states in the Valence-Arousal-Dominance space [11].

Table 1 shows that some researchers, studying EEG-based functional connectivity in the brain, have reported a relationship between specific brain areas and emotional states. Studies that take at-single-electrode-level analysis into account have shown that asymmetric activity at the frontal site in the alpha band is associated with emotion. Ekman and Davidson found that enjoyment activated the brain's left frontal parts [12]. Another study found a left frontal activity reduction when volunteers adopted fear expressions [13]. Increased power in theta bands at the frontal midline is associated with pleasurable emotions, and the opposite has been observed with unpleasant feelings [14].

Table 1. Frequency bands associations [15],[16].

Band	State Association	Potential Localization	Stimuli
Gamma rhythm (above 30 HZ)	Positive valence. These waves are correlated with positive spiritual feelings. Arousal increases with high-intensity visual stimuli.	Different sensory and non-sensory cortical networks.	These waves appear caused by attention, multi-sensory information, memory, and consciousness.
Beta (13 to 30 Hz)	They are related to visual self-induced positive and negative emotions. These waves are associated with alertness and problem-solving.	Motor cortex.	They are stimulated by motor activity, motor imagination, or tactile stimulation. Beta power increases during the tension of scalp muscles, which are also involved in frowning and smiling.
Alpha (8 to 13 Hz)	They are linked to relaxed and wakeful states, feelings of conscious awareness, and learning.	Parietal and occipital regions. Asymmetries reported: that rightward-lateralization of frontal alpha power during positive emotions, compared to negative or withdrawal-related emotions, originates from leftward-lateralization of prefrontal structures.	These waves appear during relaxation periods with eyes shut while remaining still awake. They represent the visual cortex in a repose state. These waves slow down when falling asleep and accelerate when opening the eyes, moving, or even when thinking about the intention to move.
Theta (4 to 7 Hz)	They appear in relaxation states, allowing better concentration in those cases. These waves also correlate with anxious feelings.	The front central head region is associated with the hippocampal theta waves.	Theta oscillations are involved in memory encoding and retrieval. Additionally, individuals that experience higher emotional arousal in a reward situation reveal an increase of theta waves in their EEG [16]. Theta coma

			waves appear in patients with brain damage.
Delta (0 to 4 Hz)	They are present in the deep NREM 3 sleep stage. Since adolescence, their presence during sleep declines with advancing age.	Frontal, temporal, and occipital regions.	Deep sleep. These waves also have been found in continuous attention tasks [17].

Several studies confirm that frequency bands are related to affective responses. However, emotions are complex processes. The authors in [14] assert that recognizing different emotional states may be more valid if EEG-based functional connectivity is examined rather than a single analysis at the electrode level. Correlation, coherence, and phase synchronization indices between EEG electrodes to estimate functional connectivity between different brain locations. Likewise, differential entropy (DE) and its derivatives like differential asymmetry (DASM), rational asymmetry (RASM), and differential caudality (DCAU) measure functional dissimilarities. Such features are calculated through logarithmic power spectral density for a fixed-length EEG sequence, plus the differences and ratios between DE features of hemispheric asymmetry electrodes [18].

The growing consensus seems to be that a simple mapping between emotions and specific brain structures is inconsistent with observations of different emotions activating the same structure or one emotion activating several structures [19]. Additionally, functional connectivity between brain regions or signal complexity measures may help detect and describe emotional states [20].

2.3. EEG-Based BCI Systems for Emotion Recognition

Figure 3 presents the structure of an EEG-based BCI system for emotion recognition. Such structure comprises the following processes: signal acquisition, preprocessing, feature extraction, feature selection, classification, and performance evaluation, which are explained in detail in this subchapter.

Figure 3. Processes of an EEG-based BCI for emotion recognition.

2.4. Signal Acquisition

Inexpensive wearable EEG helmets and headsets that use noninvasive electrodes along the scalp can efficiently acquire EEG signals. EEG is a recording of electrical activity in the brain over time [21]. Thus, electrodes capture signals, amplify them, and send them to a computer (or mobile device) for storage and processing. Currently, various low-cost EEG-based BCI devices are available on the market [22]. However, many current models of EEG-based BCI become uncomfortable after continued use. Therefore, it is still necessary to improve their usability.

Public Databases

Alternatively, there are also public databases with EEG data for affective information. Table 2 presents a list of available datasets related to emotion recognition. Such datasets are convenient for research, and several emotion recognition studies have used them.

Table 2. Publicly available datasets.

Source	Dataset	Number of Channels	Emotion Elicitation	Number of Participants	Target Emotions
[19]	DEAP	32 EEG channels	Music videos	32	Valence, arousal, dominance, liking
[23]	eNTERFACE'06	54 EEG channels	Selected images from IAPS.	5	Calm, positive, exciting, negative exciting
[24]	headIT	-	Recall past emotions	31	Positive valence (joy, happiness) or

					negative valence (sadness, anger)
[25]	SEED	62 channels	Film clips	12	Positive, negative, neutral
[26]	SEED-IV	62 channels	72 film clips	15	Happy, sad, neutral, fear
[27]	Mahnob-HCI-tagging	32 channels	Fragments of movies and pictures.	30	Valence and arousal rated with the self-assessment manikin
[28]	EEG Alpha Waves dataset	16 channels	Resting-state eyes open/closed experimental protocol	20	Relaxation
[29]	DREAMER	14 channels	Film clips	23	Rating 1 to 5 to valence, arousal, and dominance
[30]	AMIGOS	14 channels	Videos	40	Valence, arousal, dominance, familiarity, and liking
[31]	RCLS	64 channels	Native Chinese Affective Video System	14	Happy, sad, and neutral

Emotion Elicitation

The International Affective Picture System (IAPS) [32] and the International Affective Digitized Sound System (IADS) [33] are the most popular resources for emotion elicitation. These datasets provide emotional stimuli in a standardized way. Hence, IAPS and IADS are helpful for experimental research.

IAPS consists of 1200 images divided into 20 sets of 60 photos. Valence and arousal values are tagged for each photograph. IADS' latest version provides 167 digitally recorded natural sounds familiar in daily life, with sounds labeled for valence, arousal, and dominance. Participants labeled the dataset using the Self-Assessment Manikin system [11]. IAPS and IADS stimuli are accessible with labeled information, which is convenient for constructing ground truth for emotion assessment [34].

Other researchers used movie clips, which have also been shown to provoke emotions. In [35], the authors state that emotions using visual or auditory stimuli are similar. However, results obtained through affective labeling of multimedia may not be generalizable to more interactive situations or everyday circumstances. Thus, new studies using interactive emotional stimuli to ensure the generalizability of results for BCI are welcomed.

Numerous experiments stimulated emotions in different settings, but they did not use EEG devices. However, they collected other physiological indicators like heart rate, skin galvanic changes, and respiration rate. Conceptually, such paradigms could be helpful if they are replicated for EEG signal acquisition. Possible experiments include stress during interviews to detect anger, anxiety, rejection, and depression. Exposure to odorants triggers emotions, such as anger, disgust, fear, happiness, sadness, and surprise. Harassment provokes fear. A threat of short-circuit or a sudden backward-tilting chair elicits fear. A thread of shock causes anxiety. Naturally, these EEG-based BCIs experiments should take into account ethical considerations.

To our knowledge, only a few studies have used more interactive conditions where participants played games or used flight simulators to induce emotions [36][37]. Alternatively, some authors have successfully used auto-induced emotions through memory recall [38].

Normalization

EEG signals vary widely in amplitude depending on age, sex, and other factors like changes in subjects' alertness during the day. Hence, it is necessary to normalize measured values to deal with this variability.

There are three possible approaches to normalization. The first is to record reference conditions without stimulus on the subject. The values obtained can be normalized by subtracting the reference value, then dividing by the reference value (or subtracting the reference value), and then dividing by that same value. The second approach also requires reference conditions. Those values are included in the feature vector, which will have twice the characteristics that make up the "baseline matrix." The third approach normalizes the data separately by obtaining a specific range, for example, between −1 and 1. This method applied to each feature independently ensures that all characteristics have the same value ranges [39],[40].

The effect of normalization and its influence on the entire process of emotion recognition is not yet evident. However, some studies show that normalization allows the characteristics to be generalized in cross-subject emotion recognition. Tangentially, data normalization helps machine learning algorithms' efficiency due to faster convergence.

2.5. Preprocessing

EEG signals' preprocessing relates to signal cleaning and enhancement. EEG signals are weak and easily contaminated by internal and external noise sources. Thus, these processes are essential to avoid noise contamination affecting posterior classification. The body itself may produce electrical impulses through blinking, eye or muscular movement, or even heartbeats that blend with EEG signals. It should be carefully considered whether these artifacts should be removed because they may have relevant emotional state information and could improve emotion recognition algorithms' performance. If filters are used, it is necessary to be cautious to avoid signal distortions.

The three commonly used filter types in EEG are (1) low-pass filters, (2) high-pass filters, and (3) notch filters. The first two filters filter frequencies between 1 and 50–60 Hz. Filters such as Butterworth, Chebyshev, or inverse Chebyshev for EEG signal processing are preferred [40]. Each of them has specific features that need to be analyzed. A Butterworth filter has a flat response in the passband, the stopband, and a broad transition zone. The Chebyshev filter has a ripple on the passband, and a steeper transition, so it is monotonic on the stopband. The inverse Chevishev has a flat response in the passband. It is narrow in the transition and has a ripple in the stopband. To avoid this problem, a Butterworth phase zero filter should be used to prevent a phase shift because this filter goes forward and backward over the signal.

Another preprocessing objective is to clean the noise that may correspond to low-frequency signals generated by an external source, such as power line interference [21]. Notch filters are used to stop the passage of a specific frequency rather than a frequency range. This filter is designed to eliminate frequencies originated by electrical networks, and it typically ranges from 50 to 60 Hz depending on the electrical signal's frequency in the specific country.

All these filters are appropriate for artifact elimination in EEG signals. However, as previously noted, care must be taken when using filters. Generally, filters could distort the EEG signal's waveform and structure in the time domain. Hence, filtering should be kept to a minimum to avoid the loss of EEG signal information.

Nevertheless, preprocessing helps to separate different signals and sources. Table 3 shows methods used for preprocessing EEG signals [41] and the percentage in which they are mentioned in the literature from 2015 to 2022. Independent Component Analysis (ICA) and Principal Component Analysis (PCA) are tools that apply blind source analysis to isolate the source signal from noise when using multi-channel recordings. ICA and PCA can be used for artifact removal and noise reduction. Common Average Reference (CAR) is suitable for noise reduction. SL is applied for spatial filtering to improve the signal's spatial resolution. The Common Spatial Patterns (CSP) algorithm finds spatial filters to distinguish signals corresponding to muscular movements.

Therefore, each of the discussed preprocessing algorithms has its advantages and limitations. In Table 3, we can observe from the percentage of the usage column that the most utilized algorithms for preprocessing are PCA (50.1%), ICA (26.8%), and CSP (17.7%). These values are obtained by computing the proportion of each algorithm used in the different implementations' processes in the 60 papers we analyzed in the literature review.

Table 3. Frequently used pre-preprocessing methods of EEG signals.

Preprocessing Method	**Main Characteristics**	**Advantages**	**Limitations**	**Literature's Usage Statistics % (2015–2022)**
Independent component analysis (ICA) [42]	ICA separates artifacts from EEG signals into independent components based on the data's characteristics without relying on reference channels. It decomposes the multi-channel EEG data into temporal separate and spatial-fixed components. It has been applied for ocular artifact extraction.	ICA efficiently separates artifacts from noise components. ICA decomposes signals into temporal independent and spatially fixed components.	ICA is successful only under specific conditions where one of the signals is greater than the others. The quality of the corrected signals depends strongly on the quality of the artifacts.	26.8
Common Average Reference (CAR) [43], [44]	CAR is used to generate a reference for each channel. The algorithm obtains an average of all the recordings on every electrode and then uses it as a reference. The result is an improvement in the Signal to Noise Ratio quality.	CAR outperforms standard types of electrical referencing, reducing noise by >30%.	The average calculation may present problems for finite sample density and incomplete head coverage.	5.0
Surface Laplacian (SL) [45]–[48]	SL is a way of viewing the EEG data with high spatial resolution. It estimates current density entering or leaving the scalp through the skull, considering the	SL estimates are reference-free, meaning that any EEG recording	It is sensitive to artifacts and spline patterns.	0.4

Preprocessing Method	Main Characteristics	Advantages	Limitations	Literature's Usage Statistics % (2015–2022)
	volume conductor's outer shape, and does not require details of volume conduction.	reference scheme will render the same SL estimates. SL enhances the spatial resolution of the EEG signal. SL does not require any additional assumptions about functional neuroanatomy.		
Principal Component Analysis (PCA) [35,50–55]	PCA finds patterns in data. It can be pictured as a rotation of the coordinate axes so that they are not along with single time points. Still, linear combinations of sets of time points collectively represent a pattern within the signal. PCA rotates the axes to maximize the variance within the data along the first axis, maintaining their orthogonality.	PCA helps in the reduction of feature dimensions. The ranking will be done, and it will help classify data.	PCA does not eliminate noise, but it can reduce it. PCA compresses data compared to ICA and allows for data separation.	50.1
Common Spatial Patterns (CSP) [55]–[57]	CSP applies spatial filters that discriminate different classes of EEG signals. For instance, those corresponding to different motor activity types. CSP also estimates covariance matrices.	CSP does not require a priori selection of sub-specific bands and knowledge of these bands.	CSP requires many electrodes. Changes in electrode location may affect classification accuracies.	17.7

2.6. Feature Extraction

Once signals are noise-free, the BCI needs to extract essential features, which will be fed to the classifier. Features can be computed in the domain of (1) time, (2) frequency, (3) time-frequency, or (4) space, as shown in Table 4 [31,38,39]. This table presents the most popular techniques used for feature extraction, their domain, advantages, and limitations.

Time-domain features include the event-related potential (ERP), Hjorth features, and higher-order crossing (HOC) [58], [60], independent component analysis (ICA), principal component analysis (PCA), and Higuchi's fractal dimensions (FD) as a measure of signal complexity and self-similarity in this domain. There are also statistical measures, such as power, mean, standard deviation, variance, skewness, kurtosis, relative band energy, and entropy. The latter evaluates signal randomness [61].

The most popular frequency-domain method is the Fast Fourier transform (FFT). Auto-regressive (AR) modeling is an alternative to Fourier-based methods for computing the frequency spectrum of a signal [62],[63].

The time-frequency domain exploits variations in time and frequency, which are very descriptive of the neural activities. For this, wavelet transform (WT) and wavelet packet decomposition (WPD) are used [62].

The spatial information provided in the description of EEG signals' characteristics is also considered broader. For this dimension, signals are referenced to digitally linked ears (DLE) values, which are calculated in terms of the left and right earlobes as follows:

$$V_e^{DLE} = V_e - \frac{1}{2}(V_{A1} + V_{A2}), \tag{1}$$

VA1 and VA2 are the reference voltages on the left and right earlobe. These references allow finding the absolute potential corresponding to a given scalp area, where an electrode is located. Consequently, each channel contains spatial information of the location pertinent to its source.

The surface Laplacian (SL) algorithm dramatically reduces volume conduction effects for spatial computation. SL also improves EEG spatial resolution by reducing the distortion produced by volume conduction and reference electrodes [47].

Figure 4 shows EEG signals in the time domain, the frequency domain, and spatial information.

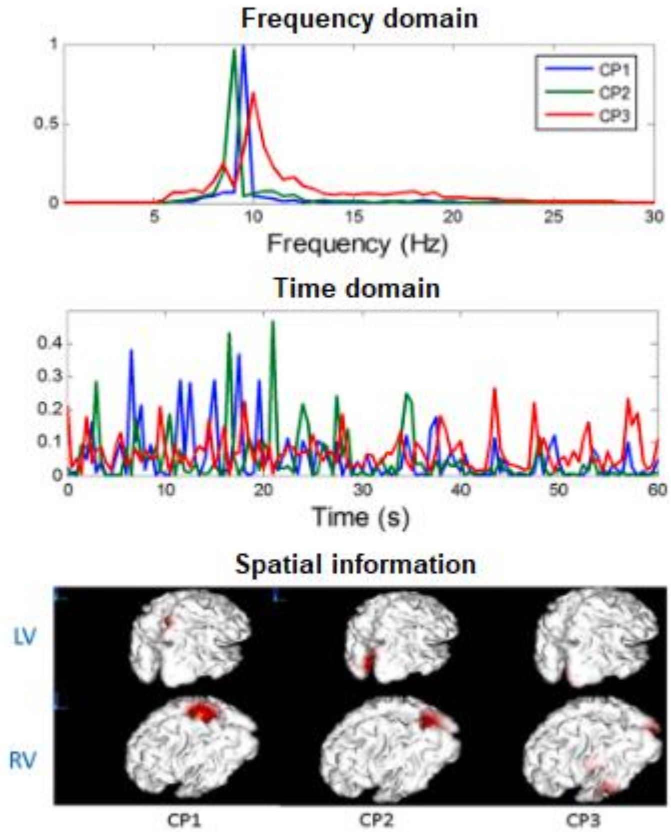


Figure 4. Frequency domain, time domain, and spatial information [63].

Table 4. Feature extraction algorithms.

Feature Extraction Method	Main Characteristics	Domain	Advantages	Limitations	Literature's usage statistics % (2015–2022)
ERP [17], [21], [64]–[69]	The brain responds to a sensory, cognitive, or motor event. Two sub-classifications are (1) evoked potentials and (2) induced potentials.	Time	It has an excellent temporal resolution. ERPs provide a measure of the processing between a stimulus and a response.	ERP has a poor spatial resolution, so it is not useful for research questions related to the activity location.	2.6
Hjorth Features [52,59,60]	These are statistical indicators whose parameters are	Time	Low computational cost appropriate for	Possible statistical bias in signal parameter calculations	17.1

Feature Extraction Method	Main Characteristics	Domain	Advantages	Limitations	Literature's usage statistics % (2015–2022)
	normalized slope descriptors. These indicators are activity (variance of a time function), mobility (mean frequency of the proportion of standard deviation of the power spectrum), and complexity (change in frequency compared to the signal's similarity to a pure sine wave).		real-time analysis.		
Statistical Measures [39,40,42,52,61–70]	Signal statistics: power, mean, standard deviation, variance, kurtosis, relative band energy.	Time	Low computational cost.	-	8.7
DE [1,10,11,15,59,68,71–84]	Entropy shows scattering in data. Differential Entropy can reflect spatial signal variations.	Time–spatial	Entropy and derivate indexes reflect the intra-cortical information flow.		5.1
HOC [1,2,42,63,85–88]	Oscillation in times series can be represented by counts of axis crossing and its differences. HOC displays a monotone property whose rate of increase discriminates	Time	HOC reveals the oscillatory pattern of the EEG signal providing a feature set that conveys emotional information to the classification space.	The training process is time-consuming due to the dependence of the HOC order on different channels and different channel	2.2

Feature Extraction Method	Main Characteristics	Domain	Advantages	Limitations	Literature's usage statistics % (2015–2022)
	between processes.			combinations [60].	
ICA [20,37,53,69,89–91]	ICA is a signal enhancing method and a feature extraction algorithm. ICA separates components that are independent of each other based on the statistical independence principle.	Time. There is also a FastICA in the frequency domain.	ICA efficiently separates artifacts from noise components. ICA decomposes signals into temporal independent and spatially fixed components.	ICA is only useful under specific conditions (one of the signals is greater than the others). The quality of the corrected signals depends strongly on the quality of the isolated artifacts.	11.1
PCA [33,40,52,69,92–95]	The PCA algorithm is mainly used for feature extraction but could also be used for feature extraction. It reduces the dimensionality of the signals creating new uncorrelated variables.	Time	PCA reduces data dimensionality without information loss.	PCA assumes that the data is linear and continuous.	19.7
WT [48]	The WT method represents the original EEG signal with secured and straightforward building blocks known as wavelets, which can be discrete or continuous.	Time-frequency	WT describes the features of the signal within a specified frequency domain and localized time domain properties. It is used to analyze irregular data patterns. It uses variable windows, wide for low	High computational and memory requirements.	26.0

Feature Extraction Method	Main Characteristics	Domain	Advantages	Limitations	Literature's usage statistics % (2015–2022)
			frequencies and narrow for high frequencies.		
AR [48]	AR is used for feature extraction in the frequency domain. AR estimates the EEG's power spectrum density (PSD) using a parametric approach. The estimation of PSD is achieved by calculating the coefficients or parameters of the linear system under consideration.	Frequency domain	AR is used for feature extraction in the frequency domain. AR limits the leakage problem in the spectral domain and improves frequency resolution.	The order of the model in the spectral estimation is challenging to select. It is susceptible to biases and variability.	1.6
WPD [95]	WPD generates a sub-band tree structuring since a full binary tree can characterize the decomposition process. WPD decomposes the original signals orthogonally and independently from each other and satisfies the law of conservation of energy. The energy distribution is extracted as the feature.	Time-frequency	WPD can analyze non-stationary signals such as EEG.	WPD uses a high computational time to analyze the signals.	1.6
FFT [48]	FFT is an analysis method	Frequency	FFT has a higher speed	FFT has low-frequency	2.2

Feature Extraction Method	Main Characteristics	Domain	Advantages	Limitations	Literature's usage statistics % (2015–2022)
	in the frequency domain. EEG signal characteristics are reviewed and computed using power spectral density (PSD) estimation to selectively represent the EEG sample signal.		than all the available methods for real-time applications. It is a useful tool for stationary signal processing.	resolution and high spectral loss of information, making it hard to find the actual frequency of the signal.	
Functional EEG connectivity indices [14]	EEG-based functional connectivity is estimated in the frequency bands for all pairs of electrodes using correlation, coherence, and phase synchronization index. Repeated variance measures for each frequency band were used to determine different connectivity indices among all pairs.	Frequency	Connectivity indices at each frequency band can be used as features to recognize emotional states.	Difficult to generalize and distinguish individual differences in functional brain activity.	1.3
Rhythm [13],[56]	Detection of repeating patterns in the frequency band or "rhythm."	Frequency	Specific band rhythms contribute to emotion recognition.	-	0.1
Graph Regularized Sparse Linear Regularized GRSLR [31]	This method applies a graph regularization and a sparse regularization on the transform matrix	Frequency	It can simultaneously cope with sparse transform matrix learning while	-	0.2

Feature Extraction Method	Main Characteristics	Domain	Advantages	Limitations	Literature's usage statistics % (2015–2022)
	of linear regression		preserving the intrinsic manifold of the data samples.		
Granger causality [63], [96]	This feature is a statistical concept of causation that is based on prediction.	Frequency	The authors can analyze the brain's underlying structural connectivity.	These features only give information about the linear characteristics of signals.	0.5

According to [97], emotions emerge as the synchronization of various subsystems. Several authors use synchronized activity indexes in different parts of the brain. The efficiency of these indexes has been demonstrated in [98], calculating the correlation dimension of a group of EEG signals. In [98], other methods were used to calculate the synchronization of different brain areas. Synchronized indexes are a promising method for emotion recognition that deserves further research.

Table 4 shows the most commonly used algorithms and their respective mention percentages in the literature: (1) WT (26%), (2) PCA (19.7%), (3) Hjorth (17%), (4) ICA (11.3%), and (5) statistical measures (8.6%).

2.7. Feature Selection

The feature selection process is vital because it obtains the signal's properties that best classify the EEG characteristics. In BCI systems, the feature vector generally has high dimensionality [99]. Feature selection reduces the number of input variables for the classifier (not to be confused with dimensionality reduction). While both processes decrease the data's attributes, dimensionality reduction combines features to reduce their quantity.

A feature selection method does not change characteristics but excludes some according to specific usefulness criteria. Feature selection methods aim to achieve the best results by processing the least amount of data. It removes attributes that do not contribute to the classification because they are irrelevant (or redundant) for simpler classification models (which are faster and perform better). Additionally, feature selection methods reduce the overfitting likelihood in regular datasets, flexible (complex) models, or when the dataset has too many features but not enough observations. Flexible models correspond to more complex mathematical expressions that may work well in training but could present problems in new test data (overfitting).

One classification of feature selection methods based on the number of variables divides them into two classes: (1) Univariate and (2) multivariate. Univariate methods consider the input features one by one. Multivariate methods consider whole groups of characteristics together [100].

Another classification distinguishes feature selection methods as filtering, wrapper, and built-in algorithms [73].

- Filter methods evaluate features using the data's intrinsic properties. Additionally, most filtering methods are univariate, so each feature is self-evaluated. These methods are appropriate for large data sets because they are less computationally expensive.
- Wrapping methods depend on classifier types when selecting new features based on their impact on characteristics already chosen. Only features that increase accuracy are chosen.
- Built-in methods run internally in the classifier algorithms, such as deep learning. This type of process requires less computation than wrapper methods.

Examples of Feature Selection Algorithms

The following are some examples of algorithms for feature selection:

- Effect-size (ES)-based feature selection is a filter method.

  ES-based univariate: Cohen's is an appropriate effect size for comparing two means [101]. So, if two groups' means do not differ by 0.2 standard deviations or more, the difference is trivial, even if it is statistically significant. The effect size is calculated by taking the difference between the two groups and dividing it by the standard deviation of one of the groups. Univariate methods may discard features that could have provided helpful information.

  ES-based multivariate helps remove several features with redundant information, therefore selecting fewer features while retaining the most information [58]. It considers all the dependencies between characteristics when evaluating them—for example, calculating the Mahalanobis distance using the covariance structure of the noise.

Min-redundancy max-relevance (mRMR) is a wrapper method [102]. This algorithm compares the mutual information between each feature with each class at the output. Mutual information between two random variables x and y is calculated to identify if they are statistically independent [58]. mRMR maximizes $I(x_i, y)$ between each characteristic x_i and the target vector y; and minimizes the average mutual information $I(x_i, y_i)$ between two characteristics.

- Genetic algorithms reduce the feature vector's dimensionality using evolutionary methods, leaving only more informative features [2,86,97].
- Stepwise discriminant analysis SDA [73]. SDA extends the statistical tool for discriminant analysis that includes the stepwise technique.
- Fisher score is a feature selection technique to calculate the interrelation between output classes and each feature using statistical measures [102].

Table 5 shows feature selection algorithms and their usage percentage in the literature according to the systematic review prepared in the present thesis, which is shown in Section 3. Genetic algorithms are frequently used (32.3%), followed by SDA (17.7%), wrapper methods (15.6%), and mRMR (11.5%).

Table 5. Feature selection methods used in the literature (2015–2020) in (%).

Feature Selection Method	Literature's Usage Statistics % (2015–2022)
min-Redundancy Max-Relevance mRMR	11.5%
Univariate	5.3%
Multivariate	6.3%
Genetic Algorithms	32.3%
Stepwise Discriminant Analysis SDA	17.6%
Fisher score	7.3%
Wrapper methods	15.6%
Built-in methods	4.1%

2.8. Classification Algorithms

Model frameworks can categorize classification algorithms [56],[57]. The model's categories may be (1) generative-discriminative, (2) static-dynamic, (3) stable-unstable, and (4) regularized [103]–[105].

Two selection approaches for the classifier work best under certain conditions in emotion recognition [56]. The first identifies the best classifier for a given BCI device. The second specifies the best classifier for a given set of features.

For synchronous BCIs (cue-paced), dynamic classifiers and ensemble combinations have shown better performances than Support Vector Machine (SVM). For asynchronous BCIs (self-paced), the authors in this field have not determined an optimal classifier. However, dynamic classifiers perform better than static classifiers [56] because they better identify the onset of mental processes.

From the second approach, discriminative classifiers have been found to perform better than generative classifiers, principally in the presence of noise or outliers. Dynamic classifiers and SVM generally handle high dimensionality in the features better. If a small training set exists, simple techniques like Linear Discriminant Analysis (LDA) classifiers may yield satisfactory results [58].

Generative Discriminative

These classification models generally have supervised learning problems that fit the data's probability. A generative model specifies the distribution of each class using the joint probability distribution p(x,y) and Bayes theorem. A discriminative model finds the decision boundary between the categories using the conditional probability distribution p(y|x). Such a model includes the following classifiers: Naïve Bayes, Bayesian networks, Markov random fields, and hidden Markov models (HMM).

Static-Dynamic Classification

If the training time is offline or online, a classifier is categorized as static or dynamic. A static model trains the data once offline and then uses the trained model to classify a single feature vector. In a dynamic model, the system is online and updated continually. Thus, dynamic models can obtain a sequence of feature vectors and catch temporal dynamics [99].

Multilayer perceptron (MLP) can be considered a static classifier. Likewise, an example of a dynamic classifier is hidden Markov methods (HMM) because it can classify a sequence of feature vectors.

Stable Unstable

Stable classifiers usually have low complexity and do not affect their performance with minor variations of the training set. For example, k Nearest Neighbors (kNN) is a standard stable classifier. Unstable classifiers have high complexity and present considerable changes in performance with minor variations of the training set. Examples of unstable classifiers are linear support vector machine (SVM), multi-layer perceptron (MLP), and bilinear recurrent neural network (BLR-NN) [99].

Regularized

Regularization consists of carefully controlling classifier complexity to prevent overtraining. These classifiers have excellent generalization performance. Examples of these classifiers are regularized Fisher LDA (RF-LDA), linear SVM, and radial basis function kernel for support vector machine (RBF-SVM) [106].

General Taxonomy of Classification Algorithms

Another taxonomy divides classifiers using their properties to distinguish them into general algorithms: linear, neural networks, nonlinear Bayesian, nearest neighbor classifiers, and combinations of systems (ensemble). Most of the more specialized algorithms can be generated from these general types. Table 6 shows this taxonomy criterion with five different categories of general classifiers: (1) Linear, (2) neural networks, (3) nonlinear Bayesian, (4) nearest neighbor classifiers, and (5) combinations of classifiers or ensemble [44,56,58].

All general classifiers have characteristics of each of the previously mentioned framework models. For instance, SVM is discriminant, static, stable, and regularized; HMM is generative, dynamic, unstable, and not regularized; and kNN is discriminant, static, stable, and not regularized.

Consequently, the suggested guidelines for classifier selection are also applicable in this categorization. Table 6 presents the usage statistics of these classifiers in the 2015–2020 literature. The following are the most noteworthy classifiers: Neural networks CNN (46.16%), Linear classifiers SVM (30.3%), LDA (5.5%), Nearest Neighbors kNN (4.5%), and Ensembled classifier AdaBoost (3.9%).

Table 6. Categories of general classifiers.

Category of Classifier	Description	Examples of Algorithms in the Category	Advantages	Limitations	Literature's Usage Statistics % (2015–2022)
Linear	Discriminant algorithms that use linear functions (hyperplanes) to separate classes.	Linear Discriminant Analysis LDA [65]. Bayesian Linear Discriminant Analysis BLDA. Support Vector Machine SVM [107],[108]. Graph Regularized Sparse Linear Regularized GRSLR [31].	These algorithms have reasonable classification accuracy and generalization properties.	Linear algorithms tend to have poor outcomes in processing complex nonlinear EEG data.	LDA 5.20 BLDA 1.40 SVM 30.30 GRSLR 0.02
Neural networks (NN)	NN are discriminant algorithms that recognize the underlying relationship of data resembling the human brain operation.	Multilayer Perceptron MLP [109]. Long Short-term Memory Recurrent Neural Network LSTM-RNN [66–69]. Domain Adversarial Neural Network DANN [110]. Convolutional Neural Network CNN [68,70–73,[112]–[113]. Complex-Valued Convolutional Neural Network	NN generally yields good classification accuracy	Sensitive to overfitting with noisy and non-stationary data as EEGs.	MLP 1.60 LSTM 1.10 DANN0.20 CNN 46.36 CVCNN 0.40 GSCNN 0.40 GSLTFCNN 0.02 CapsNet-NN 0.10 GELM–NN 0.10 FBC-CNN 0.10

		CVCNN [107]. Gated-Shape Convolutional Neural Network GSCNN [107]. Global Space Local Time Filter Convolutional Neural Network GSLTFCNN [107]. CapsNet-NN Genetic Extreme Learning Machine GELM–NN [87].			
Nonlinear Bayesian classifier	Generative classifiers produce nonlinear decision boundaries.	Bayes quadratic BC [111]. Hidden Markov Model HMM [50],[114].	Generative classifiers reject uncertain samples efficiently.	For Bayes quadratic, the covariance matrix cannot be estimated accurately if the dimensionality is vast and there are insufficient training sample patterns.	BC 0.10 HMM 0.30
Nearest neighbor classifiers	Discriminative algorithms that classify cases based on their similarity to other samples	k-Nearest Neighbors kNN [115]. Mahalanobis Distance MD [116].	kNN has excellent performance with low-dimensional feature vectors. Mahalanobis Distance is a simple but efficient clear, suitable even for asynchronous BCI.	kNN has reduced performance for classifying high dimension feature vectors or noise distorted features.	KNN 4.5 MD 0.1
Combination of classifiers	Combined classifiers using	Ensemble methods can	Variance reduction	Quality measures are	Ensemble 2.1 Random-Forest 1.1

(ensemble-learning)	boosting, voting, or stacking. Boosting consists of several cascading classifiers. In voting, classifiers have scores, which yield a combined score per class, and a final class label. Stacking uses classifiers as meta-classifier inputs.	combine almost any type of classifier [117]. Random Forest [9], [118]. Bagging Tree BT [113], [117]. XGBoost [119] AdaBoost [120]	leads to an increase in classification accuracy.	application-dependent.	BT 0.2 XGBoost 0.4 AdaBoost 3.9

2.9. Performance Evaluation

Results must be reported consistently so that different research groups can understand and compare them. Hence, evaluation procedures need to be chosen and described accurately. Evaluation of the classifier's execution involves addressing performance measures, error estimation, and statistical significance testing [121]. Performance measures and error estimation configure the fulfillment rate of the classifier's function. The most recommended performance evaluation measures are shown in Table 7. They are confusion matrix, accuracy, error rating, and other measures obtained from the confusion matrix, such as the recall, specificity, precision, Area Under the Curve (AUC), and F-measure. Other performance evaluation coefficients are Cohen's kappa (k) [122], information transfer rate (ITR) [65], and written symbol rate (WSR) [122].

Performance evaluation and error estimation may need to complement a significance evaluation. This situation is because high accuracy can have little impact if the sample size is too small or classes are imbalanced (labeled EEG signals typically are). Therefore, significance classification is essential. General approaches can handle arbitrary class distributions to verify accuracy values significantly above certain levels. Used methods are the theoretical level of random classification and adjusted Wald confidence interval for classification accuracy.

The theoretical level of the random classification test is the sum of the products between the experimental results' classification probability and the probability calculated if all the categorization randomly occurs. This approach can only be used after performing the classification [123].

Adjusted Wald confidence interval gives the lower and upper confidence limits for the probability of the correct classification, which specifies the intervals for the classifier performance evaluation index [124].

Table 7. Conventional performance evaluation methods for BCI.

Performance Evaluation	Main characteristics	Advantages	Limitations
Confusion matrix	The confusion matrix presents the number of correct and erroneous classifications specifying the erroneously categorized class.	The confusion matrix gives insights into the classifier's error types (correct and incorrect predictions for each class). It is a good option for reporting results in M-class classification.	Results are difficult to compare and discuss. Instead, some authors use some parameters extracted from the confusion matrix.
Accuracy and error rate	The accuracy p is the probability of correct classification in a certain number of repeated measures. The error rate is $e = 1 - p$ and corresponds to the probability that an incorrect classification has been made.	It works well if the classes are balanced, i.e., an equal number of samples belong to each class.	Accuracy and error rate do not consider whether the dataset is balanced or not. If one class occurs more than another, the evaluation may appear high for accuracy even though the classification is not performing well. These parameters depend on the number of classes and the number of cases. In a 2-class problem, the chance level is 50%, but with a confidence level depending on the number of cases.
Cohen's kappa (k)	k is an agreement evaluation between nominal scales. This index measures the agreement between a true class and a classifier output. 1 is a perfect agreement, and 0 is a pure chance agreement.	Cohen's kappa returns the theoretical chance level of a classifier. This index evaluates the classifier realistically. If k is low, the confusion matrix would not have a meaningful classification even with high accuracy values. This coefficient presents more information than simple percentages because it uses the	This coefficient has to be interpreted appropriately. It is necessary to report the bias and prevalence of the k value and test the significance for a minimum acceptable level of agreement.

		entire confusion matrix.	
Sensitivity or Recall	Sensitivity, also called Recall, identifies the true positive rate for describing the accuracy of classification results. It evaluates the proportion of correctly identified true positives related to the sum of true positives plus false negatives.	Sensitivity measures how often a classifier correctly categorizes a positive result.	The Recall should not be used when the positive class is larger (imbalanced dataset), and correct detection of positive samples is less critical to the problem.
Specificity	Specificity is the ability to identify a true negative rate. It measures the proportion of correctly identified true negatives over the sum of the true negatives plus false positives. The False Positive Rate (FPR) equals 1 – Specificity.	Specificity measures how often a classifier correctly categorizes a negative result.	Specificity focuses on one class only, and the majority class biases it.
Precision	Precision also referred to as Positive Predicted Value, is calculated as 1 – False Detection Rate (F). The false detection rate is the ratio between false positives and the sum of true positives and false positives.	Precision measures the fraction of correct classifications.	Precision should not be used when the positive class is larger (imbalanced dataset), and correct detection of positive samples is less critical to the problem.
ROC	The ROC curve is a Sensitivity plot vs. the False Positive Rate. The area under the ROC curve measures how well a parameter can distinguish between a true positive and a true negative.	ROC curve provides a measure of the classifier performance across different significance levels.	ROC is not recommended when the negative class is smaller but more important. The Precision and Recall will primarily reflect the ability to predict the positive class if it is larger in an imbalanced dataset.
F-Measure	F-Measure is the harmonic mean of Precision and Recall. It is useful because as the Precision increases, Recall decreases, and vice versa.	F-measure can handle imbalanced data. F-measure (like ROC and kappa) measures the classifier performance across different significance levels.	F-measure does not generally take into account true negatives. True negatives can change without affecting the F-measure.
Pearson correlation coefficient	Pearson's correlation coefficient (r) quantifies the degree of a ratio between the true and predicted values by a value ranking from −1 to +1.	Pearson's correlation is a valid way to measure the performance of a regression algorithm.	Pearson's correlation ignores any bias between the true and the predicted values.

Information transfer rate (ITR)	As BCI is a channel from the brain to a device, it is possible to estimate the bits transmitted from the brain. ITR is a standard metric for measuring the information sent within a given time in bits per second.	ITR is a metric that contributes to the criteria to evaluate a BCI System.	ITR is often misreported due to inadequate understanding of many considerations as delays are necessary to process data, present feedback, and clear the screen. TR is best suited for synchronous BCIs over user-paced BCI.

3. LITERATURE REVIEW

The present work started with developing a survey that followed the guidelines of Kitchenham presented in [125]. This guide suggests three phases with their respective activities. Therefore, we considered planning, conducting, and reporting phases in our literature research.

In the planning phase, the research questions and search chains are defined. In the conducting the review phase, we established the search strategy. We used Semanticscholar.org to find sources because it links to journals and conference proceedings' major databases and set the search to journals and conferences with a date range from 2015 to 2022. In this phase, we also conduct the searches, obtain initial results and apply inclusion and exclusion criteria to them. Additionally, we used quality assessment filters. The final papers were used to extract information in tables for the reporting phase.

Based on the goals of our research. We define the following literature review's objectives:

Obj1: Which computational systems with their respective performance results for emotion recognition EEG-based BCI devices are in the literature between the date range?

Obj2: Which emotion elicitation methods are used in computational systems for emotion recognition applying EEG-based BCI devices can be found in the literature in the specified date range?

Obj3: How many electrodes use the EEG-based BCI devices in computational systems for emotion recognition?

Obj4: Are any poker games used as emotion elicitation methods in computational systems for emotion recognition using EEG-based BCI devices?

Obj5: Are any stock market activities used as emotion elicitation methods in computational systems for emotion recognition using EEG-based BCI devices?

The analyzed papers' were found using the following search chains with the respective result number for journal and conference papers in the date range:

SCh1: Computational systems emotion recognition EEG-based BCI devices. This search got 113 papers.

SCh2: Emotion elicitation methods emotion recognition EEG-based BCI devices. This search obtained 67 papers.

SCh3: Computational systems algorithms emotion recognition EEG-based BCI devices number of electrodes. This search produced four papers.

SCh4: Poker games emotion elicitation emotion recognition EEG-based BCI. This chain generated seven papers

SCh5: Stock market activities emotion elicitation emotion recognition EEG-based BCI. This search obtained two papers.

The obtained studies also allowed a comparison of results while considering the classified number of emotions.

Adding the results of each search, we obtained 193 papers. Each article was read to have complete information to guide the application of inclusion and exclusion filters. The inclusion criteria were: (1) The articles were published in the considered period from 2015 through 2022 in peer-reviewed journals and conferences, (2) they constitute emotion recognition systems that used EEG-based BCI devices with a focus on computational intelligence applications, and (3) they include experimental setups and performance evaluations. We applied exclusion criteria in this literature review stage and eliminated redundant papers.

With these considerations, 36 journal studies and 24 conference papers were selected. Finally, from these 60 papers, we applied additional quality assessments using the Kitchenham methodology. With these quality criteria, we eliminated articles that did not have detailed results and documents that were medical studies for diagnosis or evaluation because they have a different perspective than ours. We obtained a sample of 35 articles at the end of these filters to show a summary of technical details, components, and algorithms, as presented in Table 8. As illustrated in Tables 3, 4, 5, 6, 8, and 9, we extracted this group's statistical data about computational techniques to detect trends and perform a comparative analysis.

Applying the Kitchenham methodology to select these 35 articles, we evaluated the papers using quality assessments that include an evaluation of different criteria as the use of datasets with varying forms of emotion elicitation, types and number of feature extraction and selection methods, number of classification algorithms used, number and variety of classified emotions (valence-arousal space or discrete emotions), and level of accuracy results. With this scope, Table 8 summarizes the research in this field from 2015 to 2022.

The following components characterize the systems presented in Table 8: (1) Stimulus type; (2) databases generated by the paper's authors or publicly available; (3) the number of participants; (4) extraction and selection of characteristics; (5) features; (6) classification algorithms; (7) number and types of classes; and (8) performance evaluation.

The applied preprocessing methods are mostly similar in the reviewed studies. Their primary preprocessing methods are standard, i.e., artifacts removal outside the EEG frequency bands and a notch filter for 50 or 60 Hz to eliminate electrical noise. For the above, this information was omitted in Table 8.

3.1. Emotion Elicitation Methods

This section analyzes research papers that used different resources to provoke emotions in their subjects. These stimuli are music videos, film clips, music tracks, self-induced disgust (produced by remembering an unpleasant odor), and risky situations in a flight simulator as examples of active elicitation of emotions. EEG-based BCI systems frequently use public DEAP and SEED databases that apply music videos and film clips as stimuli, respectively. Different stimuli provoke emotions that affect different brain areas and produce EEG signals that can be recognized concerning specific emotions. Figure 5 shows the frequency in which different emotion elicitation methods are applied to generate datasets used in the reviewed systems.

Few research papers resort to more elaborate platforms to provoke "real life" emotions. However, such methods have been applied to other physiological responses (other than EEG like skin conductance, respiration, electrocardiogram (ECG), and facial expressions, among others) [126]. Some authors state that stimuli that provoke wide-ranging emotions could make exploring the brain's mechanisms activated for specific emotion generation challenging. In this sense, focusing on a particular emotion could improve our understanding of such mechanisms. For our research sample, we highlighted research pieces that study emotions, such as dislike and disgust, separately [38],[127].

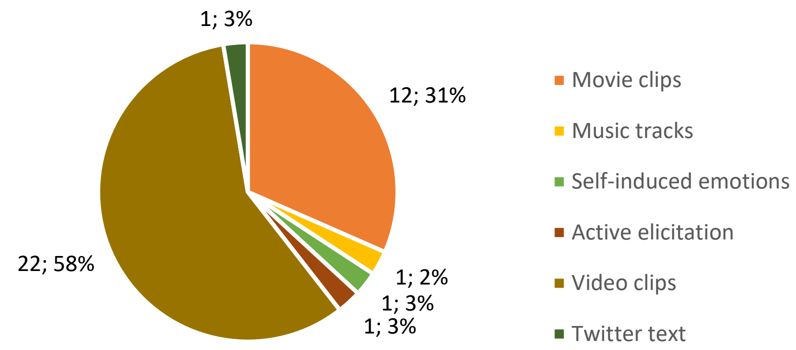


Figure 5. Emotion elicitation methods.

3.2. Number of Participants to Generate Datasets

Figure 6 presents the number of participants in the experiments to obtain EEG datasets to train and test the emotion recognition systems. Most of the systems used a number of subjects ranging from 31–40 (53%) and 11–20 (32%). For more than 40 participants, there are no datasets. The targeted studies used EEG data from healthy individuals.

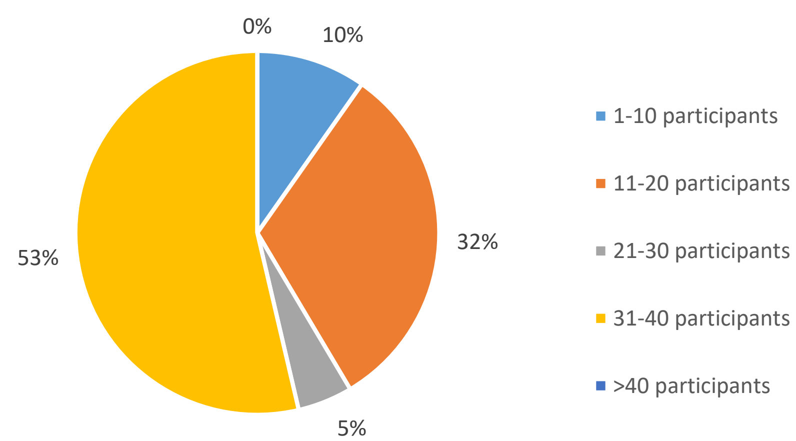


Figure 6. Number of participants in EEG datasets.

Table 8. Emotion recognition systems using BCI [1]

Reference/Year	Stimuli	EEG Data	Feature Extraction	Feature Selection	Features	Classification	Emotions	Accuracy
[128]/2016	-	DEAP	Computation in the time domain, Hjorth, Higuchi, FFT	mRMR	Statistical features, BP, Hjorth, FD	RBF NN SVM	3 class/Arousal 3 class/Valence	Arousal/60.7% Valence/62.33%
[78]/2015	15 movie clips	Own dataset/15 participants	DBN	-	DE, DASM, RASM, DCAU, from Delta, Theta, Alpha, Beta, and Gamma.	kNN LR SVM DBNs	Positive Neutral Negative.	SVM/83.99% DBN/86.08%
[38]/2015	Self-induced emotions	Own dataset/10 participants	WT	PCA	Eigenvalues vector	SVM	Disgust	Avg. 90.2%
[129]/2018	Video clips	Own dataset/10 participants	Higuchi	-	FD	RBF SVM	Happy Calm Angry	Avg. 60%
[130]/2017	Video clips	Own dataset/30 participants	STFT, ERD, ERS	LDA	PSD	LIBSVM	Joy Amusement Tenderness Anger Disgust Fear Sadness Neutrality	Neutrality 81.26% 3 Positive emotions 86.43% 4 Negative emotions 65.09%
[127]/2020	-	DEAP	DFT, DWT	-	PSD, Logarithmic compression of Power Bands, LFCC, PSD, DW	NB CART kNN RBF SVM SMO	Dislike	Avg. SMO/81.1% NB/63.55% kNN/86.73% CAR/74.08%

Reference/Year	Stimuli	EEG Data	Feature Extraction	Feature Selection	Features	Classification	Emotions	Accuracy
[79]/2019	-	DEAP and SEED-IV	Statistics in the time domain, FFT, DWT	-	PSD, Energy, DE, Statistical features	SVM	HAHV HALV LALV LAHV	Avg DEAP/79% Avg.SEED/76.5%
[13]/2016	Music tracks	Own dataset/30 participants	STFT, WT	-	PSD, BP Entropy, Energy, Statistical features, Wavelets	SVM MLP kNN	Happy Sad Love Anger	Avg. SVM/75.62% MLP/78.11% kNN/72.81%
[84]/2017	-	SEED	FFT and electrode location	Max Pooling	DE, DASM, RASM, DCAU	SVM ELM Own NN method	Positive Negative Neutral	Avg. SVM/74.59% ELM/74.37% Own NN/86.71%
[48]/2019	Video clips	Own dataset/16 participants	STFT, WT, Hjorth, AR	-	PSD, BP, Quadratic mean, AR Parameters, Hjorth	SVM	Happy Sad Fear Relaxed	Avg. 90.41%
[131]/2019	-	DEAP	WT	-	Wavelets	LSTM RNN	Valence Arousal	Avg. 59.03%
[132]/2018	-	SEED	Embedded LSTM	-	DE	BiDANN	Positive Negative Neutral	Avg. 92.38%
[113]/2019	-	DEAP	Statistics in the time domain and FFT		Statistical characteristics. PSD	BT SVM LDA BLDA CNN	Valence Arousal	Avg. for combination features AUC BT/0.9254 BLDA/0.8093 SVM/0.7460 LDA/0.5147 CVCNN/0.9997 GSCNN/1 GSCNN/1

Reference/Year	Stimuli	EEG Data	Feature Extraction	Feature Selection	Features	Classification	Emotions	Accuracy
[120]/2017	-	DEAP	Statistics in the time domain and FFT	GA	Statistical characteristics, PSD, and nonlinear dynamic characteristics	AdaBoost	Joy Sadness	95.84%
[133]/2019	-	DEAP	STFT, NMI	-	Inter-channel connection matrix based on NMI	SVM	HAHV HALV LALV LAHV	Arousal/73.64% Valence/74.41%
[73]/2018	-	SEED	FFT	SDA	Delta, Theta, Alpha, Beta, and, Gamma	LDA	Positive Negative Neutral	Avg. 93.21%
[114]/2019	-	SEED	FFT	-	Electrodes-frequency Distribution Maps (EFDMs)	CNN	Positive Negative Neutral	Avg. 82.16%
[85]/2019	-	SEED/ DEAP/ MAHNOB-HCI	Statistics in the time domain and FFT	Fisher-score, classifier-dependent structure (wrapper), mRMR, SFEW	EEG-based network patterns (ENP) PSD, DE, ASM, DASM, RASM, DACU, ENP, PSD + ENP, DE + ENP	SVM GELM	Positive Negative Neutral	Best feature F1 SEED/DE+ENP gamma 0.88 DEAP/PSD+ENP gamma 0.62 MAHNOB/PSD+ENP Gamma 0.68
[96]/2019	-	DEAP	Embedded	Sparse group lasso	Granger causality feature	CapsNet Neural Network	Valence-arousal	Arousal/87.37% Valence/88.09%
[31]/2019	Video clips	Own dataset RCLS/14 participants. SEED	Statistics in the time domain, WT	-	HOC, FD, Statistics, Hjorth, Wavelets	GRSLR	Happy Sad Neutral	81.13%

Reference/Year	Stimuli	EEG Data	Feature Extraction	Feature Selection	Features	Classification	Emotions	Accuracy
[134]/2019	-	DEAP	Statistics in the time domain, FFT, WT	Correlation matrix, information gain, and sequential feature elimination	Statistical measures, Hjorth, Autoregressive parameters, frequency bands, the ratio between frequency bands, wavelet domain features	XGBoost	Valence, arousal, dominance, and liking	Valence/75.97% Arousal/74.20% Dominance/75.23% Liking 76.42%
[135]/2015	-	DEAP	Frequency phase information	Sequential feature elimination	Derived features of bispectrum	SVM	Low/high valence, low/high arousal	Low-high arousal/64.84% Low-high valence/61.17%
[136]/2016	-	DEAP	Higuchi, FFT	-	FD, PSD	SVM	Valence, arousal	Valence/86.91% Arousal/87.70%
[137]/2017	-	DEAP	DWT	-	Discrete wavelets	kNN	Valence, arousal	Valence/84.05% Arousal/86.75%
[138]/2015	-	DEAP	RBM - embedded	-	Raw signal-6 channels	Deep-Learning	Happy, calm, sad, scared	Avg. 75%
[139]/2017	-	DEAP	DWT	Best classification performance for channel selection	Discrete wavelets	MLP kNN	Positive, negative	MLP/77.14% kNN/72.92%
[140]/2017	-	DEAP	Embedded LSTM	-	-	LSTM NN	Low/high valence, Low/high arousal, Low/high liking	Low-high valence/85.45% Low-high arousal/85.65% Low-high liking/87.99%
[141]/2018	-	DEAP	Embedded CNN	-	-	3D-CNN	Valence, arousal	Valence/87.44% Arousal/88.49%

Reference/Year	Stimuli	EEG Data	Feature Extraction	Feature Selection	Features	Classification	Emotions	Accuracy
[142]/2018	-	DEAP	FFT, phase computations, Pearson correlation	-	PSD, phase, phase synchronization, Pearson correlation	CNN	Valence	Low-high valence/96.41%
[37]/2019	Flight simulator	Own dataset/8 participants	Statistics in time domain, and WT	-	Statistical measures, DE, Wavelets	ANN	Happy, Sad, Angry, Surprise, Scared	Avg. 53.18%
[143]/2021	-	SEED DEAP	CCWT	DE, Mutual Information Matrix	Time and frequency behavior	SVM	SEED: positive, negative, neutral DEAP: 4 classes HAHV HALV LALV LAHV	SEED: 96.3% DEAP: 81.1%
[144]/2021	Text (Twitter data)	Own dataset/2 participants	STFT	-	Frequency bands	RF Decision Tree SVM	Happiness Sadness Fear Anger	RF: 98% Decision Tree: 88% SVM: 32%
[145]/2021	-	DEAP DREAMER	Deep Forest - Embedded	-	Spatial and temporal	Deep Forest	DEAP: Low/high Valence, Low/high Arousal DREAMER: Low/high Valence, Low/high Arousal, Low/high Dominance	DEAP Low-high valence/97.59%, Low-high arousal/97.53% DREAMER Low-high valence/89.03% Low-high arousal/: 90.41% Low-high dominance/ 89.89%

Reference/Year	Stimuli	EEG Data	Feature Extraction	Feature Selection	Features	Classification	Emotions	Accuracy
[146]/2022	-	DEAP DREAMER AMIGOS SEED	Embedded CNN	ReliefF	Embedded features in 10 different channels for each dataset	SVM	Low/high Valence, Low/high Arousal, Low/high Dominance.	DEAP Low-high valence/ 83.26% Low-high arousal/ 83.85% Low-high dominance/ 88.58% DREAMER Low-high valence/ 90.76% Low-high arousal/ 92.92% Low-high dominance/ 92.97% AMIGOS Low-high valence/ 88.54% Low-high arousal/ 91.51% Low-high dominance/ 90.34% SEED Low-high valence/ 88.19%
[147]/2022	-	DEAP	Embedded FBCCNN	-	Frequency band correlation features	FBCCNN	Valence - Arousal	70.34%

[1] Autoregressive Parameter (AR). Bagging Tree (BT). Band Power (BP). Bayesian linear discriminant analysis (BLDA). Bi-hemispheres Domain Adversarial Neural Network (BiDANN). Convolutional Neural Network (CNN). Complex Continuous Wavelet Transform (CCWT), Complex-Valued Convolutional Neural Network (CVCNN). Gated-Shape Convolutional Neural Network (GSCNN). Global Space Local Time Filter Convolutional Neural Network (GSLTFCNN). Deep Belief Networks (DBNs). Differential entropy (DE). DE feature Differential Asymmetry (DASM). DE feature Rational Assimetry (RASM). DE feature Differential Caudality (DCAU). Electrooculography (EOG). Electromyogram (EMG). Event-Related

Desynchronization (ERD) and Synchronization (ERS). Feature selection and weighting method (SFEW). Fractal dimensions (FD). Frequency Band Correlation Convolutional Neural Network. Genetic Algorithm (GA). Graph regularized Extreme Learning Machine (GELM) NN. Graph Regularized Sparse Linear Regularized (GRSLR). High Order Crossing (HOC). Linear Discriminant Analysis (LDA). Logistic Regression (LR). Long short-term memory Recurrent Neural Network (LSTM RNN). Minimum-Redundancy-Maximum-Relevance (mRMR). Normalized Mutual Information (NMI). Principal Component Analysis (PCA). Radial Basis Function (RBF). Random Forest (RF). Short-Time Fourier Transform (STFT). Stepwise Discriminant Analysis (SDA). Support Vector Machine (SVM). Wavelet Transform (WT),

3.3. Datasets

Figure 7 presents the usage percentage of datasets used in emotion recognition. DEEP and SEED are publicly available databases and most frequently used (48% and 23% of applications, respectively). Other studies used self-generated datasets (23%), which are typically not freely accessible. The MAHNOB-HCI and RCLS public datasets appeared in our research sample for the literature review, with a participation of 3% each.

Systems that use public databases offer some comparability, but contrast is limited even if the same characteristics are handled. Still, such public databases could eventually lead to findings if objective comparisons are performed.

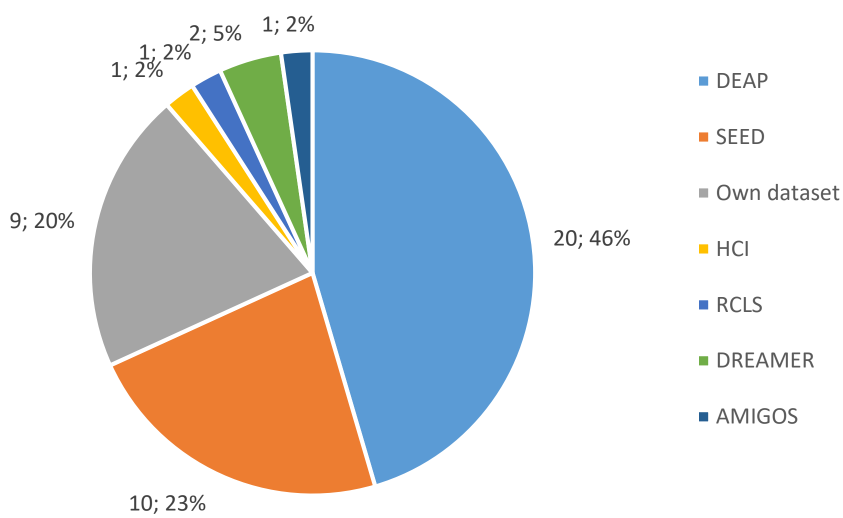


Figure 7. EEG datasets for emotion recognition.

3.4. Most used Feature Extraction algorithms

Most systems use feature extraction methods in the time, frequency, time-frequency, or space domains. A small percentage of works evaluate the functional connectivity (or differences) in the observed activity between brain regions when emotions are provoked. Features with non-redundant information combined from different domains yield better classification results. However, it is still unclear if features work better alone or in combination with each other or which type of features are more relevant for emotion recognition.

Our review found that researchers addressed these issues by developing feature extraction algorithms that outperform the classic frequency bands and extract as much information as possible from brain signals. Further developments should be connected to a comprehensive understanding of the brain's neurophysiology.

Figure 8 presents the domains of the used features. Frequency domain features are the most frequently used (38.6%) and appear more often as time domain (24%) or time-frequency domain features (19.3%). Asymmetry characteristics between electrode pairs (by each hemisphere) are used in 1.8% of the occasions. Additionally, raw data (without features) is used for deep learning classifiers (15.8%) that produce embedded or internal features.

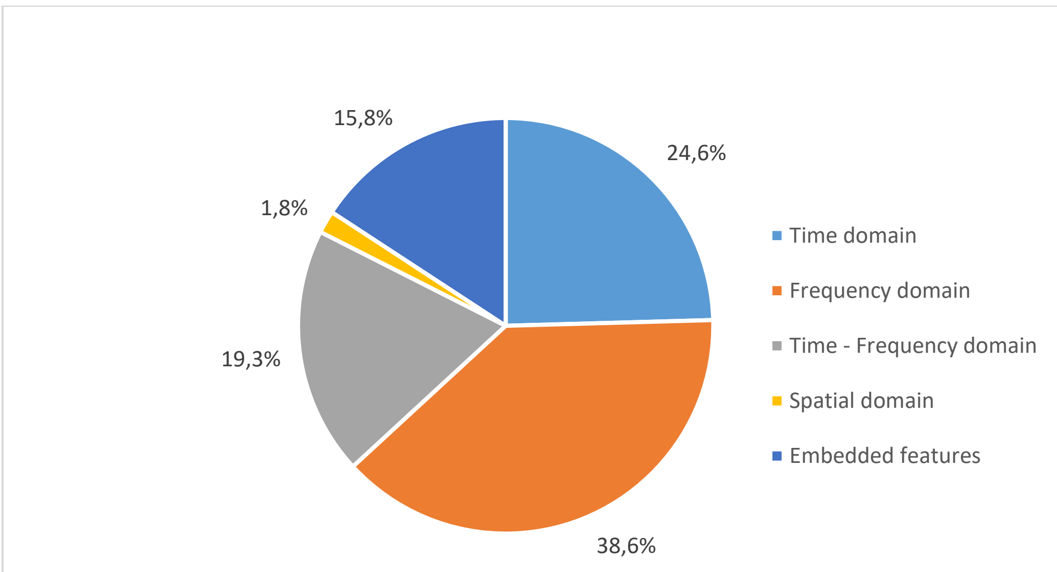


Figure 8. The domain of used features.

In the 35 papers shown in Table 8, we calculated the percentage of use of the feature extraction algorithms. Figure 9 presents these percentages.

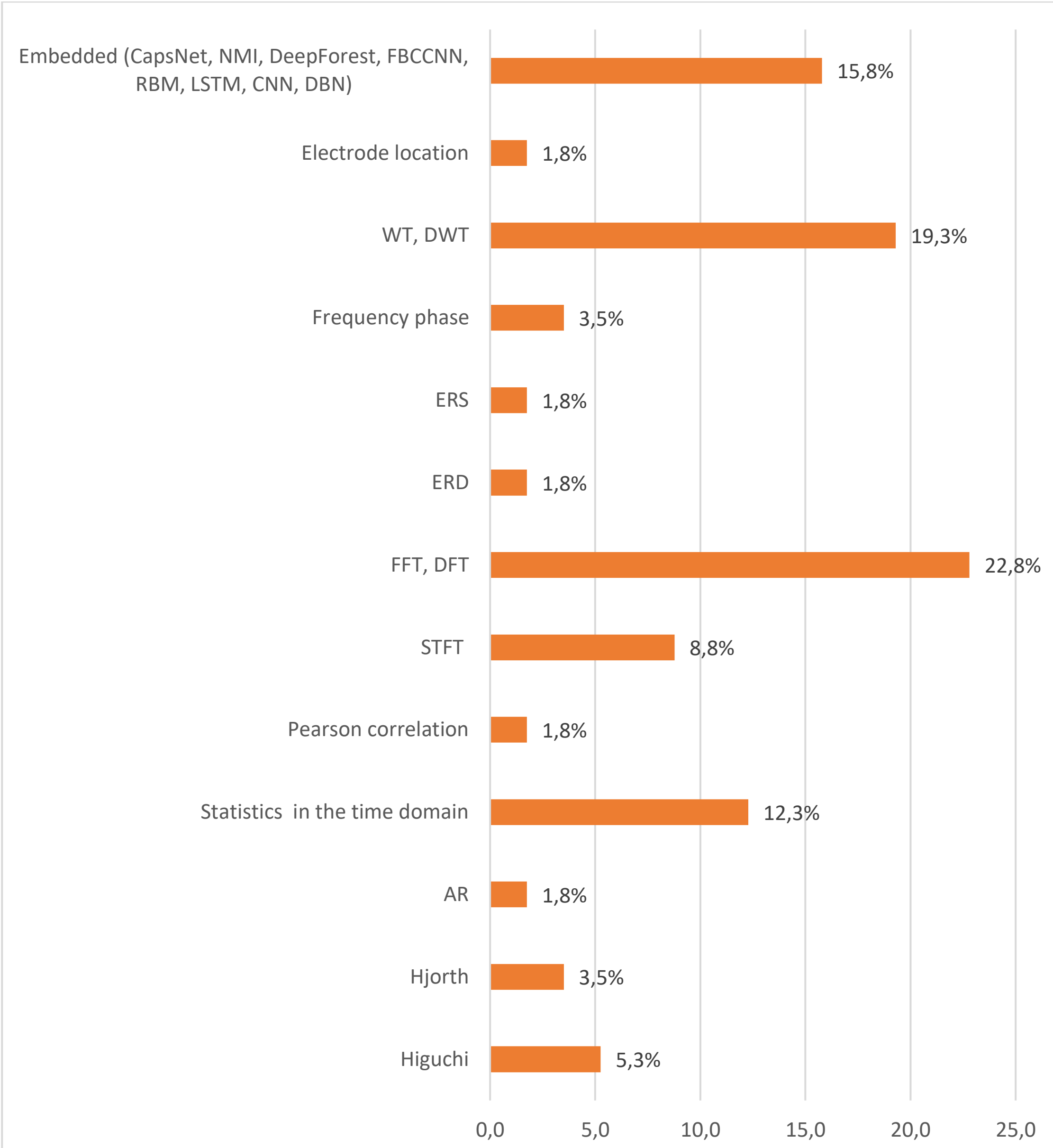


Figure 9. Percentage of the use of algorithms for feature extraction from Table 8.

We observed an increasing presence of algorithms embedded in neural networks like Restricted Boltzmann Machine (RBM), Deep Belief Networks (DBN), Normalized Moment of Inertia (NMI), Capsule Network (CapsNet), Convolutional Neural Networks (CNN), Deep Forest, Frequency Band Correlation CNN (FBCCNN), and Long Short Term Memory (LSTM) (15.8%) that are used to extract signal features automatically from raw data. This approach yields a good enough classifier performance because it preserves information and avoids the risk of removing essential emotion-related signal features.

1.8% of the studies use spatial features with signal vs. electrode location information.

Features in the time-frequency domain are extracted using Wavelets (WT) and Discrete Wavelet Transform in 19.3% of the analyzed studies.

Features in the frequency domain are Frequency Phase with a presence of 3.5%, Event Related Desynchronization (ERD) with 1.8% of the occurrences, Event Related Synchronization (ERS) with 1.8%, Fast Fourier Transform and Discrete Fast Fourier were used in 22.8% of the occasions, and Short Time Fourier Transform (STFT) that appeared 8.8%.

Features in the time domain are Pearson Correlation with 1.8%, Statistics with 12.3% of occurrences, Autoregression (AR) with 1.8%, Hjorth that appeared 3.5%, and Higuchi with 5.3% of the studies.

3.5. Most used Feature Selection Methods

It is worth noting that 61.3% of the systems presented in Table 8 do not use a feature selection method. Table 9 lists the systems that utilized feature selection algorithms. Interestingly, virtually every system uses a different algorithm except for the methods minimum redundancy maximum relevance (mRMR) and recursive feature elimination, which are utilized for two other schemes.

Table 9. Systems in Table 8 using feature selection algorithms.

Feature Selection Algorithm	Reference
mRMR	[80,126]
PCA	[38]
LDA	[128]
Max Pooling	[79]
Genetic Algorithm	[118]
SDA	[75]
Fisher-score	[80]
SFEW	[80]
Sparse group lasso	[96]
Correlation matrix	[132]
Information gain	[132]
Recursive feature elimination	[132,133]
Best classification performance for channel selection	[137]
ReliefF	[48]

3.6. Classifiers

Figure 10 shows that most classifiers were linear (48%) and neural networks (41%); a few papers used nearest neighbors (7%) and ensemble methods (5%). Consequently, it is worth mentioning that the following algorithms have become increasingly popular for EEG-based emotion recognition applications:

- Linear classifiers, such as naïve Bayes (NB), logistic regression (LR), support vector machine (SVM), and linear discriminant analysis (LDA) (48% of use).
- Neural networks like multi-layer perceptron (MLP), radial basis function RBF, convolutional neural network (CNN), deep belief networks (DBN), extreme learning method (ELM), graph regularized extreme learning machine (GELM), long short term memory (LSTM), domain adversarial neural network (DANN), CapsNet, and graph regularized sparse linear regularized (GRSLR) (41% of use).
- Ensemble classifiers like random forest, CART, bagging tree, Adaboost, and XGBoost are less used (5%). The same situation occurs with the kNN algorithm despite its consistently good performance results, probably because it works better with a simpler feature vector (7%).

This review did not find studies that applied non-linear Bayesian classifiers as hidden Markov models (HMM) during our considered period.

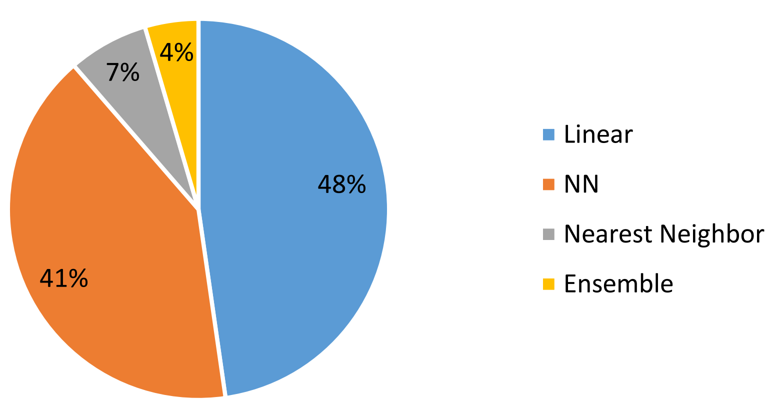


Figure 10. Classifiers’ usage

3.7. Performance vs. the Number of Classes-Emotions

The performance of almost all systems was evaluated using accuracy, except for two systems that used the area under the curve (AUC), and the other presented an F1 measure. Unfortunately, EEG datasets are usually unbalanced, with one or two labeled emotions more numerous than the others, which is somewhat problematic for this approach. Thus, this situation could lead to biased classifications. Moreover, EEG datasets are typically unbalanced, and performance measures should be calculated to contextualize their outcomes. Unbalanced datasets are those in which one or some classes are significantly more numerous than the others. This situation causes a classification bias for the more frequent categories. In this scenario, the minority classes fail to be recognized, but the classifier's performance seems high despite this. Treatment to prevent this problem varies greatly from study to study because they could use different criteria. In our view, this is one of the reasons for such results are not entirely comparable among the various studies. In our opinion, this is why such results are not entirely comparable among different studies.

Figure 11 presents the relationship between systems and the number of classified emotions. Most systems use the VA or VAD spaces and classify each dimension as a bi-class (for instance, valence positive and negative; arousal high-value and low value) or tri-class problem (valence positive, neutral, and negative; arousal and dominance high-value and low-value).

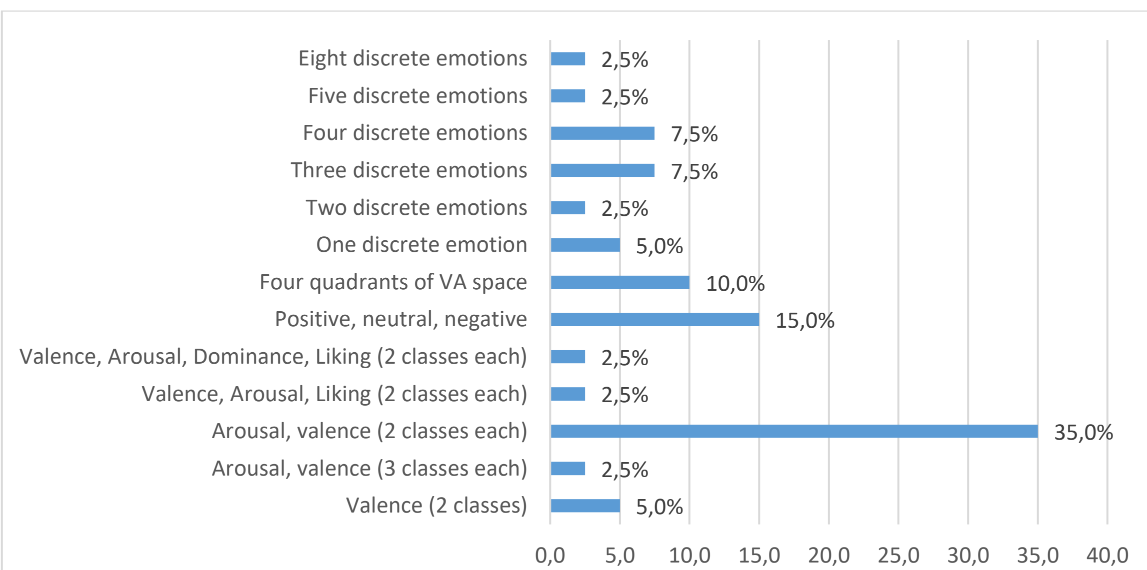


Figure 11. Percentage of systems with different numbers of classified emotions.

The two most used classification criteria are arousal and valence, categorized separately, with 35.0% of occurrences. On the other hand, 15.0% categorized valence with three classes: Positive, neutral, and negative. Then, 7.5% classified three discrete emotions: sadness, love, and anger. Moreover, 5.0% ranked valence as two classes (positive and negative), four discrete emotions (happy, sad, fear, and relaxed) were used by 7.5% of studies, and one discrete emotion (disgust in one system and dislike in other) had 5% of occurrences. Feelings located in one of four quadrants of the VA space (high valence-high arousal, high valence–low arousal, low valence–high arousal, and low valence–low arousal) had 10% of use.

Classifier performance should be evaluated, considering that accuracy would be inversely proportional to the number of detected emotions. In other words, classification accuracy should be higher than a random classification process with an equal chance for each class. As the number of classification classes increases, a random classification process would yield a lower accuracy. Consequently, a classification between two classes must be greater than 50% to be better than a random result; in the same way, a classification between three categories must be greater than 33% to be considered better than a slipshod result, and so on. Therefore, such accuracy metrics should provide the classification performance benchmark for our evaluations.

Although the systems' performance results depend on many factors, it is possible to find some relationship between the number of classes, the type of emotions classified, and the accuracy obtained (Figure 12). The best results are obtained with two classes, either discrete emotions or positive or negative values in a dimensional space. The second-best value recognizes one negative discrete emotion like dislike or disgust. The result that the classification of one emotion does not obtain the best performance value could be explained by the fact that in our review, we observed that negative emotions are more challenging to classify and tend to yield smaller performance values.

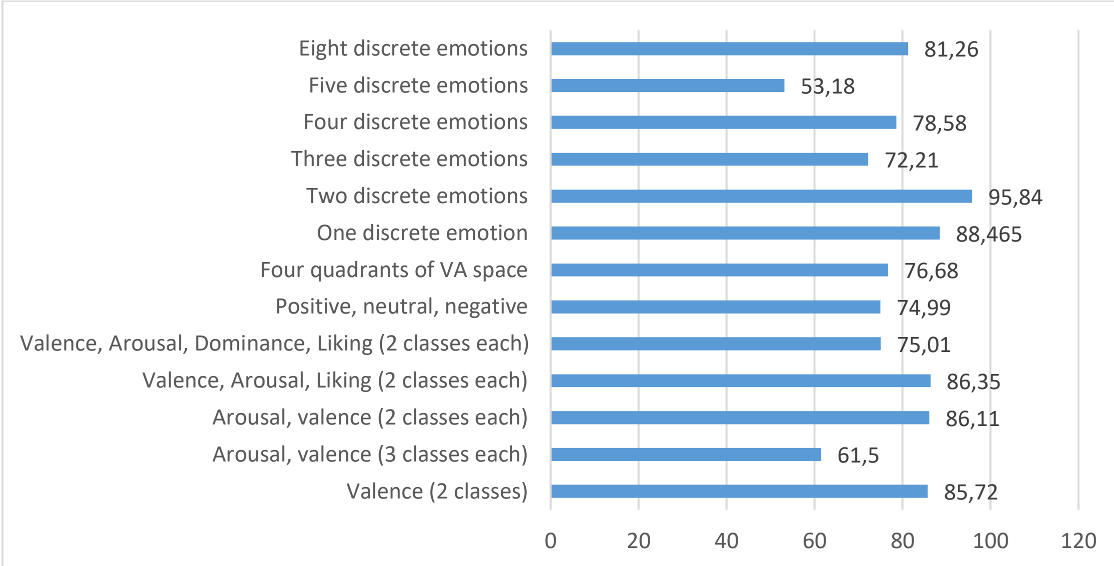


Figure 12. Accuracy vs. types and number of classified emotions.

It is crucial to notice that the average accuracy for our type of results for four quadrants in the VA space is 76.68%.

Comparing approaches and results obtained through different BCI-based systems is complex. Each system uses diverse experimental methods for emotion elicitation, protocols to detect EEG signals, datasets, extraction and selection of features, and classification algorithms; generally speaking, each implementation has different settings. Ideally, systems should be tested under similar conditions, but that scenario is not yet available. However, we have performed a comparative analysis to extract trends, bearing in mind such limitations.

3.8 Remarks

EEG signals are reliable information that cannot be simulated or faked. Decoding EEG and relating these signals to specific emotions is a complex problem. Affective states do not have a simple mapping with particular brain structures because different emotions activate the same brain locations, or conversely, a single emotion can activate several structures.

In recent years, EEG-based BCI emotion recognition has been a computing field that has generated much interest. Significant advances in developing low-cost BCI devices with increasingly better usability have encouraged numerous research studies.

In this section, we reviewed the different algorithms and processes that can be part of EEG-based BCI emotion recognition systems: (1) Emotion elicitation, (2) signal acquisition, (3) feature extraction and selection, (4) classification techniques, and (5) performance evaluation. For our survey on this topic, we mined different databases. We selected 35 studies from a computer science perspective to gain insight into the state of the art and suggest future research efforts.

This literature review responds to the questions raised in the Introduction chapter of the present document regarding the EEG emotion recognition datasets available, the emotion elicitation methods, and the number of channels they use. Also, the most used feature extraction and selection algorithms, classification approaches, and their respective values for performance evaluation of the different systems that are presented in the literature in the established date range search.

This study shows that computational methods still do not have standards for various applications. Researchers are continuing to look for new solutions in an ongoing effort. The relationship between brain signals and emotions is a complex problem, and novel methods and new implementations for emotion recognition are continuously presented. We expect that many existing challenges will soon be solved and pave the way for a vast area of possible applications using EEG-based emotion recognition.

The hypothesis established in the Introduction section (page 15) was generated from the niches recognized in this literature review.

4. DATA ACQUISITION FOR EMOTION RECOGNITION

As was mentioned in Chapter 3, in emotion recognition, datasets are generated with several emotion elicitation methods. For instance, the participants watch music video clips, and pieces of films, listen to sounds or music, view images, dyadic interactions, video games, flight simulators, virtual reality immersion, and execute memory recall to auto-induced emotion.

Such approaches have different characteristics and merits. The most frequently utilized among the previously listed are the passive methods, i.e., those that involve films, music, and images. However, only a few studies used interactive stimuli such as games or a flight simulator to induce emotions [148][149]. To the best of our knowledge, currently, none of the procedures for emotion elicitation are applied in an environment resembling everyday work activities in a competitive environment such as trading in the stock market. Also, the recognition of emotions of individuals playing poker has not been studied using EEG-based BCI devices, either.

In this context, two datasets were generated using BCI devices to sensor EEG signals while the experiment's participants were engaged in emotion-provoking activities. The first dataset is related to emotions induced by trading in the stock market; the second one has information recollected from poker players. Our research proposes methods of emotion elicitation that proved helpful in provoking emotions.

In our searches, for the date range from 2015 to 2022, two results appear when in semanticscholar.com, we enter the chain: "Stock market activities emotion elicitation emotion recognition EEG-based BCI." The first one corresponds to an experiment that studies the EEG signals of just one investor influenced by their emotions and behaviors while making an investment decision [150]; the second one is unrelated to stock market activities as a source of emotion elicitation. Therefore, according to our systematic literature review, the proposed method using people carrying out stock trading has not been explored enough to provoke emotions interactively, that is, with the active participation of individuals in the experiments.

Likewise, when we enter the string: "Poker games emotion elicitation emotion recognition EEG-based BCI," seven results appear, the first is our Review [151], the third article is a review that cites us [152], and the other five results have to do with multimodal elicitation of emotions using other different methods.

Therefore, according to our systematic literature review, these two methods: participation in poker games and stock trading, have not been extensively used in other works. We chose these methods precisely because they are not previously investigated and allow for active participation of the subjects in the experiments versus studies that use a passive stimulus to elicit emotions. One of the contributions of our work is to research whether emotions can first be induced using active elicitation methods so they can later be recognized through models obtained from EEG signals.

For the data acquisition, we will use two different BCI devices with 8 and 14 channels to see the influence of diverse channel numbers in the results. To compare the classification applied in the two generated datasets, we will use the same number of registers for categorization, i.e., the corresponding 12 participants and 20 minutes of experimentation (Stock-Emotion) and 6 participants and 40 minutes for game playing (Poker). The data acquisition protocols for the two dataset acquisition have the same scheme because they consisted of the EEG data collection in two periods. First, a baseline with the participants relaxed, and second for the participants doing stock trading or poker activities, with one minute for auto-labeling emotions.

4.1 Dataset generated from stock market participants

This section describes the generation of a dataset with a group of people executing paper trading in the stock market. Paper trading is virtually simulated trading (i.e., no real money is at stake) that uses live stock market data to record paper trades (i.e., real transactions). Thus, paper trading is virtually the same as real stock market trading, but without actual money at risk. Paper trading is useful for our area of interest because it triggers emotions (like actual stock market trading). Naturally, financial decisions have consequences for individuals and society in general.

Stock market trading consistently induces feelings in participants due to its money component. Simulated trading is as emotionally charged as real trading because money has a strong emotional connotation for most humans [153]. Money represents variables such as resources, lifestyle, survival odds, value, status, health, and even the likelihood of leaving offspring, to name a few [154]. Thus, participants have strong emotions tied to money, making real and even simulated trading a powerful emotional stimulus. In particular, paper trading is also emotionally charged because it evidences the participant's ability (or lack thereof) to be profitable with real money [153].

Research suggests that emotions play a role in stock market trading because they influence decisions. Traders often cannot ignore such emotions, which sometimes distracts them from their work process when real or even simulated money is at risk [153].

Emotional states also manifest through changes in confidence and risk tolerance [155]. If the trader has a constructive mindset, performance will likely increase. Likewise, if emotions are too intense, trading performance can worsen because stock market trading requires continuous decision-making while objectively analyzing potential risks and rewards.

Emotions allow humans to respond to stimuli quickly and without rational thought. Evolutionarily, this offers a cheap way to react to preprogrammed scenarios, such as fight or flight emotions. Likewise, emotions can generate biases that influence decision-making, often unbeknownst to humans [156]. Likely, the brain's parts that create emotions evolved first and directly connected with the body, unlike the prefrontal cortex with several separated neuronal layers [157]. Thus, emotions are often harder to curtail through rational thought, but they can be managed through recognition.

Traders often report making decisions because they feel right at first but regret them shortly after. Excellent examples of emotional choices like these are “panic selling” or “fear of missing out” [158]. These emotions cause traders to sell at market bottoms or buy market tops. These trading decisions occur due to emotions, expressed as the psychological pain avoidance of losing money. However, such decisions are often inadequate after logical scrutiny. For example, buying oversold and undervalued assets probably has a positive long-term expectancy. However, doing so is typically emotionally difficult for traders due to the previously explained dynamics.

Therefore, traders need to recognize how their emotions affect their trading process. This way, traders could use emotions to their advantage and not take over their systematic methods in their trading strategy. Thus, our approach aims to understand how traders can get the best of both worlds in trading through emotions and rational thought.

We see a similar dynamic in other peak performance activities, such as playing chess or poker [159]. For instance, chess grandmasters are reportedly capable of observing a chess position and instantly “feeling” who is winning or losing, without much need for calculating further moves. This example shows the need for intuition (emotion or feeling) through experience, combined with a disciplined, rational process (i.e., calculating chess or poker moves).

In particular, stock market trading triggers three key emotions: 1) fear, 2) hope, and 3) regret [160]. These emotions directly affect the participant's trading, and these effects are detectable through EEG readings [161].

Applying the present research results may improve trading performance and profits because they are likely linked to emotional management and peak performance techniques. Finally, we believe that enhancing trading performance supports a more efficient market, which is generally understood to be positive for society. The present research results can also be applied in other competitive workplace scenarios.

The above explained the reasons why we aimed to capture critical emotions related to trading activities, such as fear, sorrow, hope, and a calm (relaxed state) [158].

4.1.1 Data acquisition for Stock – Emotion

We used a brain-computer interface (BCI) Ultracortex Mark IV EEG headset to record the experiment's participants' EEG signal. The device had 8-channel dry electrodes for recording brain signals. It uses the Cyton Bio-sensing board with an 8-channel neural interface and a 32-bit processor to collect the EEG signal. The board communicates wirelessly with a computer using a USB dongle (Figures 13.a and 13.b). The Cyton Bluetooth headband device allows for an open-source brain-computer interface that facilitates EEG data procurement and analysis.

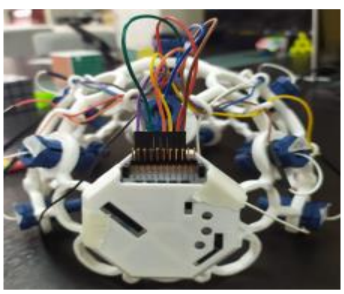
We applied the 10-20 electrode placement diagram with channels 1-8 of the OpenBCI default setting [162]. Figure 13.c and Table 10 show the systems' electrode location, and the channel's positions are presented inside a circle. Therefore, the Stock-Emotion dataset has EEG recordings using eight electrodes, 4 in each cerebral hemisphere. Table 10 shows Cyton's channels and eight electrode position

The dataset corresponds to EEG signals of 12 healthy participants between the ages of 20 to 50 years old, male and female (figure 14). Based on the trends found in the literature, it was decided to perform this experiment using 12 participants. In the databases that can be accessed publicly, there is a range from 5 to 32 participants, as seen in Table 2. Also, Figure 6 shows that 31% of the studies analyzed in the systematic review use datasets obtained with

11 to 20 participants. Moreover, one of the most used and useful open datasets is SEED, which was generated using 12 participants. Additionally, in the analysis of the results presented later in this document at the end of section 5.3, it is shown that the results obtained with 12 participants are statistically consistent.

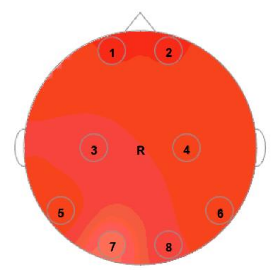
The EEG brain waves of the subjects were recorded while they traded in the United States stock markets. The experiment used paper trading, i.e., live market data and simulated money, as presented in Figure 15. The investigation required that each of the 12 participants have a session of paper trading. Participants self-reported their state of mind in one-minute intervals.

The protocol followed for the recordings started with a two-minute reference point, where the participants were asked to relax. This initial recording was labeled as a calm state. After this, the subjects initiated the trading process.



(a)

(b)



(c)

Figure 13. (a) OpenBCI Ultracortex headset, (b) Cyton Biosensing board, (c) Electrode location.

Table 10. Cyton's channels and eight electrode's position

Electrode's position	Channel
1	FP1
2	FP2
3	C3
4	C4
5	P7
6	*P8*
7	O1
8	O2

Figure 14. A participant with the Open-BCI headset.

Figure 15. Open-BCI EEG signals.

Figure 15 is composed of three sections. The data for each electrode is a time series presented on the left side. On the upper right side is data of each electrode's frequency plot. On the lower right side are shown accelerometer data that sense the head's orientation (or change in direction); in the figure, it is shown that the position of the participant's head was stable. Additionally, photo capture of the participant is presented in the right lower corner.

The participants were previously trained in a standardized trading methodology that used RSI, MACD, Keltner channel indicators, and momentum and mean reversal trading techniques.

Our goal was to give participants the same theoretical trading framework and analyze their EEG readings as they dived into the markets. We set up a carrot and stick reward dynamic for participants. As measured by their profits, the top-performing individuals were promised to receive a payout from the bottom participants. Here, we aimed to enhance the risk and reward dynamics inherent to stock trading and influence the participant's emotional states.

Participants tagged the data with a self-reported state of mind. The data was labeled using self-assessment manikin [11], presented in figure 16, to define valence – arousal in the VA space. We used the quadrants derived from the variables 1) valence and 2) arousal according to Russell's circumplex model [10].

This self-labeling method is used to generate datasets, such as DEAP and SEED. In our experiments, we have an interruption every minute to allow self-labeling using the manikin.

In this thesis, since the emotion elicitation methods are active processes in which the participant must remain focused, self-labeling can cause a loss of concentration in the activity. But this does not affect the experiment since our objective is not to achieve efficiency in the participant's activity; our purpose is to capture EEG signals and relate them to the recognition of emotions.

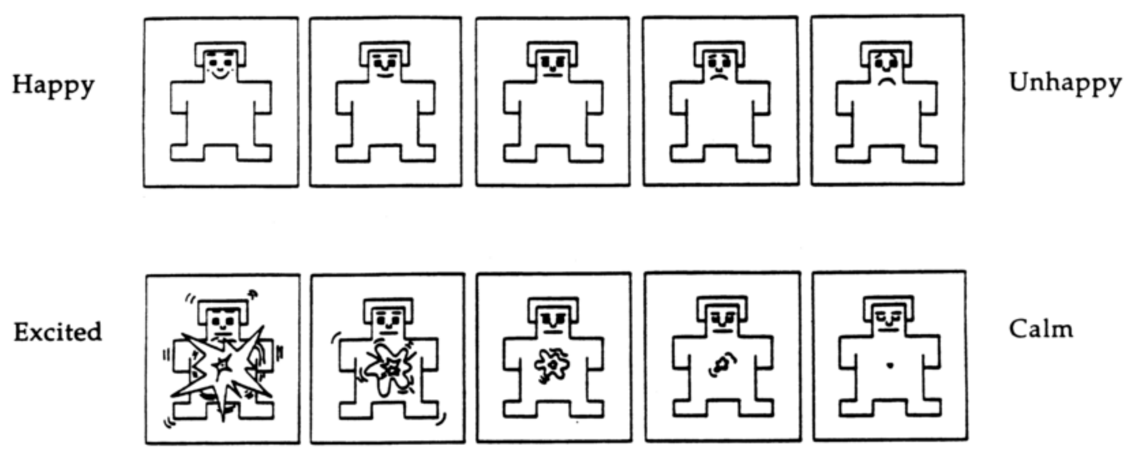


Figure 16. Self-assessment manikin for valence and arousal

As already mentioned, stock market activities trigger three key emotions: fear, hope, and regret [149]. Then we translate the EEG readings to emotional states related to these emotions. As figure 16 depicts, emotional valence describes if the emotion is positive or negative, lower values refer to a negative valence and higher values to a positive one; emotional arousal refers to the intensity of the feeling. Fear is an emotion of negative valence and high arousal LVHA; hope is a feeling of high valence and high arousal HVHA, and regret is an emotion that produces a negative valence and low arousal LVLA [10]. Figure 17 shows the Valence-Arousal plane and the three related key emotions of stock trading (fear, hope, and regret). In this figure, it is also located a relaxed state with positive valence and low arousal. A calmed condition is possibly the ideal affective state for a trader to make objective decisions.

One of our work's contributions is the delivery of the dataset called Stock-Emotion. It introduces an interactive emotion elicitation method that resembles a specific work scenario and constitutes an emotionally relevant stimulus. This method efficiently motivates affective states that a machine learning system could recognize. Also, the Stock-Emotion database could be used for general emotion recognition purposes.

As already stated, the experiments' design with 12 participants is justified by referencing similar studies analyzed in related work reviews where at least 30.1% of researchers execute experiments between 11 to 20 individuals. Regarding the duration of the sessions and the size of our dataset, it can be observed that the present work carries out 20-minute sessions with each participant, with self-labeling of emotions every one minute. This protocol is defined from the study of the systematic review of the literature since the most used databases, such as DEAP and SEED, use self-labeling every minute. On the other hand, an adult's average attention span focusing on a single task is about 20 minutes [163], which is why each session with a participant was determined to last that long.

Our dataset did not contain HVLA entries (relaxed state) except those obtained in the two-minute baseline. Therefore, the amateur participants never felt something resembling a calm state during trading. This outcome is expected for novice traders.

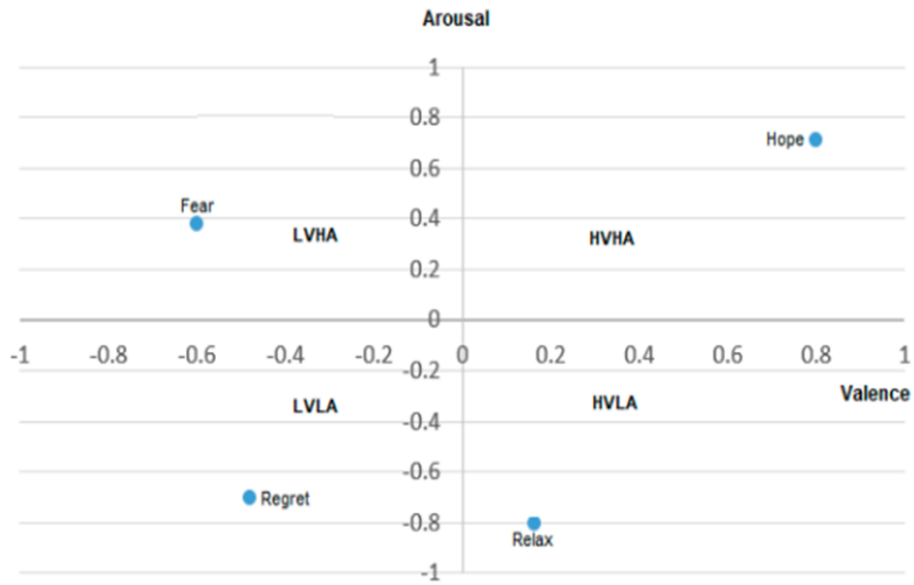


Figure 17. Russell's Circumplex Model with three key emotions and a relaxed state.

On the topic of the labeling frequency of segments of the dataset, it is worth noting that in the DEAP dataset, emotions are tagged for every one-minute long music video, with three seconds separating each clip. So, in this well-known dataset, one minute was considered enough to provoke, recognize, and tag an emotion. In our experiment, the participants labeled their emotions every one minute, deemed sufficient to detect their current feelings.

Since emotions gradually change in humans and are possible higher in the last section of the time window, processing was made for the one-minute EEG records before a label. It was also executed, focusing on the final 30 seconds of EEG data from the one-minute segment. Classification accuracy was evaluated to see if it may lead to better performance, as Koelstra et al. have suggested this strategy in their work [164].

Figure 18 shows that the Stock-emotion database does not have High Valence - Low Arousal (relaxed state) labels, except for the first two-minute baseline (not included in the figure). This finding contrast with the DEAP dataset, which has 21% in HVLA because it has another emotion elicitation method (music videos).

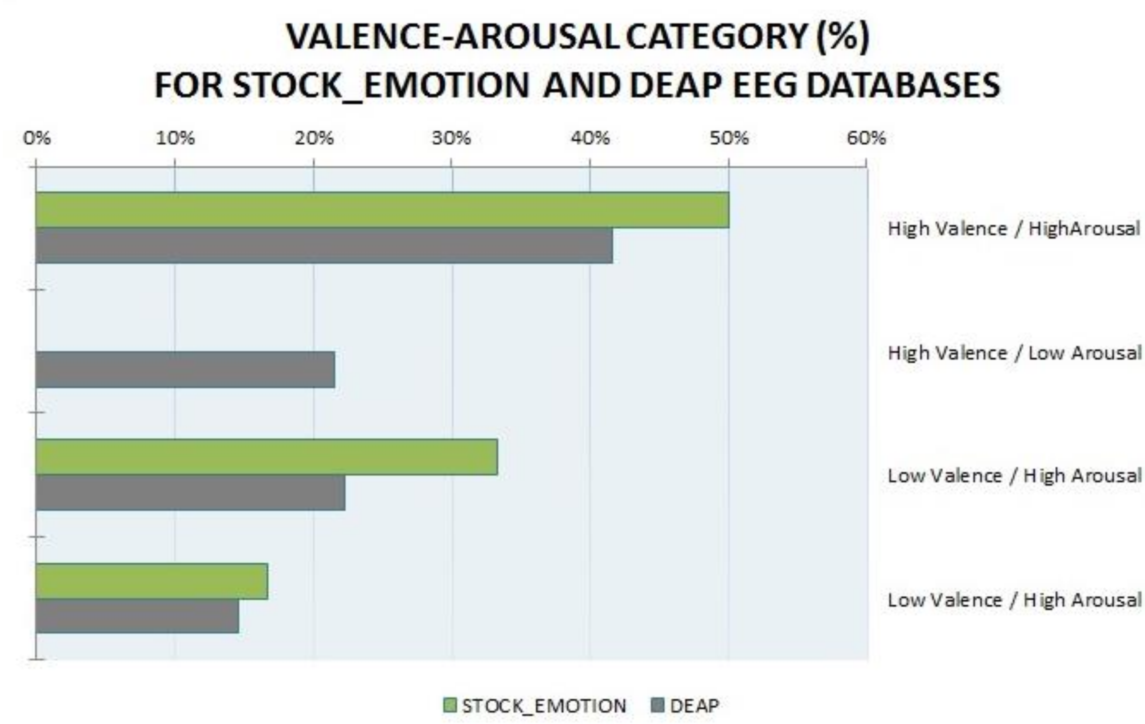


Figure 18. Valence – Arousal results Stock-Emotions vs. DEAP results

These results show that our elicitation method efficiently generates these three critical emotions related to stock trading in competitive markets. The following section presents an extension of this primary dataset to test emotion elicitation methods further using poker games.

4.1.2 Preprocessing

In addition to the preprocessing embedded in the Cyton Open-BCI device, artifact removal is carried out using two Butterworth filters with zero-phase to preserve frequencies between 1 Hz and 80 Hz to eliminate the noise generated from eye movement and heartbeat. These two filters: one high-pass followed by a low-pass, causes less distortion or loss of information than one band-pass filter. Besides, a zero-phase filter can prevent going forward and backward over the signal to eliminate a phase shift. Additionally, a 60 Hz notch filter was utilized to remove electrical noise contamination.

4.1.3 Data balancing

The dataset is unbalanced with minority classes that are ignored in the classification; this means that the number of records for each class is unequal, with a majority in some classes and fewer occurrences in others. For this reason, it is necessary to balance the number of samples to increase the sensitivity toward these categories. A type of data augmentation is used for minority classes known as Synthetic Minority Oversampling Technique or SMOTE. This technique was introduced by [165] that combines over-sampling the minority classes and under-sampling the majority classes to achieve better classifier performance. The implementation available in the SKLearn library is used.

4.2 Dataset generated from poker players

This study employs a novel emotion elicitation method to recollect EEG signals using a Brain-Computer Interface (BCI) commercial device with 14 channels for emotion recognition using machine learning algorithms. At this point in the research, we wanted to address another problem: the influence of the number of electrodes used in the BCI device to capture the EEG signals. Some studies indicate that using a smaller number of electrodes, even with a single-channel EEG, it is possible to have enough information to perform emotion recognition [166].

Therefore, we took advantage of the opportunity to test another sensor with 14 channels to see if the higher number of channels allows higher accuracy in emotion recognition compared to the OpenBCI with only eight electrodes. On the other hand, this new device, being of a newer generation than the OpenBCI, is more comfortable to use for an extended time. This last consideration is important because the average time an online poker tournament lasts around 45 minutes.

We obtained EEG signals of 6 participants playing a complete Texas hold ’em poker game consisting of several hands in a sit-and-go online tournament. As was said, a match has an average game duration of 45 minutes. The experiment with a poker game is a ludic activity that will more easily retain attention for a longer time than trading, and we want each participant to join in an entire contest to have the experience of winning or losing several games until they are eliminated or win. The time available for each experiment per participant would be about 45 minutes which is more than double that of the trading experiment; therefore, to have a similar amount of data in volume, half the number of participants is used, i.e., six subjects. This dataset is also imbalanced; we used SMOTE to oversample minority classes to obtain the same number of registers for each class. This information is used as input for an emotion recognition system that uses machine learning algorithms such as kNN, Random Forest, and neural networks.

Poker was chosen because it is a game that provokes both positive and negative emotions associated with the sudden changes of the hands. People capable of detecting and distinguishing their feelings have more possibilities to control them and make objective decisions to win over their opponents [167]. This type of emotion and the necessity of directing them to keep objective is also present in other environments that provoke intense emotions as working in the stock market [151].

4.2.1 Dataset acquisition

The experiment corresponds to the register of real-time emotions using an EEG-based BCI device to record EEG readings while the participants play a complete Texas hold ’em poker game consisting of several hands in a sit-and-go online tournament with real money at stake. The experiment joined six subjects (three men and three women) playing online Texas Hold ’em poker during a complete sit-and-go tournament with several hands.

The experimental setup has the following procedure: as a baseline, the player is asked to relax with the eyes open for 15 seconds, 5 seconds of separation, and 15 more seconds for relaxation with the eyes closed. After a hint signal, the game starts. The instants in which the player reports emotions are marked with one of the valence – arousal space quadrants. Self-assessment manikin [11] is used to evaluate valence (positive to negative feelings) and arousal (from calm to excited) to build a two-dimensional emotional space with four quadrants corresponding to positive valence and low arousal (for instance, calmness), positive valence, and high arousal (for illustration, happiness), negative valence and low arousal (for example, sadness), and negative valence and high arousal (for instance, anger).

We used a brain-computer interface (BCI) EPOCx headset with EmotivPRO software to record the EEG signal from the studied participants. Figure 19 shows that it has 14-channel wet electrodes to record brain signals. EPOCx has 14 sensors for the channels AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF4 with two references: CMS and DRL, according to the 10-20 systems diagram. Figure 20 presents a participant during one experiment with the EPOCx headset. In the upper part of Figure 21, the FFT for the F7 and AF4 channels is observed, and in the lower section, a comparison of values between these two channels for the Theta, Alpha, Low Beta, High Beta, and Gamma bands.

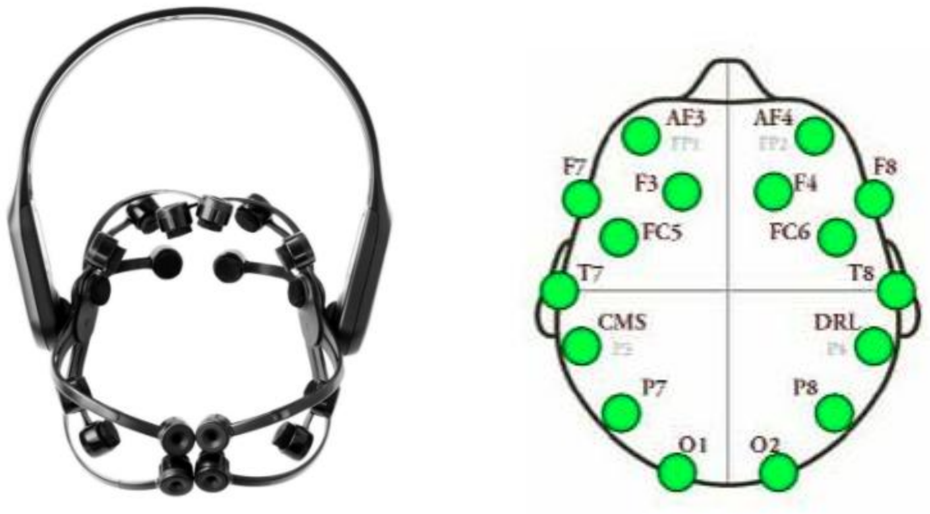


Figure 19. EPOCx headset and Electrode location.

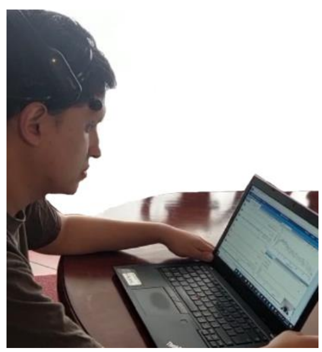


Figure 20. A participant with an EPOCx headset

Figure 21. FFT and band comparison between channels F7 and AF3 for one participant

4.2.2 Preprocessing

The generated bandwidth goes from 0.16 to 43 Hz, and the EmotivPRO software uses two notch filters to eliminate electrical noise at 50Hz and 60 Hz. Additional filtering is added with a built-in digital 5th order sync filter. Figure 19 shows the device and the position of the electrodes. The output of the EPOCx device has a complete filter system that produces a virtually clean EEG output ready for processing without additional filtering.

We obtained data transformed in CSV format from the BCI device with EmotivPro software in three files, one corresponding to raw data, the second corresponding to the labels, and a third is a JSON file with descriptive details.

4.2.3 Data balancing

The dataset for poker games is also unbalanced, with different numbers of registers for each class. We used SMOTE again to avoid a bias in the classification toward the majority classes as we did with the Stock-Emotion dataset.

5. EXPERIMENTATION AND RESULTS

This chapter presents the most relevant experiments and best results obtained using the two datasets generated for this research and different combinations of input data, features, and machine learning algorithms to infer the optimal setup for emotion recognition using EEG signals.

One of the main contributions of this work is the use of active methods for emotion elicitation to generate an EEG dataset because this field could benefit from further exploration. As far as we know, the proposed methods have not been used before, and they are one of the novelties of our research.

On the other hand, the methodology for choosing algorithms for extraction and selection of features and classification of samples was defined using the information obtained in the literature revision, focusing on the methods less commonly employed that obtained more appealing results in our preliminary experiments. Then, based on these data, additional systematic experimentation was carried out with different approaches until the best performance in terms of classification accuracy was found.

Different tests were performed using features extracted in the time and frequency domain. The best results obtained in the time domain were Higher-Order Crossing (HOC) and statistical measures as the variance. In the frequency domain, the best results corresponded to the use of five bands (alpha, beta, gamma, delta, theta), Differential Entropy (DE), Differential Asymmetry (DASM), and Rational Asymmetry (RASM).

For feature selection, a method was developed combining two criteria: computing the independence of pairs of features using the Mutual Information Matrix, a filter method. Additionally, a wrapper technique is used to measure the relevance of the features related to the classification results using Chi-Square Statistics applied in trials.

Algorithms like MLP, Naïve Bayes, SVM, and ensemble classifiers such as Random Forest were examined for classification results. In the literature, we detected that the most used algorithms were linear as Naïve Bayes and SVM, but in our experimentation, the best yields were using KNN and Random Forest, which are less used. In the same way, good results were obtained with neural networks MLP. Therefore, these algorithms were chosen to be included in the systems developed for emotion recognition.

Table 14 presents performance results using different classification algorithms with all the initial features and a selection of characteristics using 10-fold cross validation.

5.1 Feature extraction

An initial satisfactory setup included a feature vector formed by statistical characteristics in the time domain, Higher-Order Crossing feature, power bands, Differential Entropy (DE) for each frequency band, and the combination of symmetrical electrodes: Differential Asymmetry (DASM) and Rational Asymmetry (RASM).

5.1.1 Time -domain features

Statistical-based characteristics were extracted as features in the time-domain. The algorithm HOC was chosen because it is a feature extraction technique in the time domain that obtains relevant information but is not frequently used, according to the results we got in the literature review

Statistical features: In the time-domain, information about EEG signals can be extracted using statistical calculations [168]. We selected essential features that contain the most relevant signal characteristics: mean and variance because mean is the average amplitude of the signal, and variance is a measure of dispersion. These features are defined for a raw signal: (X(n), n=1, …, N), as follows:

Mean of the raw signal:

$$\mu_x = \frac{1}{N} \Sigma_{n-1}^{N} X(n) \tag{2}$$

The variance of the raw signal:

$$\sigma_x^z = \frac{1}{N} \Sigma_{n-1}^{N} (x(n) - \mu_x)^2 \tag{3}$$

HOC: Higher-Order Crossing is a feature calculated as the number of oscillations through the zero-crossing. This feature helps indicate the pattern of periodic change in the EEG signal. When a filter is applied to the time series signal, it changes its oscillation and zero-crossing count. To compute HOC, it is necessary to make an iterative process where a specific sequence of high-pass filters is applied to the original signal and to count the resulting zero-crossings, resulting in a HOC sequence. Different types of HOC sequences can be obtained using different filter designs.

HOC is estimated by counting zero-crossings with the difference operator Di of each filtered signal, as shown in equation 4, where M denotes de maximum order of the estimated HOC. HOC counts the various high-pass filtered time series changes in sign, using the difference operator D to detect these changes, i.e., the zero-crossings. The HOC features are defined by the sequence of D as follows:

$$HOC = [D_1, D_2, \ldots, D_M] \quad (4)$$

Where M is the maximum order of the filtered signal where the number of zero-crossings was calculated, we used M=5, which gives enough information without being computationally costly.

5.1.2 Frequency-domain features

Additionally, the signals were converted to the frequency domain using the Fast Fourier Transform FFT to obtain more features. Then filters were applied to obtain the respective band of frequencies. We extracted five frequency-domain features: gamma rhythm (above 30Hz), beta (13 to 30 HZ), alpha (8 to 13 Hz), theta (4 to 7 Hz), and delta (0 to 4 Hz) waves using power spectral density PSD ratio divided by total PSD.

Appropriate feature extraction is the key to constructing an efficient emotion recognition model. The features are expected to have the essential properties to discriminate among signals. We obtained features related to the electrode location and characteristics in the frequency domain.

The signals are referenced to digitally linked ears (DLE) values for EEG spatial information, calculated in the left and right earlobes (1) shown in Section 2.6 of this thesis.

V_{A1} and V_{A2} are the reference voltages on the left and right earlobe. This way, the EEG data is broken down, considering each electrode. Thus, each channel contains spatial evidence of the location of its source.

First, it is considered information obtained using FFT. Again, frequency filters in Python were used to separate the five bands. For example, Figure 22 shows five frequency bands from one of the experiment’s subjects.

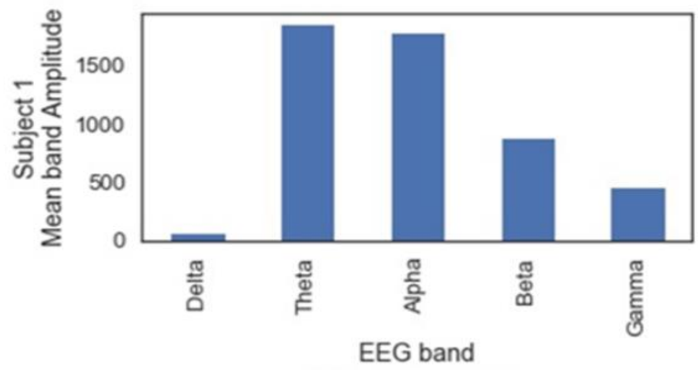


Figure 22. Frequency bands for subject 1, trial 1

EEG signals are highly complex and non-linear. Spectral entropy reflects the amount of non-linearity present in the EEG signal. It is a way to quantify, in a statistical sense, the amount of uncertainty or randomness in the pattern. This value serves to quantify the amount of information contained in the signal and its complexity. The differences in entropy generated by different emotional states are features proven to discriminate emotions [77]. For this reason, we computed as features Differential Entropy (DE) for each frequency band and the combination for symmetrical electrodes: Differential Asymmetry (DASM) and Rational Asymmetry (RASM).

In the frequency domain, Differential Entropy (DE) and its derivatives, Differential Asymmetry (DASM) and Rational Asymmetry (RASM), measure functional dissimilarities. These features are calculated from the logarithmic power spectral density for a fixed length EEG sequence and the difference and ratios between the DE features of hemispheric asymmetry electrodes [77]. These attributes are related to the frequency domain and take spatial considerations into account.

In the present work, we first transformed the signal into the frequency domain and then calculated differential entropy over each band as a relevant feature. DASM and RASM were computed to measure the DE's differences between both cerebral hemispheres.

DE is defined in (5) and (6).

$$h(X) = -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right) dx \qquad (5)$$

$$h(X) = \frac{1}{2}\log(2\pi e\sigma^2) \qquad (6)$$

Where the series X is a Gauss distribution N(μ, σ^2) where $\mu = mean, \sigma^2 = variance$ for an EEG sequence of fixed length, DE is equivalent to the logarithm of the variance in a particular frequency band [81]. DE will correspond to a feature in each of the five power bands.

DASM and RASM were calculated for each trial as the differences and ratios, respectively, between DE of 4 pairs of electrodes located in the left and right cerebral hemisphere to calculate h($X_i$left) and h($X_i$right) to obtain DASM and RASM in equations (7) and (8). Each pair of electrodes are located in the right and the left cerebral hemispheres, as shown in Table 11.

DASM = h($X_i$left) – h($X_i$right) (7)

RASM = h($X_i$left) / h($X_i$right) (8)

Figure 23 shows the distribution of the initial 20 features (Power Bands-Mean Band Amplitude (MBA), DE, DASM, and RASM, for each band) and their respective labels for a sample case. The values of the 20 features are shown as computed for each quadrant in the valence-arousal space. These values of the quadrants are the four output labels HVLA, HVHA, LVLA, and LVHA.

Table 11. Pairs of symmetrical electrodes located in each cerebral hemisphere

Pair No.	1	2	3	4
Electrode position	1,2	3,4	5,6	7,8
Channel Left	FP1	C3	P7	O1
Channel Right	FP2	C4	P8	O2

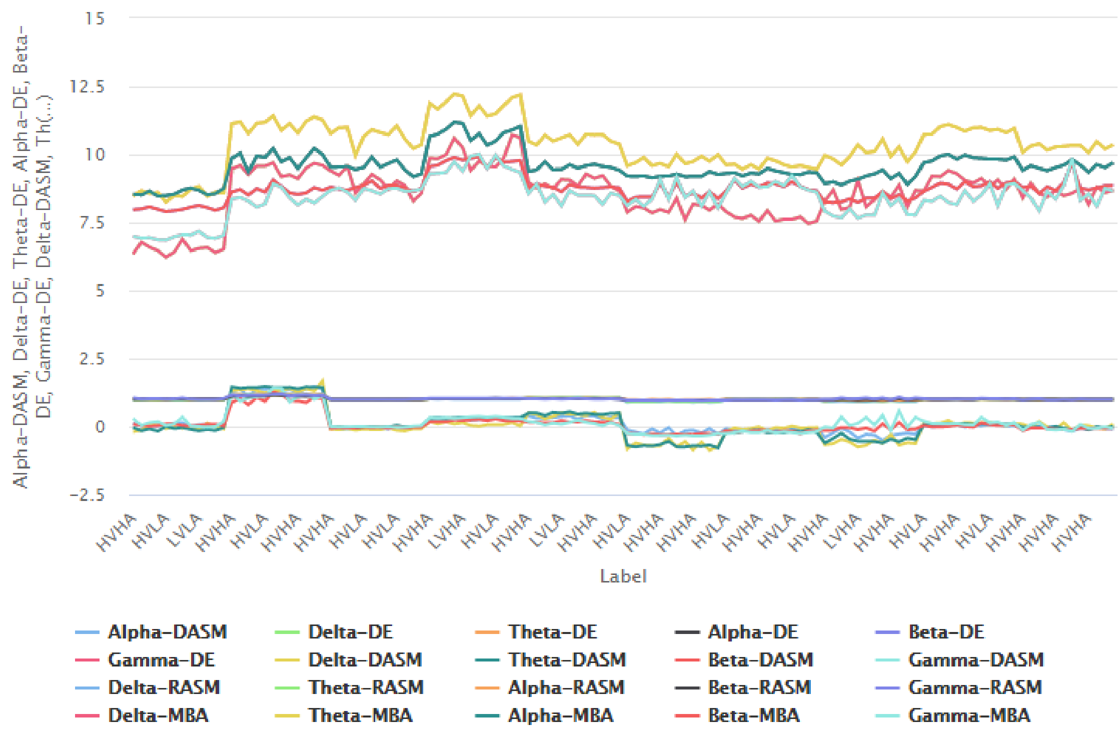


Figure 23. Initial frequency features distribution vs. labels

5.2 Classification Algorithms and Parameters

To cover a wide range of algorithm types, we used the nearest neighbor, neural network, and ensemble classifiers, namely: kNN, MLP, and Random Forest, respectively.

For classification, we selected the following algorithms: Random Forest [118], KNN [75], SVM [79], MLP [111], and a Convolutional Neural Network with one dimension (1DCNN) [113]. We chose the mentioned algorithms using three criteria: (1) because of their better performances compared with other tested, (2) since each represents different types of approaches, therefore we cover an ample range of types of machine learning algorithms: kNN is simple, non-parametric, Random Forest is an ensemble algorithm, MLP is a neuronal network, and we use it with only three layers, and 1DCNN is deep learning, and (3) because in the literature review we found that these are the algorithms that performed better for four quadrants in the valence-arousal space categorization as presented in Table 8 and summarized in Figure 12.

Additionally, the following classifiers were tested: Naive Bayes, Gradient Boosted Trees, Gaussian Process Classifiers, and AdaBoost. These algorithms generally generate lower accuracy values than kNN, ANN, Random Forest, and 1DCNN (Annex H).

We applied 10-fold cross-evaluation because it measures an accurate performance estimate [169] computed after ten iterations of different train-testing splits. This evaluation metric allows training and testing in multiple data splits to evaluate the algorithm performances in ten different training and testing data. Cross-validation is a preferred evaluation method over hold-out because this last method depends on only one train-test split. In our trials, the results from hold-out tend to be higher than those obtained with 10-fold cross-validation; despite this, we favored the cross-evaluation method for its consistency.

In the result tables of this document, we consider accuracy and not the other possible evaluation method derived from the confusion matrix because accuracy is the preferred method to evaluate classifiers' performance, as we can confirm from Table 8 with data extracted from the literature review. Also, in our results (Annex H), accuracy is an excellent metric because we apply the algorithms in a balanced dataset with almost the same occurrences for each class. That is why in the classification results, the values of precision and recall are in the same range as the obtained accuracies. Therefore, for our experiments, the analysis of the mean accuracies is good enough to evaluate the classifiers' performances. However, for further reference, the results of the other metrics are presented in Annex H for one hold-out sample.

The algorithms were implemented using Python's SciKit-Learn (SKLearn) library [170], Tensor Flow, and Keras [171]. We obtained the parameters for each algorithm experimentally, maximizing the best performance. Some of these experiments are shown in Annex H.

All algorithms have the features or raw data as input vectors and classify the label emotion corresponding to the respective quadrant in the valence - arousal space.

The accuracy, Recall, and F1 are computed from the confusion matrix as indicated by (9), (10), and (11), where TP = true positive, FP = false positive, TN = true negative, and FN = false negative, as shown in equation 5.

$$Accuracy = \frac{TP}{TP+FP+TN+FN} \quad (9)$$

$$Recall = \frac{TP}{TP+FN} \quad (10)$$

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (11)$$

For our experiments, in the confusion matrix, the values 0,1,2, and 3 correspond to each quadrant of the emotion representation in the space valence-arousal (Annex H). The accuracy value in the result tables is computed as the average of the 10-fold cross validation accuracies. Detailed results with accuracies for each fold are also displayed in Annex H.

Next, we will describe the algorithms' parameters that will be used to classify the features obtained in the two datasets: Stock-emotions and Poker. They will also be used to test a subset of the DEAP dataset.

5.2.1 kNN parameters

For kNN, we tried for 3, 5, and 7 nearest neighbors (Annex H). For instance, for one of the datasets (Poker), we obtained mean 10-fold cross-validation accuracies of 85.15%, 82.06%, and 77.71% for 3, 5, and 7 nearest neighbors, respectively. Therefore, the best performances are achieved for 3 neighbors.

The chosen parameters for this classifier are the following:

Three neighbors: three nearest points are considered.

Uniform weights: all neighbors are weighted equally.

Euclidean distance is used for the computation of the nearest neighbors.

5.2.2 MLP parameters

The network has the feature vector as input and four outputs corresponding to each quadrant in the valence-arousal space.

To define the parameters, we tried with a different number of nodes in three hidden layers. For example, in the Poker dataset with hidden layer sizes of (150,100,50) and (100,50,10), we obtained mean 10-fold cross-validation accuracies of 80.83% and 72.50%, respectively. Also, we tried other activation functions for the chosen configuration of hidden layers and number of nodes: Hidden layer sizes = (150,100,50) with ReLU and Tanh activation functions, and we got mean 10-fold cross-validation accuracies of 96.55% and 73.21%, respectively. (Annex H).

After these trials, the preferred configuration is the following:

The classifier has three hidden layers with 100, 150, and 50 nodes, respectively.

The activation function for the hidden layers is a rectified linear unit (ReLU).

The solver is Adam which refers to a stochastic gradient-based optimizer because it works well and converges faster for large datasets like our EEG database.[172].

The maximum number of iterations=500: The solver iterates until convergence or this number of iterations. We found that fewer iterations do not always converge. Experimenting with the limits of convergence, we establish this number of iterations.

5.2.3 Random Forest parameters

This algorithm uses a set of decision trees. Each tree performs a classification, and the class with the highest number of votes is included in the model prediction. This model works very well based on the logic that multiple classifiers that function as a committee work better than a single classifier. On the other hand, even though many trees may have the wrong classification, others are included in the correct categorization. The final decision will be accurate if the trees have a low correlation.

To define the parameters of this classifier, we tried 10, 100, and 300 trees in the forest (n_estimators), and we obtained mean 10-fold cross-validation accuracies of 73.57%, 77.89%, and 62.14%, respectively (Annex H).

After these trials, the chosen parameters are:

Number of trees in the forest (n_estimators)=100

The criterion for estimating a split's quality is Gini impurity and entropy for the information gain [173].

Maximum depth of the tree=none; the nodes are expanded until all leaves are pure or contain less than two split samples.

5.2.4 CNN parameters

This algorithm was implemented as a neural convolutional network of one dimension (1DCNN). The proposed model is shown in Figure 24.

L1 and L4 are convolutional layers with 16 and 32 filters.

L2 and L5 are Batch Normalization layers to help the convergence of the model and avoid overfitting.

L3 and L6 are average pool layers that reduce the dimensionality of the data using a 2x2 window.

L7 is the flatten layer to connect to the classifier transforming the output of the last pool layer into a one-dimension input vector.

L8, L10, and L12 through L12 are dense layers of the neuronal classifier with activation functions tanh, tanh, and softmax, respectively. Layer 12 has four outputs representing the labels for each quadrant in the valence – arousal space.

L9 and L11 are dropout layers with 50% neurons randomly switched off to prevent overfitting.

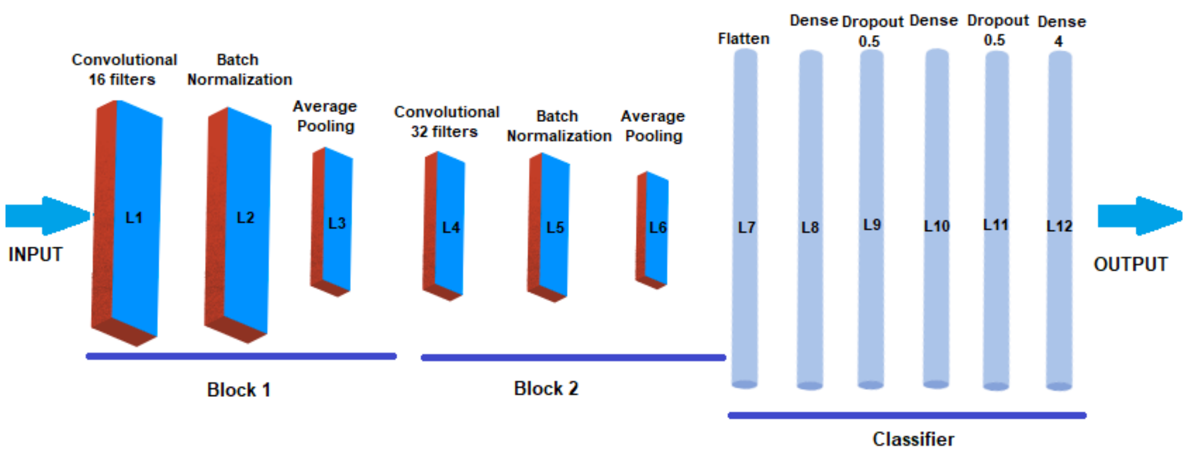


Figure 24. 1DCNN network configuration

5.3 Classification without feature selection applied in the Stock-Emotion dataset

The feature vector was composed of the five frequency bands, DASM and RASM values, statistical measures, and HOC series applied in the stock market-related dataset. The classifier inputs were processed using feature scaling to normalize them. This process helped the algorithms to converge more efficiently. The specified algorithms are applied, i.e., kNN, Random Forest, MLP, and 1DCNN.

Table 12 shows the accuracy results for 10-fold cross validation with the Stock-Emotion dataset and the selected algorithms.

Table 12. Accuracy results

Dataset	Classifier	Mean Accuracy (%)
Stock-Emotion	kNN	83.81
	MLP	86.81
	Random Forest	85.32
	1DCNN	84.65
DEAP	kNN	51.24

Using the detailed parameters, the MLP algorithm had the best classification accuracy of 86.81%. kNN had an accuracy of 83.81%, Random Forest 85.32%, and 1DCNN results are 84.65%. All these results are better than the average value of 76.68% obtained from state-of-the-art for four quadrant classification in the valence –arousal space (Figure 12).

We also tested the KNN algorithm in a subset of the DEAP dataset considering the corresponding eight electrodes and obtained an accuracy of 51.24%. The same features were obtained in DEAP in the corresponding eight electrodes, and we fed them into a kNN classifier. The objective is to test the possibility of recognizing emotions using our generated datasets vs. a subset with eight channels from the DEAP public dataset. Table 12 presents the results. As shown, our model developed on our Stock-Emotion dataset has higher accuracy than executed in the DEAP dataset. Nevertheless, the model's metrics on the DEAP dataset with our approach give a lower performance than the state-of-the-art and lower than other studies using complete DEAP data. This lower performance may be produced due to using only eight channels instead of the original 32 used for DEAP.

The participants for the generation of the Stock-Emotion dataset mostly tagged three classes: High Valence – High Arousal, Low Valence – High Arousal, Low Valence – Low Arousal. Each class is related to one fundamental emotion associated with stock trading: hope, fear, and regret, respectively. High Valence – Low Arousal was never labeled for the participants during trading because they were all beginners in stock trading, and neither was in a focused or relaxed emotional state. However, this dataset also has High Valence – Low Arousal samples, but they correspond to the baseline where participants were asked to relax with open and closed eyes. Due to the differences in occurrences for the different classes, the datasets were first balanced using SMOTE before the classification process.

5.4 Classification without feature selection applied in the Poker-game dataset

Using the 14-channel EPOCx headset and amateur poker games participants, we obtained data transformed into CSV format. With this information, we linked the raw signals to the labels, transformed them into the frequency domain, and then applied filters to obtain the respective frequency band. We extracted the same five frequency bands and the HOC values and statistics in the time domain, DE, RASM, and DASM features, as we did before, to facilitate the comparison of results. Using two devices, Open-BCI for Stock-Emotion and EPOCx for Poker games, may allow us to compare results linked to a minimum change in the number of electrodes. Keep in mind that Open-BCI has eight electrodes, and EPOCx has 14 channels.

The same four classifiers are used in this dataset: a non-parametric KNN, a neural network classifier MLP, a deep learning classifier 1DCNN, and an ensemble classifier Random Forest, according to the criteria explained in section 5.2. The results of the initial trials with SVM, Gaussian Naïve Bayes, Decision Tree, and Ada Boost Classifier are shown in Annex H.

The parameters for each algorithm are the same as described in Section 5.2. The classifiers' parameters not described are set to default values. Results for each classifier are shown in Table 13 for mean accuracy for 10-fold cross validation.

Table 13. Accuracy results

Dataset	Classifier	Accuracy (%)
Poker game	kNN	85.15
	MLP	77.89
	Random Forest	96.55
	1DCNN	86.78
DEAP	kNN	58.22

Table 13 shows that Random Forest has better accuracy (96.55%), followed by 1DCNN (86.78%), MLP (77.89%), and kNN (85.15%). All the algorithms have high accuracy and work satisfactorily for a four-class quadrant classification compared with state-of-the-art performances that show an accuracy of 76.68% (Figure 12).

Again, we replicate this setup in the DEAP dataset, using the data for the corresponding 14 channels, consuming the same kNN classifier. The accuracy results in our dataset were higher than those obtained using the DEAP database. Again, the model's metrics on the DEAP dataset with our approach give a lower performance than the state-of-the-art and lower than other studies using complete DEAP data. This lower performance may be produced due to using only eight channels instead of the original 32 used for DEAP.

5.5 Feature selection and classification results for Stock-emotion and Poker datasets

This section discusses the methods for selecting the most relevant features to enter into different algorithms to build an optimal emotion recognition model. This section will use the features described in Section 5.1 as the original input vector, and then we will apply feature selection methods to test new models using the two datasets developed in our experiments: Stock-Emotion and Poker.

5.5.1 Feature Selection

As mentioned, features with information from the EEG signal were extracted, taking into account the frequency domain and spatial information related to the electrodes' symmetrical location. Power bands, DE, DASM, and RASM, were calculated for each of the five frequency bands of the EEG signal: Delta (0-4Hz), Theta (4-7Hz), Alpha (8-12Hz), Beta (12-30Hz), and Gamma (30-80Hz). Consequently, this yielded a vector of characteristics made up of 20 attributes.

Feature selection may allow us to obtain an optimized feature vector. We aimed to get a classification with less possibility of overfitting, reducing the number of features. We used feature selection methods to feed the classifiers and compare performance: filter, wrapper, and embedded methods. The mutual information matrix is used as a filter algorithm that evaluates each pair of features' correlation; the goal is to have information to discard redundant features. Chi-square statistics, a wrapper method, is used after trial classification to help eliminate features that are not significant since they do not influence the classification; here, the objective is to discard irrelevant characteristics. Then, we combined the results of filter and wrapper methods to select no redundant and relevant features.

Therefore, we test if the features are statistically independent in the first step. The mutual information between two random variables x and y is calculated and defined using (12).

$$I(x;y) = \iint p(x,y) \log \frac{p(x,y)}{p(x)p(y)} dxdy \qquad (12)$$

Where p(x) and p(y) are the marginal probability density functions of x and y, respectively, and p(x, y) is their joint probability function. If I(x, y) equals zero, the two random variables x and y are statistically independent [98].

Figure 25 presents each attribute's mutual information matrix values in the Stock-Emotion dataset. MBA represents Mean Band Amplitude. This matrix can detect a correlation between the different pairs of features. In this figure, red cells show more correlation between features present in rows and columns, and blue cells show no correlation. In the middle of these extremes are intermediate values. If two features are not statistically independent, i.e., they have a high correlation, eliminating one should be considered.

It is also essential to consider the chi-squared statistics results, which evaluate each feature's significance related to the output classes. In our proposed feature selection method, Mutual Information Matrix and Chi-square statistics are algorithms that complement each other strengths.

The Chi-square method is a statistical approach to evaluate the dependency between two variables and differences in distributions. This algorithm evaluates features individually to determine their influence over the output classes, so it is a method that is applied after classification results.

The features were chosen considering those not easily accessed because they are not already developed in known libraries. Additionally, the features were selected if we found they provide sufficient information for the classification; this can be deducted using the chi-square value.

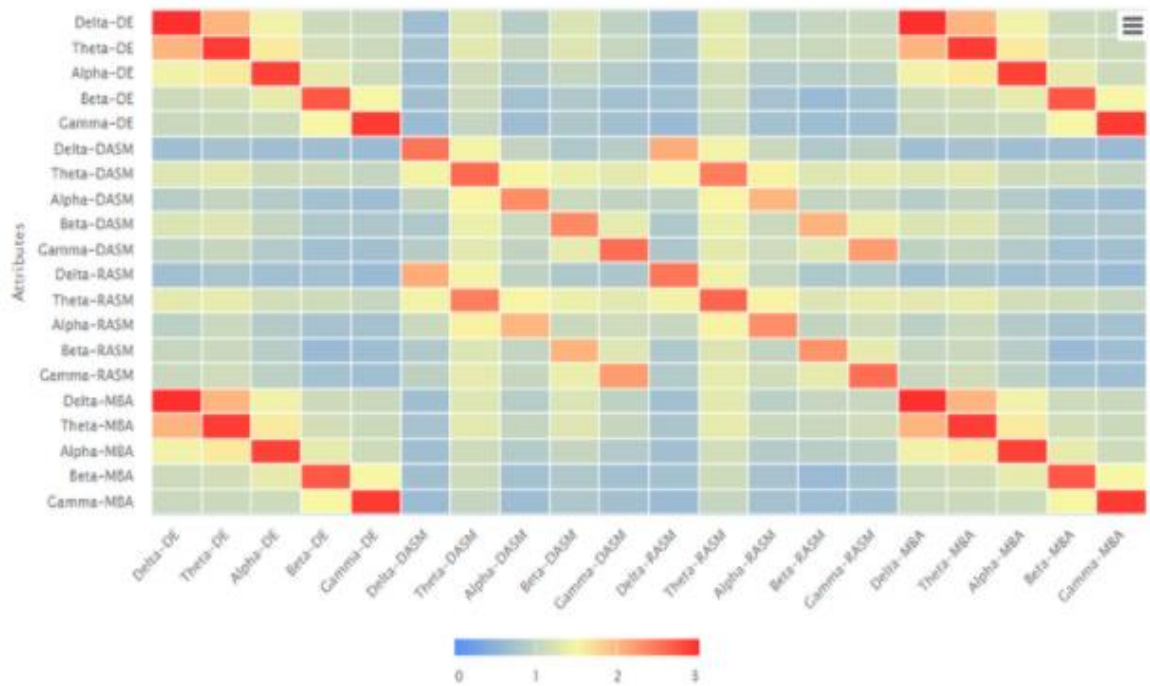


Figure 25. Mutual Information Matrix

The algorithm evaluates each feature's significance concerning the output classification. The higher the Chi-square value, the more the classifier response depends on the feature, and the attribute should be selected. If a feature is not significant for the response, it may be discarded from the model [174].

The chi-square statistic is defined in (13).

$$\chi_c^2 = \sum \frac{(Oi - Ei)^2}{Ei} \tag{13}$$

Where: i = number of intervals, c = degrees of freedom, O = observed value(s), E = expected value(s). The larger the chi-square value, the more significant the feature is, i.e., it has more influence on the emotion classification.

Chi-square statistics show the weight of the features in the classification. Less significant features appear to be the indexes RASM and DASM for the gamma and delta bands, respectively (Figure 26). Delta and its differential entropy have the most weight in the classification, followed by the Alpha and Gamma bands and their differential entropy. These findings align with assertions made in other works. In [175], the authors stated that the delta band increased the synchronized activity between the two cerebral hemispheres, particularly in the emotions with negative valence; this feature can be associated with the power band's amplitude and the differential entropy. In [164], a negative correlation was reported in the Alpha and Gamma bands for arousal.

Figure 27 shows the remaining features after the selection made considering Mutual Information Matrix and Chi-square statistics. According to the matrix in Figure 25, there is a strong correlation between the power bands and their differential entropy. Thus, power bands were omitted from the remaining features. The values of the seven remaining features are shown as computed for each quadrant in the valence-arousal space. These values of the quadrants are the four output labels HVLA, HVHA, LVLA, and LVHA.

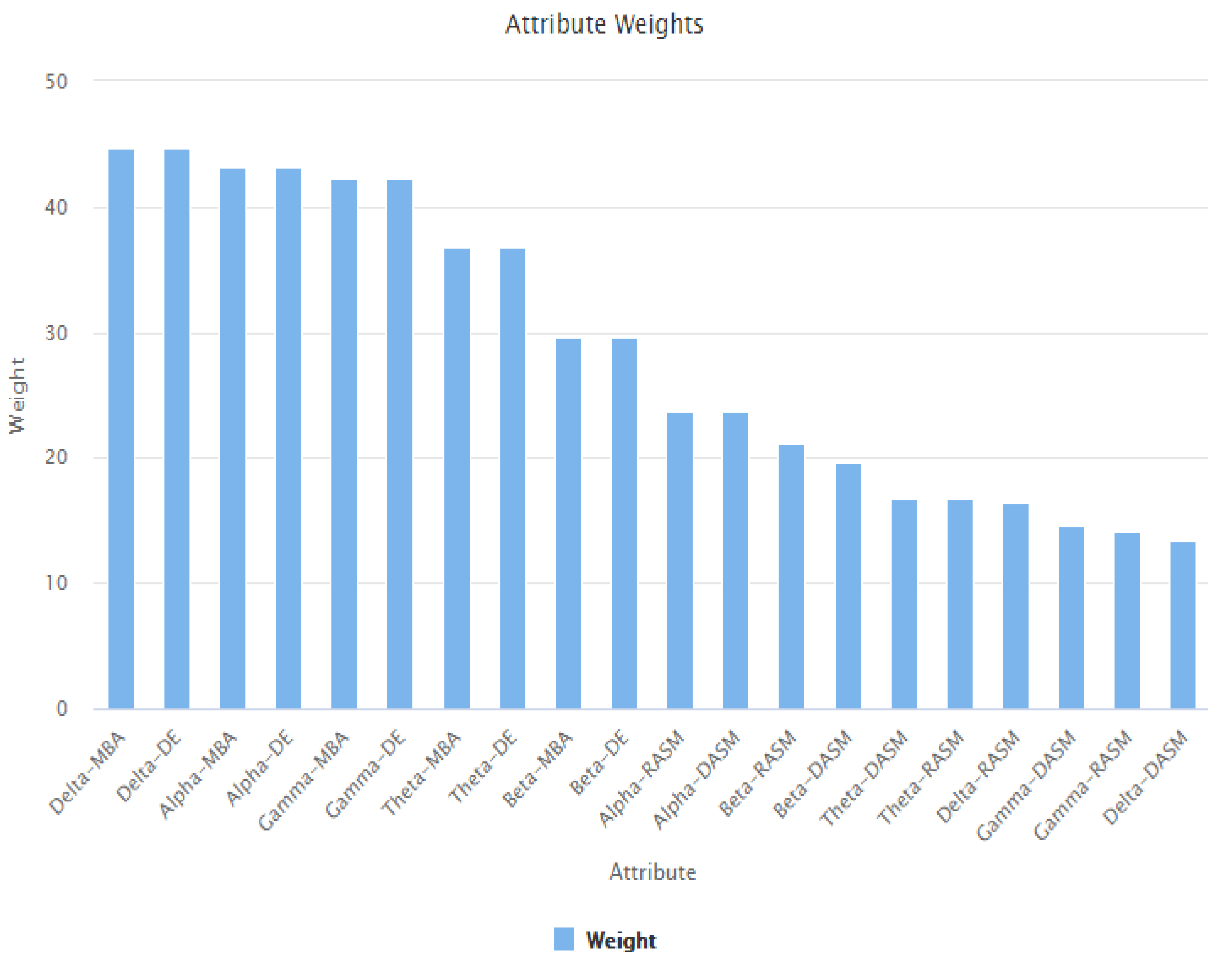


Figure 26. Attribute weight by chi-square statistics**.**

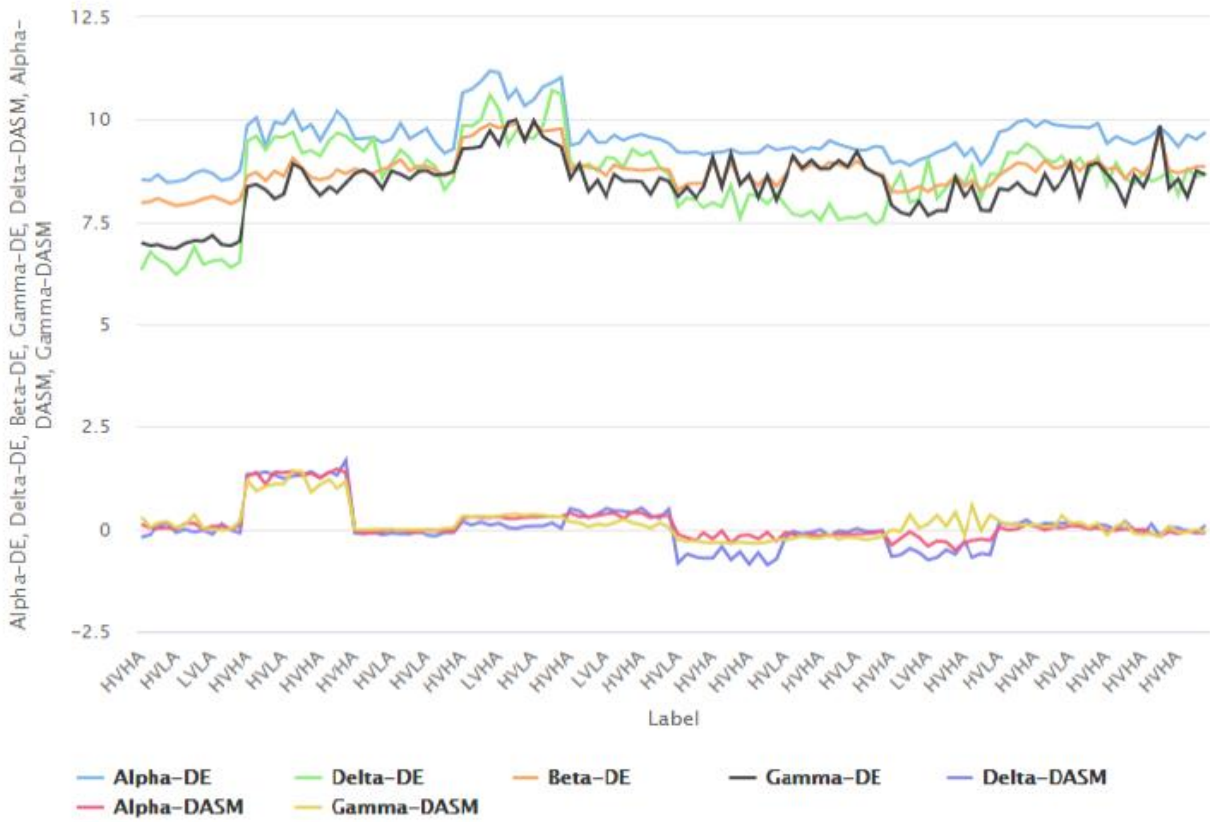


Figure 27. Remaining features after feature selection vs. labels**.**

5.5.2 Classification with feature selection

A classification algorithm is used to infer the emotional states of the experiment's participants. The system uses two phases: 1) training to generate models and 2) testing to evaluate them in new data. The classes that the algorithms recognize correspond to the four quadrants of the Valence-Arousal space, related to the critical emotions associated with stock trading: fear (LVHA), hope (HVHA), sorrow, or regret (LVLA), and a calm state (HVLA).

The system trains the Stock-Emotion, Poker game, and DEAP databases. Training and testing are carried out using 10-fold cross-validation.

Necessary adaptations are made to test the models in DEAP. Firstly, DEAP has 32 EEG channels configured with the 10-20 system. From DEAP, only eight electrode data are considered to correspond to the eight channels configuration of the Cyton BCI device used to obtain Stock-Emotion data and the 14 channels for the EPOCx device. This correspondence is possible because all three devices use the same 10-20 system for electrode location.

Furthermore, from the attributes from DEAP, only valence and arousal labels are utilized because our approach does not consider dominance or liking. Finally, physiological signals were not included.

For classification, again, we use Random Forest, KNN, SVM, MLP, and 1DCNN. The system performance was evaluated using 10-fold cross-validation using the tests carried out in Stock-Emotion and DEAP.

This thesis used Random Forest, KNN, MLP, and 1DCNN algorithms to classify emotions using previously described feature vectors. The purpose is to compare results from different types of classifiers using an original set of features and a subset of these characteristics.

The classifiers were trained and tested using the described original feature described in section 5.5.1: five frequency bands, DE for every band, DASM y RASM for eight pairs of channels in each frequency band, for a total of 20 initial features. The process also used a feature selection procedure that combines filter and wrapper algorithms to obtain a feature vector with only seven chosen features. Filter methods for feature selection evaluate the feature relevance by their correlation, so the characteristics with redundant information can be eliminated. Wrapper methods measure the influence of the features in the classification. kNN and 1DCNN classifier, were also used to evaluate results [99] when we fed them with raw data.

The system was trained and tested for comparison using all the initial features, processing the selected attributes, and using raw data. The selection process was successful if the classification performance was maintained or improved using the selected features versus the original ones. Then, it would be possible to state that only the redundant information was eliminated while retaining the features relevant to emotion recognition.

The parameters of the algorithms are described in section 5.2.

In summary, classification training and testing were done with raw data, original features, and selected features. The results with the original features (without feature selection) were shown in sections 5.1 and 5.2 obtained in the three databases: Stock-Emotion, Poker datasets, and a subset with eight and 14 channels of the DEAP dataset. In this section, we again examined the results of the application of the algorithms in the three datasets, adding classification with feature selection and processing with raw data.

Stock-Emotion, Poker game, and DEAP are unbalanced datasets with more labels in two classes than in the other two, introducing a bias in the classification. For these reasons, for all the experiments, we balanced the number of samples in each class using SMOTE and data augmentation in the minority classes [165]. Again, to avoid a lack of significance in our results, we applied 10-fold cross-validation.

Table 14 shows mean accuracies for the experiments from the previous section that used the original feature vector without feature selection. We add classification with an input vector with feature selection. Table 14 also presents accuracy values for classification made with raw data. The mean accuracies obtained in our work with our datasets Stock-Emotion and Poker are comparable and even better than the State-of-the-Art's best results, corresponding to four classes in the quadrants of the valence-arousal space (76.68%) (Figure 12).

It is worth mentioning that authors in [176] experimentally determined that extracting information from eight channels is needed to obtain enough performance in an emotion recognition system for practical use. In the present work, that hypothesis is confirmed between a range of electrode variation because, without feature selection, the results are in the same value range using Stock-Emotion and Poker datasets with eighth and 14 channels, respectively, i.e., a six electrode variation. Nevertheless, the performance with the DEAP dataset is diminished, so maybe the change from 32 channels to eight and 14 influences the results because is a 24 and 18 electrode decrease.

The goal of applying the algorithms in a subset of the DEAP dataset was to test how good the three datasets, Stock-Emotion, Poker, and DEAP, are at using emotion recognition algorithms. The results with Stock-Emotion and Poker were consistently better in the classification results. But again, these results may be due to the DEAP´s decreased number of electrodes.

On the other hand, Table 14 demonstrates that feature selection maintains or improves accuracy in most of the experiments. In the Stock-Emotion dataset for kNN, Random Forest, and MLP, the mean accuracy with the original feature vector is 83.81%, 86.81%, and 85.32%, respectively. It improves using the selected features with the following values: 84.78%, 90.60%, and 91.30% for kNN, Random Forest, and MLP. In the Poker dataset for Random Forest, MLP, and 1DCNN with the original features, we have mean accuracies of 77.89%, 80.83%, and 86.78%, respectively. With input vector with selected characteristics, we have higher mean accuracies of 78.28%, 83.27%, and 86.79%, respectively. Mean accuracies for input vectors with selected features decrease sightly only using 1DCNN for Stock-Emotions and kNN for Poker dataset with 84.65% versus 84.36% (minimal difference) and 85.15% versus 82.06% for original features versus features chosen, respectively. One possible reason for this improvement is that in the proposed feature selection process, we eliminate the redundant and the irrelevant features, i.e., the pairwise features correlated between them (redundant) and the features that are not influential concerning the target classes (irrelevant). In most cases, we believe that removing redundant and irrelevant features reduces noise and increases the contrast among classes. It is worth noting that all classification results using our original and selected features are better than the average value of 76.68% of the state-of-art (Figure 12).

The results using raw data applying kNN and Random Forest in the Stock-Emotion dataset are 30.85% and 51.10%, respectively. Mean accuracies using raw data in the Poker dataset with kNN and Random Forest are 38.36% and 51.62%, respectively. Mean accuracies for algorithm kNN for 8-channels DEAP and 14-channels DEAP are 47.81% and 45.01%, respectively. All of these numbers correspond to low accuracies and demonstrate the value of extracting features from the raw data because it improves the performance of the classifiers. Our justification for this is the same as above: the contrast among classes improves when removing data that does not contribute to the classification's pattern recognition.

Additionally, it is interesting to notice that using neural networks with embedded feature extraction processes produces better mean accuracies obtained from raw data. MLP presents a mean accuracy of 64.30% for the Stock-Emotion dataset; 1DCNN has an accuracy of 67.17%%. MLP gives a mean accuracy of 64.11% for the Poker dataset, and the 1DCNN result is 68.87%. For the DEAP dataset subset of eight channels, 1DCNN has a mean accuracy of 59.90%. For the DEAP dataset subset of 14 electrodes, 1DCNN presents an accuracy of 60.10%. These values are higher than those obtained in raw data with kNN and Random Forest that do not have an embedded feature extraction process. Neural network algorithms work to assign weights to each input, while Random Forest and KNN do not have this process. This last consideration explains the better results with raw data with MLP and 1DCNN.

Also, feature extraction reduces the program´s execution times by several days, as seen in Table 15, which presents data for execution time for kNN and 1DCNN for Stock-Emotions and

Poker datasets with original features, selected features, and raw data. The execution time is critical, considering that most emotion recognition applications involve real-time programs.

Results validated the proposed feature extraction process that gives better results than raw data classification. Also, the results validate our proposed feature selection method because combining Mutual Information Matrix and Chi-square statistics for feature selection maintains or improves performances. The algorithms respond better and better accuracies are obtained when they are fed with relevant information, not redundant, using features that extract pertinent information from the raw data. We justify these results with the reasoning that removing redundant and irrelevant data reduces noise and increases the contrast between classes. With our proposed method, the chosen features have two characteristics: they do not correlate among them, i.e., they are not redundant, and they directly affect the classification results. We achieve these types of informative features using proper feature extraction and later a combination of feature selection techniques: the mutual information matrix, where it is shown that the selected features do not correlate with each other, and their chi-squared values confirm that results depend on them; meaning that the selected features are not redundant and relevant, respectively.

Table 14. Classifiers Performances

Dataset	Classifier	Accuracy with initial feature selection (%)	Accuracy with – raw data	Accuracy with feature selection (%)
Stock-Emotion	KNN	83.81	30.85	84.78
	Random Forest	86.81	51.10	90.60
	MLP	85.32	64.30	91.30
	1DCNN	84.65	67.17	84.36
Poker game	KNN	85.15	38.36	82.06
	Random Forest	77.89	51.62	78.28
	MLP	80.83	64.11	83.27
	1DCNN	86.78	68.87	86.79
DEAP (8 channels for Stock Emotion comparison)	KNN	51.24	47.81	50.55
	1DCNN	67.40	59.90	65.07
DEAP (14 channels for POKER comparison)	KNN	58.22	45.01	50.73
	1DCNN	64.86	60.10	64.10

Table 15. Programs execution times for KNN in Stock-Emotion and Poker Datasets

	KNN execution time (secs.)	
	Stock-Emotion Dataset	Poker Dataset
Original features	8.75	2.88
With feature selection	8.83	2.3
Raw data	342647	363413

With these outcomes, we compare the performance of the systems. Making a compilation of what has been done, we can say that our systems use two different datasets with a different number of channels: 8 for the Stock-Emotion and 14 for Poker. Stock-Emotion and Poker datasets were generated through two active approaches to emotion elicitation: trading in a competitive stock market and online poker games following a similar protocol for data acquisition; two initial minutes of relaxation for baseline and emotion auto-labeling each minute. The datasets were preprocessed to eliminate artifacts and to balance the classes. The emotions provoked by stock trading activities and online poker games were tagged using the self-assessment manikin method. The user-made labels were translated to four discrete states in each of the four quadrants of the Valence-Arousal space: HVHA, HVLA, LVLA, LVHA, to further differentiate emotional states in our classification models. We experimented with classification with the original features and also proposed a system that included a selection process.

Feature extraction is essential when manipulating EEG signals involving large amounts of data. We combined time, frequency, and spatial information to define the calculated features. Time attributes were extracted using statistical methods. The mean amplitudes of five frequency bands (Delta, Theta, Alpha, Beta, and Gamma) were calculated in the frequency domain. Additionally, Differential Entropy (DE) and its derivatives Differential Asymmetry (DASM) and Rational Asymmetry (RASM) were obtained to measure functional dissimilarities between the cerebral hemispheres (location information). The original vector had 20 features, with interesting features in time, frequency, and spatial domains. The classification results with these original features are better than the state-of-the-art accuracy values, demonstrating that these 20 characteristics extract essential information to recognize emotions in the valence-arousal space's quadrants.

However, we aimed to optimize this input feature vector and tried to reduce the number of input characteristics using a proposed combination of methods for feature selection.

After feature selection, seven characteristics were chosen using Mutual Information Matrix and Chi-squared statistics: the differential entropies for Delta, Alpha, Gamma y Beta bands, and the DASM for Delta, Alpha, and Gamma bands were the most influential, i.e., not redundant and relevant, in the four-classes emotion classification.

The findings concerning each feature's relevance in the classification align with assertions made by other papers corresponding to the relationship among information associated with frequency bands and emotions. For example, our dataset does not have calm state-related samples (except for baseline), and interesting enough, the Theta band and its related features were eliminated in the selection process. This finding supports what was already mentioned in [14], which linked this band with the presence of mindful states.

Another chosen feature was DASM obtained in the alpha band associated with asymmetric activity at the frontal site, which is connected with emotion. This result agrees with [12]. On the other hand, RASM features were eliminated because they had redundant information already present in DASM.

The results show that combining a filter (Mutual Information Matrix) with a wrapper method (Chi-square statistics) for feature selection allows us to get a valid feature selection process. According to the results, it is possible to state that the selected features carry the relevant information for classifying the emotions in one of the four quadrants of the valence-arousal space.

Using three criteria explained in section 5.2, we chose the following classifiers: kNN, Random Forest, MLP, and 1DCNN. Outcomes with raw data, original features, and selected features were compared. Feature extraction greatly enhanced the classifiers' performances because raw data have many irrelevant patterns that affects emotion recognition. Then, feature selection also helped, in most cases, to improve the classifiers' performance.

The MLP and 1DCNN algorithms are neural networks that have embedded feature extraction and selection methods; therefore, they work better in raw data than the classifiers that are not neuronal networks. All the classifiers perform worse with raw data than with features probably because too much data corresponding not to one minute but to every second with 128 samples each can be bad for classification. This amount of data for each label (128 x

60 rows of data instead of one row for a label) can cause overfitting, lower performance, and produce an inefficient model. This effect can be justified by reasoning that any regions of the data can be presented with patterns that correspond to few or no observations, making the classifier lose accuracy in the prediction for overfitting to conform to more frequent patterns in the data.

The results show that combining a filter (Mutual Information Matrix) with a wrapper method (Chi-square statistics) for feature selection allows us to get a valid feature selection process. According to the results, it is possible to state that the selected features carry the relevant information for classifying the emotions in one of the four quadrants of the valence-arousal space.

5.5.3 Statistical analysis of the significance of the results

Using the EEG data generated for 12 participants during at least 20 minutes of experimentation gives satisfactory classification results but is also justified with statistical analysis of the results as presented below. In this section, we statistically demonstrate that 12 participants producing 20 minutes of EEG signals for eight channels and 128 sampling frequencies (trading experiments) are enough data to create a dataset of sufficient size to be statistically significative. Therefore, this demonstration also applies to a dataset of a similar size for 6 participants in experiments that collect around 45 minutes of EEG signals for 14 channels and 128 sampling frequencies (poker games experiments).

Each observation is modeled as X ~ Ber(p). It means a Bernoulli variable with an unknown parameter (accuracy) "p." We are modeling a prediction in which each input record was classified correctly (=1) or incorrectly (=0). Observation is not the input vector in this case but whether the prediction with that input vector was correct or incorrect. X is the variable with which we are modeling the accuracy of this algorithm that follows a Bernoulli distribution.

For instance, we took 149619 samples of $X \in \{0,1\}$ and $\sum_{i=1}^{n} Xi$ = 105341, where n=149619. We also assumed the samples were independent and identically distributed, given that they come from the same subject.

Each Xi takes 1 when correctly classified and 0 when it is incorrect.

We can therefore calculate our accuracy estimator as follows in (14):

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} Xi = 0.7041 \tag{14}$$

According to the Law of Large Numbers (LLN) applied in (15):

$$\bar{X}_n \xrightarrow[n \to \infty]{P, as.} p \tag{15}$$

where p is the unknown accuracy parameter that we estimate with our sample. Also, “P” and “a.s” denote convergence in probability and almost surely.

Then, according to the Central Limit Theorem (CCT) in (16):

$$\sqrt{n}\, \frac{\bar{x}n - \mathrm{P}}{\sigma} \xrightarrow[n \to \infty]{(d)} N(0,1) \tag{16}$$

Where σ is the standard deviation of the variable X and N(0,1) has a standard deviation (d).

Therefore, we have the following statistical model in (17):

$$(\{0,1], (\mathrm{Ber}(p)_{p \in \{0,1\}}) \tag{17}$$

The mean and variance of our Bernoulli variable are p and p(1-p), respectively.

We define the estimators for the mean and variance as $\hat{p}$ and $Var(\hat{p})$ in equation (18).

Where

$$Var(\hat{p}) = Var(\bar{X}_n) = Var\, \frac{1}{n} \left(\sum_{i=1}^{n} Xi\right) = \frac{Var(p)}{n} = \frac{p(1-p)}{n} \tag{18}$$

We define the confidence interval in equation (17):

$$\lim_{n \to \infty} P\,[\mathrm{I} \ni p] \geq 1 - \alpha,\ \forall\ p \in \Theta \text{ (all the universe of the possible parameters)} \tag{19}$$

The parameter p is the Bernoulli variable we are modeling. To proceed, we used the CLT, replacing variables and using the Slutsky theorem to get the confidence interval I in equations (20) and (21)

$$\sqrt{n}\, \frac{\bar{x}n - \mathrm{P}}{\sqrt{\hat{p}(1.\hat{p})}} \xrightarrow[n \to \infty]{(d)} N(0,1) \tag{20}$$

$$\mathrm{I} = \left[\bar{X}_n - \frac{q_{\alpha/2} \sqrt{\hat{p}(1.\hat{p})}}{n}, \bar{X}_n + \frac{q_{\alpha/2} \sqrt{\hat{p}(1.\hat{p})}}{n} \right] \tag{21}$$

Where:

$$\lim_{n \to \infty} P\ [\mathrm{I} \ni p] = 1 - \alpha \tag{22}$$

where ($(1-\alpha$) is confidence level, α is significance level)

Where P in equation (20) is the probability of the event's occurrence described inside the parenthesis, which is the Interval containing the parameter p. We then define the hypothesis test in (23) and (24) :

$$H_0: p = 0.7041 \tag{23}$$

$$H_1: p \neq 0.7041 \tag{24}$$

Where H_0 is rejected if $\varphi = 1$ given that $\varphi = 1\left\{\sqrt{n}\,\frac{|\bar{x}n - 0.7041|}{\sqrt{0.7041\,(1-0.7041)}} \geq q_{\alpha/2}\right\}$ (25)

Using the hypothesis test φ and the Φ function of the standard normal distribution, it is possible to compute de p-value of the mentioned test (equation (26)).

$$\text{p-value} = 2 * (1 - \Phi\ (Tn)),\ \text{where } Tn = \frac{|\bar{x}n - 0.7041|}{\sqrt{0.7041\,(1-0.7041)}} \tag{26}$$

Replacing and computing we have:

$$\text{p-value} = 0.974074 \tag{27}$$

This p-value indicates that it is impossible to reject H_0 even at a very high α. Considering that α generally has values much lower than 0.05 or 0.10, H_0 will not be rejected. This is because the p-value measures the probability of getting a more extreme value than the one you got from the experiment. If the p-value is greater than alpha, the null hypothesis is accepted. If it is less than alpha, it is rejected the null hypothesis. In this case, the data does not allow to discard that the parameter p (accuracy) is 0.7041. Moreover, the p-value is exceptionally high, indicating that the null hypothesis (p = 0.7041) is almost certainly true (because our sample size is considerable).

As an additional (and more traditional) calculation, it is also possible to compute that the accuracy (p) is in the interval [0.704094, 0.704106] with a confidence interval of (1- α) = 0.95 (95% confidence level with $\alpha = 0.05$).

Both approaches corroborate the accuracy of our experiment and are in line with the asymptotic results of the CLT and LLN when n is greater than 30 (in our case, it is well above that sample size, so our results make sense). The same consideration applies to the rest of the accuracy results. Therefore, the results obtained with the size of our datasets obtained with 12 participants are statistically significant.

This statistical analysis could be repeated for each algorithm, but the results will be similar. In other words, we can extend the results of this statistical proof to the rest of the algorithms used in this thesis. Specifically, as this statistical proof concludes that the algorithm's accuracy of 0.7041 falls within the confidence interval of (0.704094;0.704106), the precision of our estimate of the accuracy of our algorithms is extremely high (with a 95% confidence level). Extending to the rest of the algorithms, this same statistical proof will give us the same conclusions because of $n \rightarrow \infty$ in these other algorithms. We can say this because n is much greater than 30, a general rule of thumb that approximates well according to the Law of Large Numbers (LLN).

6. CONCLUSIONS

In this thesis, several contributions were made. The work was initiated with a systematic review of the literature on emotion recognition using EEG signals and machine learning to identify, evaluate, and synthesize research results in the field to find a niche to work.

One of the findings was that no public datasets were generated using interactive and dynamic ways for emotion elicitation. The most common methods to provoke emotions for recording EEG-related signals are passive because the participants only receive stimuli (video and sound) and do not respond to them. Still, they do not interact with the world around them. Therefore, obtaining and making publicly accessible datasets generated using a different strategy for emotion elicitation would be a helpful contribution in this study area to demonstrate the efficiency of the valence-arousal data acquired that way.

We developed two datasets. Stock-Emotion will be publicly available for emotion recognition using EEG signals. The proposal of interactive and dynamic ways for emotion elicitation was presented: stock market trading activities and playing of online poker games. In the first dataset, each participant was not an expert and faced different and unique market conditions (inherent to simulated live trading), so the experiment offered new and diverse circumstances to the individuals. In the same way, poker games have a random element and therefore offer variable scenarios. The participants in the poker games experiments were also amateurs.

The proposed methods provoked emotions that were tagged each minute using valence-arousal definitions. The labels were related to the key emotions that influence trading: fear (low valence - high arousal), regret or sorrow (low valence - low arousal), hope (high valence - high arousal), and a relaxation state (high valence - low arousal). This fourth state is ideal for traders since it gives more objectivity to their decision-making. Simultaneously, the other three emotions can somehow affect judgment and lead to incorrect decisions. Manikin assessment tools were used for participants to label their emotions.

It would be intended that traders and poker players become aware of their emotions to facilitate their management and achieve an optimal affective state for their activities. Interestingly, in our experiments, the traders did not label any emotion as a relaxed state, except in the baseline, but the poker players did.

Once we had the datasets, we considered that working with raw data limited the long time necessary to process the information for emotion recognition; as presented in Table 15, these raw data processing times are in the range of days. These times will make difficult to use these algorithms in real-time applications. Also, raw data presents infrequent patterns that diminish the classifiers' capacity to recognize emotions, as shown in Table 14.

We extracted features in the time, frequency, and spatial domains to obtain appropriate information for identifying emotional states and fed the classifiers considerably fewer data to enhance their performances and improve their response speed. Time is a critical criterion because the final goal for emotion recognition is to work in real-time.

Considering that feature selection is vital for analyzing EEG data because even with feature extraction, it is high-dimensional, it was clear that such a strategy was a point to be taken into account. Consequently, the present work proposed a feature selection process as a blend of two methods: Mutual Information Matrix (MIM), which presents correlations between features, and Chi-squared statistics that ponder the significance of each feature related to the output classes. First, the information of the MIM allows the elimination of redundant characteristics. Then, Chi-square statistics are applied over the mutually exclusive classes to evaluate the relationship between features and classification outputs; thus, only significant characteristics are maintained. Feature selection allows choosing the most relevant features, and with fewer but more informative inputs, it is possible to reduce processing time while maintaining or improving classification performance.

The suggested emotion elicitation methods effectively provoked emotions related to trading and poker games, and our systems were capable of recognizing them. Due to the subjective aspect of discriminating and self-labeling emotions and the complexity of EEG signals, EEG-based BCI systems for emotion recognition do not readily have high accuracy. However, the system's performance in this research is better than the state-of-the-art systems that recognize a similar number of categories for emotion recognition [2]. The state-of-art presents an accuracy of 76.68% for emotion classification in the four quadrants of the valence-arousal emotion representation space, as presented in Figure 12 in the Literature Review chapter. Our results have higher values with both the original features and with feature selection.

The importance of feature extraction and selection and the efficacy of our proposed methods, using shallow algorithms and neural networks, were also demonstrated in the results in Table 14. If we compute the average accuracy over the four classifiers' performances, we get the following results for average mean accuracy:

- For Stock-Emotion, the original features produce a value of 85.15%, and the selected features generate a performance of 87.76%. There is an improvement of selected features over original characteristics. The raw data got a mean accuracy of 53.36%, which is significative less than the ones obtained with features.

- For the Poker dataset, the original features produce a value of 82.66% versus 82.60% from selected features. In this case, for a slight difference, the better results are from the original features, but mostly we can be said that the results are maintained with and without feature selection. The raw data, on the other hand, has a performance of 55.74%, which is smaller than the results with features.

- With a subset of 8 and 14 channels of the DEAP dataset, the accuracy results with the same machine learning algorithms are minor in all cases: with original features, raw data, and selected features. To explain the better classification results for our datasets, we have to consider the influence of using a decreased number of electrodes from 32 in the original dataset to 8 and 14. Also, we might dare to speculate that in our data, being the participants motivated to have emotions actively, to take actions, and make decisions in the market or the poker game, perhaps their feelings are more intense and easier to recognize by the algorithms. While in the DEAP dataset, the emotions have a passive way of elicitation: just watching videos. However, the presented results are not conclusive in this sense.

It is important to notice that neural networks (MLP and 1DCNN) produce better results with raw data than other algorithms such as kNN or Random Forest. The explication for that is the capacity of the neural networks to extract features as embedded characteristics in the weight assignment process.

We can conclude that feature extraction and selection, using algorithms that extract necessary information from the EEG signal is better than using raw data. Moreover, feature selection with criteria that produce non-redundant and relevant information allows results that maintain or improve performance.

REFERENCES

[1] R. W. Picard, “Affective Computing for HCI,” *Proc. 8th HCI Int. Human-Computer Interact. Ergon. User Interfaces*, pp. 829–833, 1999.

[2] H. A. Elfenbein and N. Ambady, “Predicting workplace outcomes from the ability to eavesdrop on feelings,” *J. Appl. Psychol.*, vol. 87, no. 5, pp. 963–971, 2002.

[3] S. Goenaga, L. Navarro, C. G. Quintero M., and M. Pardo, “Imitating Human Emotions with a NAO Robot as Interviewer Playing the Role of Vocational Tutor,” *Electronics*, vol. 9, no. 6, p. 971, Jun. 2020.

[4] C. D. Salzman and S. Fusi, “Emotion, Cognition, and Mental State Representation in Amygdala and Prefrontal Cortex,” *Annu. Rev. Neurosci.*, vol. 33, no. 1, pp. 173–202, Jun. 2010.

[5] K. J. Panoulas, L. J. Hadjileontiadis, and S. M. Panas, “Brain-computer interface (BCI): Types, processing perspectives and applications,” in *Smart Innovation, Systems and Technologies*, 2010, vol. 3, pp. 299–321.

[6] A. Konar and A. Chakraborty, *Emotion Recognition*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2015.

[7] P. Ekman, “Ekman 1992.pdf,” *Psychological Review*, vol. 99, no. 3. pp. 550–553, 1992.

[8] G. K. Verma and U. S. Tiwary, “Affect representation and recognition in 3D continuous valence–arousal–dominance space,” *Multimed. Tools Appl.*, vol. 76, no. 2, 2017.

[9] O. Bălan, G. Moise, A. Moldoveanu, M. Leordeanu, and F. Moldoveanu, “Fear level classification based on emotional dimensions and machine learning techniques,” *Sensors (Switzerland)*, vol. 19, no. 7, pp. 1–18, 2019.

[10] J. A. Russell, “A circumplex model of affect.,” *J. Pers. Soc. Psychol.*, vol. 39, no. 6, p. 1161, 1980.

[11] M. M. Bradley and P. J. Lang, “Measuring emotion: The self-assessment manikin and the semantic differential,” *J. Behav. Ther. Exp. Psychiatry*, vol. 25, no. 1, pp. 49–59, Mar. 1994.

[12] P. Ekman and R. J. Davidson, “Voluntary Smiling Changes Regional Brain Activity,” 1993.

[13] A. M. Batti, M. Majid, S. M. Anwar, and B. Khan, “Human emotion recognition and analysis in response to audio music using brain signals,” *Comput. Human Behav.*, vol. 65, pp. 267–275, Dec. 2016.

[14] Y.-Y. Lee and S. Hsieh, “Classifying Different Emotional States by Means of EEG-Based Functional Connectivity Patterns,” *PLoS One*, vol. 9, no. 4, p. e95415, Apr. 2014.

[15] W. L. Zheng, H. T. Guo, and B. L. Lu, “Revealing critical channels and frequency bands for emotion recognition from EEG with deep belief network,” *Int. IEEE/EMBS Conf. Neural Eng. NER*, vol. 2015-July, pp. 154–157, 2015.

[16] G. G. Knyazev and J. Y. Slobodskoj-Plusnin, “Behavioural approach system as a moderator of emotional arousal elicited by reward and punishment cues,” *Pers. Individ. Dif.*, vol. 42, no. 1, pp. 49–59, 2007.

[17] E. Kirmizi-Alsan, Z. Bayraktaroglu, H. Gurvit, Y. H. Keskin, M. Emre, and T. Demiralp, “Comparative analysis of event-related potentials during Go/NoGo and CPT: Decomposition of electrophysiological markers of response inhibition and sustained attention,” *Brain Res.*, vol. 1104, no. 1, pp. 114–128, 2006.

[18] A. Hyvarinen, “New approximations of differential entropy for independent component analysis and projection pursuit,” in *Advances in Neural Information Processing Systems*, 1998, pp. 273–279.

[19] S. Hamann, “Mapping discrete and dimensional emotions onto the brain: Controversies and consensus,” *Trends in Cognitive Sciences*, vol. 16, no. 9. Elsevier Ltd, pp. 458–466, 2012.

[20] R. J. Davidson, P. Ekman, C. D. Saron, J. A. Senulis, and W. V Friesen, “Approach-Withdrawl-And-Cerebral-Asymmetry-Emotional-Expres Davidson 1990.pdf,” *Journal of personality and social psychology*, vol. 58, no. 2. pp. 330–341, 1990.

[21] S. Sanei and J. A. Chambers, *EEG Signal Processing*. John Wiley & Sons, 2013.

[22] V. Peterson, C. Galván, H. Hernández, and R. Spies, “A feasibility study of a complete low-cost consumer-grade brain-computer interface system,” *Heliyon*, vol. 6, no. 3, Mar.

2020.

[23] A. Savran *et al.*, "Emotion Detection in the Loop from Brain Signals and Facial Images," *eNTERFACE*, vol. 06, pp. 69–80, 2006.

[24] J. Onton and S. Makeig, "High-frequency broadband modulations of electroencephalographic spectra," *Front. Hum. Neurosci.*, vol. 3, no. DEC, pp. 1–18, 2009.

[25] M. Yadava, P. Kumar, R. Saini, P. P. Roy, and D. Prosad Dogra, "Analysis of EEG signals and its application to neuromarketing," *Multimed. Tools Appl.*, vol. 76, no. 18, pp. 19087–19111, 2017.

[26] W. L. Zheng, W. Liu, Y. Lu, B. L. Lu, and A. Cichocki, "EmotionMeter: A Multimodal Framework for Recognizing Human Emotions," *IEEE Trans. Cybern.*, vol. 49, no. 3, pp. 1110–1122, Mar. 2019.

[27] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 42–55, 2012.

[28] C. Grégoire, P. L. C. Rodrigues, and M. Congedo, "EEG Alpha Waves dataset," Grenoble, France, 2019.

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

[30] J. A. Miranda-Correa, M. K. Abadi, N. Sebe, and I. Patras, "AMIGOS: A Dataset for Affect, Personality and Mood Research on Individuals and Groups," *IEEE Trans. Affect. Comput.*, vol. 12, no. 2, pp. 479–493, Apr. 2021.

[31] Y. Li, W. Zheng, Z. Cui, Y. Zong, and S. Ge, "EEG Emotion Recognition Based on Graph Regularized Sparse Linear Regression," *Neural Process. Lett.*, vol. 49, no. 2, pp. 555–571, 2019.

[32] P. J. Lang *et al.*, "International affective picture system (IAPS): Technical manual and affective ratings," *NIMH Cent. Study Emot. Atten.*, pp. 39–58, 1997.

[33] W. Yang *et al.*, "Affective auditory stimulus database: An expanded version of the

International Affective Digitized Sounds (IADS-E)," *Behav. Res. Methods*, vol. 50, no. 4, pp. 1415–1429, Aug. 2018.

[34] C. Mühl, B. Allison, A. Nijholt, and G. Chanel, "A survey of affective brain computer interfaces: principles, state-of-the-art, and challenges," *Brain-Computer Interfaces*, vol. 1, no. 2, pp. 66–84, Apr. 2014.

[35] F. Zhou, X. Qu, J. Jiao, and M. G. Helander, "Emotion prediction from physiological signals: A comparison study between visual and auditory elicitors," *Interact. Comput.*, vol. 26, no. 3, pp. 285–302, 2014.

[36] F. Pallavicini, A. Ferrari, A. Pepe, and G. Garcea, "Effectiveness of virtual reality survival horror games for the emotional elicitation: Preliminary insights using Resident Evil 7: Biohazard GAMING FOR HEALTH-Video Games, Virtual Reality and Augmented Reality for the assessment and training of emotional a," no. June, 2018.

[37] V. C. C. Roza and O. A. Postolache, "Multimodal approach for emotion recognition based on simulated flight experiments," *Sensors (Switzerland)*, vol. 19, no. 24, Dec. 2019.

[38] D. Iacoviello, A. Petracca, M. Spezialetti, and G. Placidi, "A real-time classification algorithm for EEG-based BCI driven by self-induced emotions," *Comput. Methods Programs Biomed.*, vol. 122, no. 3, pp. 293–303, Dec. 2015.

[39] D. Novak, M. Mihelj, and M. Munih, "A survey of methods for data fusion and system adaptation using autonomic nervous system responses in physiological computing," *Interact. Comput.*, vol. 24, no. 3, pp. 154–172, 2012.

[40] P. A. Bustamante, N. M. Lopez Celani, M. E. Perez, and O. L. Quintero Montoya, "Recognition and regionalization of emotions in the arousal-valence plane," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015, vol. 2015-Novem, pp. 6042–6045.

[41] B. W. G. Abhang, Priyanka A. Suresh C. Mehrotra, *Introduction to EEG- and Speech-Based Emotion Recognition*. Elsevier, 2016.

[42] R. (Ricardo) Jardim-Gonçalves, Universidade Nova de Lisboa. Faculdade de Ciências e Tecnologia, Institute of Electrical and Electronics Engineers, IEEE Technology Engineering and Management Society., and IEEE International Technology

Management Conference (2017 : Madeira Islands), “2017 International Conference on Engineering, Technology and Innovation: Engineering, Technology and Innovation Management Beyond 2020: New Challenges, New Approaches, ICE/ITMC 2017 - Proceedings,” *2017 International Conference on Engineering, Technology and Innovation: Engineering, Technology and Innovation Management Beyond 2020: New Challenges, New Approaches, ICE/ITMC 2017 - Proceedings*, vol. 2018-Janua. pp. 1578–1583, 2018.

[43] M. J. Alhaddad, M. Kamel, H. Malibary, K. Thabit, F. Dahlwi, and A. Hadi, “P300 speller efficiency with common average reference,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7326 LNAI, pp. 234–241, 2012.

[44] M. J. Alhaddad, M. Kamel, H. Malibary, K. Thabit, F. Dahlwi, and A. Hadi, “LNAI 7326 - P300 Speller Efficiency with Common Average Reference,” 2012.

[45] M. Murugappan, R. Nagarajan, and S. Yaacob, “Combining spatial filtering and wavelet transform for classifying human emotions using EEG Signals,” *J. Med. Biol. Eng.*, vol. 31, no. 1, pp. 45–51, 2011.

[46] M. Murugappan and S. Murugappan, “Human emotion recognition through short time Electroencephalogram (EEG) signals using Fast Fourier Transform (FFT),” in *2013 IEEE 9th International Colloquium on Signal Processing and its Applications*, 2013, pp. 289–294.

[47] B. Burle, L. Spieser, C. Roger, L. Casini, T. Hasbroucq, and F. Vidal, “Spatial and temporal resolutions of EEG: Is it really black and white? A scalp current density view,” *Int. J. Psychophysiol.*, vol. 97, no. 3, pp. 210–220, Sep. 2015.

[48] I. Mazumder, “An Analytical Approach of EEG Analysis for Emotion Recognition,” *Proc. 3rd Int. Conf. 2019 Devices Integr. Circuit, DevIC 2019*, pp. 256–260, 2019.

[49] A. Subasi and M. I. Gursoy, “EEG signal classification using PCA, ICA, LDA and support vector machines,” *Expert Syst. Appl.*, vol. 37, no. 12, pp. 8659–8666, 2010.

[50] H. Lee and S. Choi, “PCA+HMM+SVM FOR EEG PATTERN CLASSIFICATION.”

[51] V. Doma and M. Pirouz, “A comparative analysis of machine learning methods for emotion recognition using EEG and peripheral physiological signals,” *J. Big Data*, vol. 7, no. 1, 2020.

[52] L. Shaw and A. Routray, “Statistical features extraction for multivariate pattern analysis in meditation EEG using PCA,” in *2016 IEEE EMBS International Student Conference (ISC)*, 2016, pp. 1–4.

[53] D. Iacoviello, A. Petracca°, M. Spezialetti°, and G. Placidi°, “A Real-time classification algorithm for EEG-based BCI driven by self-induced emotions.”

[54] J. Liu, H. Meng, M. Li, F. Zhang, R. Qin, and A. K. Nandi, “Emotion detection from EEG recordings based on supervised and unsupervised dimension reduction,” *Concurr. Comput.* , vol. 30, no. 23, pp. 1–13, 2018.

[55] X. Yong, R. K. Ward, and G. E. Birch, “Robust common spatial patterns for EEG signal preprocessing,” in *Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS’08 - “Personalized Healthcare through Technology,”* 2008, pp. 2087–2090.

[56] X. Li, H. Fan, H. Wang, and L. Wang, “Common spatial patterns combined with phase synchronization information for classification of EEG signals,” *Biomed. Signal Process. Control*, vol. 52, pp. 248–256, Jul. 2019.

[57] B. Interfaces, “A Tutorial on EEG Signal Processing Techniques for Mental State Recognition in Brain-Computer Interfaces,” *Guid. to Brain-Computer Music Interfacing*, no. August, 2014.

[58] R. Jenke, A. Peer, and M. Buss, “Feature Extraction and Selection for Emotion Recognition from EEG,” *IEEE Trans. Affect. Comput.*, vol. 5, no. 3, pp. 327–339, Jul. 2014.

[59] A. S. Al-Fahoum and A. A. Al-Fraihat, “Methods of EEG Signal Features Extraction Using Linear Analysis in Frequency and Time-Frequency Domains,” *ISRN Neurosci.*, vol. 2014, pp. 1–7, 2014.

[60] P. C. Petrantonakis and L. J. Hadjileontiadis, “Emotion recognition from brain signals using hybrid adaptive filtering and higher order crossings analysis,” *IEEE Trans. Affect. Comput.*, vol. 1, no. 2, pp. 81–97, Jul. 2010.

[61] S. G. Torres, Edgar P ; Torres, Edgar A; Hernandez-Alvarez Myriam; Yoo, E. A. P. Torres, E. A. P. Torres, M. Hernández-álvarez, and S. G. Yoo, “Machine Learning Analysis of EEG Measurements of Stock Trading Performance,” in *Advances in Artificial*

Intelligence, Software and Systems Engineering, Springer Nature, 2021.

[62] P. Kubben, M. Dumontier, and A. Dekker, "Fundamentals of clinical data science," *Fundam. Clin. Data Sci.*, pp. 1–219, 2018.

[63] E. Karahan, P. A. Rojas-Lopez, M. L. Bringas-Vega, P. A. Valdes-Hernandez, and P. A. Valdes-Sosa, "Tensor Analysis and Fusion of Multimodal Brain Images," *Proc. IEEE*, vol. 103, no. 9, pp. 1531–1559, Sep. 2015.

[64] I. Winkler, S. Debener, K.-R. Müller, M. Tangermann, K. R. Muller, and M. Tangermann, "On the influence of high-pass filtering on ICA-based artifact reduction in EEG-ERP," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, vol. 2015-Novem, pp. 4101–4105, 2015.

[65] Y. Zhang, G. Zhou, Q. Zhao, J. Jin, X. Wang, and A. Cichocki, "Spatial-temporal discriminant analysis for ERP-based brain-computer interface," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 21, no. 2, pp. 233–243, 2013.

[66] A. M. Brouwer, T. O. Zander, J. B. F. van Erp, J. E. Korteling, and A. W. Bronkhorst, "Using neurophysiological signals that reflect cognitive or affective state: Six recommendations to avoid common pitfalls," *Front. Neurosci.*, vol. 9, no. APR, pp. 1–11, 2015.

[67] Z. Wu, D. Yao, Y. Tang, Y. Huang, and S. Su, "Amplitude modulation of steady-state visual evoked potentials by event-related potentials in a working memory task," *J. Biol. Phys.*, vol. 36, no. 3, pp. 261–271, Jun. 2010.

[68] V. Abootalebi, M. H. Moradi, and M. A. Khalilzadeh, "A new approach for EEG feature extraction in P300-based lie detection," *Comput. Methods Programs Biomed.*, vol. 94, no. 1, pp. 48–57, Apr. 2009.

[69] P. R. Bhise, S. B. Kulkarni, and T. A. Aldhaheri, *Brain Computer Interface based EEG for Emotion Recognition System: A Systematic Review*. .

[70] X. Li, D. Song, P. Zhang, Y. Zhang, Y. Hou, and B. Hu, "Exploring EEG features in cross-subject emotion recognition," *Front. Neurosci.*, vol. 12, no. MAR, Mar. 2018.

[71] M. Z. Hossain, M. M. Kabir, and M. Shahjahan, "Feature selection of EEG data with neuro-statistical method," *2013 Int. Conf. Electr. Inf. Commun. Technol. EICT 2013*, no.

February, 2014.

[72] S. Bavkar, B. Iyer, and S. Deosarkar, “Detection of alcoholism: An EEG hybrid features and ensemble subspace K-NN based approach,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2019, vol. 11319 LNCS, pp. 161–168.

[73] E. S. Pane, A. D. Wibawa, and M. H. Pumomo, “Channel Selection of EEG Emotion Recognition using Stepwise Discriminant Analysis,” *2018 Int. Conf. Comput. Eng. Netw. Intell. Multimedia, CENIM 2018 - Proceeding*, no. 4, pp. 14–19, 2018.

[74] M. Musselman and D. Djurdjanovic, “Time-frequency distributions in the classification of epilepsy from EEG signals,” *Expert Syst. Appl.*, vol. 39, no. 13, pp. 11413–11422, 2012.

[75] H. Xu and K. N. Plataniotis, “Affect recognition using EEG signal,” *2012 IEEE 14th Int. Work. Multimed. Signal Process. MMSP 2012 - Proc.*, pp. 299–304, 2012.

[76] Y. Liu and O. Sourina, “EEG databases for emotion recognition,” *Proc. - 2013 Int. Conf. Cyberworlds, CW 2013*, pp. 302–309, 2013.

[77] R. N. Duan, J. Y. Zhu, and B. L. Lu, “Differential entropy feature for EEG-based emotion classification,” *Int. IEEE/EMBS Conf. Neural Eng. NER*, no. March, pp. 81–84, 2013.

[78] W. L. Zheng and B. L. Lu, “Investigating Critical Frequency Bands and Channels for EEG-Based Emotion Recognition with Deep Neural Networks,” *IEEE Trans. Auton. Ment. Dev.*, vol. 7, no. 3, pp. 162–175, Sep. 2015.

[79] S. Thejaswini *et al.*, “Analysis of EEG Based Emotion Detection of DEAP and SEED-IV Databases Using SVM,” *Int. J. Recent Technol. Eng.*, vol. 8, no. 1, pp. 207–211, 2019.

[80] X. H. Wang, T. Zhang, X. M. Xu, L. Chen, X. F. Xing, and C. L. P. Chen, “EEG Emotion Recognition Using Dynamical Graph Convolutional Neural Networks and Broad Learning System,” *Proc. - 2018 IEEE Int. Conf. Bioinforma. Biomed. BIBM 2018*, pp. 1240–1244, 2019.

[81] X. Wu, W.-L. Zheng, and B.-L. Lu, “Investigating EEG-Based Functional Connectivity Patterns for Multimodal Emotion Recognition,” Apr. 2020.

[82] J. Li, S. Qiu, C. Du, Y. Wang, and H. He, “Domain adaptation for eeg emotion recognition based on latent representation similarity,” *IEEE Trans. Cogn. Dev. Syst.*, vol. 12, no. 2,

pp. 344–353, 2020.

[83] W.-L. Zheng, J.-Y. Zhu, and B.-L. Lu, “Identifying Stable Patterns over Time for Emotion Recognition from EEG,” *IEEE Trans. Affect. Comput.*, vol. 10, no. 3, pp. 417–429, Jan. 2017.

[84] Y. Yang, Q. M. J. Wu, W. L. Zheng, and B. L. Lu, “EEG-based emotion recognition using hierarchical network with subnetwork nodes,” *IEEE Trans. Cogn. Dev. Syst.*, vol. 10, no. 2, pp. 408–419, Jun. 2018.

[85] P. Li *et al.*, “EEG Based Emotion Recognition by Combining Functional Connectivity Network and Local Activations,” *IEEE Trans. Biomed. Eng.*, vol. 66, no. 10, pp. 2869–2881, 2019.

[86] Z. Li, X. Tian, L. Shu, X. Xu, and B. Hu, “Emotion recognition from EEG using RASM and LSTM,” in *Communications in Computer and Information Science*, 2018, vol. 819, pp. 310–318.

[87] M. R. Mowla, R. I. Cano, K. J. Dhuyvetter, and D. E. Thompson, “Affective Brain-Computer Interfaces: A Tutorial to Choose Performance Measuring Metric,” May 2020.

[88] Z. Lan, O. Sourina, L. Wang, R. Scherer, and G. R. Muller-Putz, “Domain Adaptation Techniques for EEG-Based Emotion Recognition: A Comparative Study on Two Public Datasets,” *IEEE Trans. Cogn. Dev. Syst.*, vol. 11, no. 1, pp. 85–94, Mar. 2019.

[89] P. C. Petrantonakis and L. J. Hadjileontiadis, “Emotion recognition from EEG using higher order crossings,” *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 2, pp. 186–197, Mar. 2010.

[90] M. K. Kim, M. Kim, E. Oh, and S. P. Kim, “A review on the computational methods for emotional state estimation from the human EEG,” *Computational and Mathematical Methods in Medicine*, vol. 2013. 2013.

[91] H. J. Yoon and S. Y. Chung, “EEG-based emotion estimation using Bayesian weighted-log-posterior function and perceptron convergence algorithm,” *Comput. Biol. Med.*, vol. 43, no. 12, pp. 2230–2237, Dec. 2013.

[92] S. M. Hosni, M. E. Gadallah, S. F. Bahgat, and M. S. AbdelWahab, “Classification of EEG signals using different feature extraction techniques for mental-task BCI,” in

ICCES'07 - 2007 International Conference on Computer Engineering and Systems, 2007, pp. 220–226.

[93] X. Xing, Z. Li, T. Xu, L. Shu, B. Hu, and X. Xu, "SAE+LSTM: A new framework for emotion recognition from multi-channel EEG," *Front. Neurorobot.*, vol. 13, no. June, pp. 1–14, 2019.

[94] I. Navarro, F. Sepulveda, and B. Hubais, "A Comparison of Time, Frequency and ICA Based Features and Five Classifiers for Wrist Movement Classification in EEG Signals."

[95] W. Ting, Y. Guo-zheng, Y. Bang-hua, and S. Hong, "EEG feature extraction based on wavelet packet decomposition for brain computer interface," *Meas. J. Int. Meas. Confed.*, vol. 41, no. 6, pp. 618–625, Jul. 2008.

[96] J. Guo, F. Fang, W. Wang, and F. Ren, "EEG Emotion Recognition Based on Granger Causality and CapsNet Neural Network," *Proc. 2018 5th IEEE Int. Conf. Cloud Comput. Intell. Syst. CCIS 2018*, pp. 47–52, 2019.

[97] D. Sander, D. Grandjean, and K. R. Scherer, "A systems approach to appraisal mechanisms in emotion," *Neural Networks*, vol. 18, no. 4, pp. 317–352, 2005.

[98] G. Chanel, J. J. M. Kierkels, M. Soleymani, and T. Pun, "Short-term emotion assessment in a recall paradigm," *Int. J. Hum. Comput. Stud.*, vol. 67, no. 8, pp. 607–627, Aug. 2009.

[99] F. Lotte *et al.*, "A review of classification algorithms for EEG-based brain – computer interfaces To cite this version : A Review of Classification Algorithms for EEG-based Brain-Computer Interfaces," *Hum. Brain Mapp.*, vol. 38, no. 11, pp. 270–278, 2018.

[100] S. Sanei and J. Chambers, *EEG signal processing*. John Wiley & Sons, 2007.

[101] R. Jenke, A. Peer, and M. Buss, "Effect-size-based electrode and feature selection for emotion recognition from EEG," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013, pp. 1217–1221.

[102] A. Ella Hassanien Ahmad Taher Azar Editors, "Intelligent Systems Reference Library 74 Brain-Computer Interfaces Current Trends and Applications."

[103] L. Zhang, G. Xiong, H. Liu, H. Zou, and W. Guo, "Time-frequency representation based on time-varying autoregressive model with applications to non-stationary rotor vibration analysis," *Sadhana - Acad. Proc. Eng. Sci.*, vol. 35, no. 2, pp. 215–232, 2010.

[104] N. J. Hill and J. R. Wolpaw, “Brain–Computer Interface☆,” in *Reference Module in Biomedical Sciences*, Elsevier, 2016.

[105] M. Rashid *et al.*, “Current Status, Challenges, and Possible Solutions of EEG-Based Brain-Computer Interface: A Comprehensive Review,” *Front. Neurorobot.*, vol. 14, Jun. 2020.

[106] F. Lotte *et al.*, *A review of classification algorithms for EEG-based brain–computer interfaces*, vol. 4. IOP Publishing, 2007.

[107] S. Vaid, P. Singh, and C. Kaur, “EEG signal analysis for BCI interface: A review,” in *International Conference on Advanced Computing and Communication Technologies, ACCT*, 2015, vol. 2015-April, pp. 143–147.

[108] P. Ackermann, C. Kohlschein, J. Á. Bitsch, K. Wehrle, and S. Jeschke, “EEG-based automatic emotion recognition: Feature extraction, selection and classification methods,” *2016 IEEE 18th Int. Conf. e-Health Networking, Appl. Serv. Heal. 2016*, 2016.

[109] R. Atangana, D. Tchiotsop, G. Kenne, and L. C. DjoufackNkengfac k, “EEG Signal Classification using LDA and MLP Classifier,” *Heal. Informatics - An Int. J.*, vol. 9, no. 1, pp. 14–32, Feb. 2020.

[110] S. Srivastava, M. R. Gupta, and B. A. Frigyik, “Bayesian quadratic discriminant analysis,” *J. Mach. Learn. Res.*, vol. 8, pp. 1277–1305, 2007.

[111] P. Tzirakis, G. Trigeorgis, M. A. Nicolaou, B. W. Schuller, and S. Zafeiriou, “End-to-End Multimodal Emotion Recognition using Deep Neural Networks.”

[112] Y. Cimtay and E. Ekmekcioglu, “Investigating the use of pretrained convolutional neural network on cross-subject and cross-dataset eeg emotion recognition,” *Sensors (Switzerland)*, vol. 20, no. 7, p. 2024, 2020.

[113] J. X. Chen, P. W. Zhang, Z. J. Mao, Y. F. Huang, D. M. Jiang, and Y. N. Zhang, “Accurate EEG-Based Emotion Recognition on Combined Features Using Deep Convolutional Neural Networks,” *IEEE Access*, vol. 7, no. c, pp. 44317–44328, 2019.

[114] W. Zhang, F. Wang, Y. Jiang, Z. Xu, S. Wu, and Y. Zhang, *Cross-subject EEG-based emotion recognition with deep domain confusion*, vol. 11740 LNAI. Springer International Publishing, 2019.

[115] U. Lechner, *Scientific Workflow Scheduling for Cloud Computing Environments*, vol. 2, no. July. Springer International Publishing, 2019.

[116] F. Babiloni *et al.*, “Mahalanobis distance-based classifiers are able to recognize EEG patterns by using few EEG electrodes,” *Annu. Reports Res. React. Institute, Kyoto Univ.*, vol. 1, pp. 651–654, 2001.

[117] S. Sun, C. Zhang, and D. Zhang, “An experimental evaluation of ensemble methods for EEG signal classification,” *Pattern Recognit. Lett.*, vol. 28, no. 15, pp. 2157–2163, Nov. 2007.

[118] L. Fraiwan, K. Lweesy, N. Khasawneh, H. Wenz, and H. Dickhaus, “Automated sleep stage identification system based on time-frequency analysis of a single EEG channel and random forest classifier,” *Comput. Methods Programs Biomed.*, vol. 108, no. 1, pp. 10–19, 2012.

[119] P. Kumar, M. Valentina, E. Balas, A. Kumar Bhoi, and G.-S. Chae, “Advances in Intelligent Systems and Computing 1040 Cognitive Informatics and Soft Computing.”

[120] T. Lv, J. Yan, and H. Xu, “An EEG emotion recognition method based on AdaBoost classifier,” in *2017 Chinese Automation Congress (CAC)*, 2017, pp. 6050–6054.

[121] N. Japkowicz and M. Shah, *Evaluating Learning Algorithms*. Cambridge: Cambridge University Press, 2011.

[122] B. Z. Allison, S. Dunne, R. Leeb, J. Del R. Millán, and A. Nijholt, Eds., *Towards Practical Brain-Computer Interfaces*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013.

[123] E. Combrisson and K. Jerbi, “Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy,” *J. Neurosci. Methods*, vol. 250, pp. 126–136, Jul. 2015.

[124] D. G. Bonett and R. M. Price, “Adjusted Wald Confidence Interval for a Difference of Binomial Proportions Based on Paired Data,” *J. Educ. Behav. Stat.*, vol. 37, no. 4, pp. 479–488, 2012.

[125] Kitcheman Barbara, “Procedures for Performing Systematic Reviews,” 2004.

[126] S. D. Kreibig, “Autonomic nervous system activity in emotion: A review,” *Biol. Psychol.*, vol. 84, no. 3, pp. 394–421, 2010.

[127] F. Feradov, I. Mporas, and T. Ganchev, “Evaluation of features in detection of dislike responses to audio–visual stimuli from EEG signals,” *Computers*, vol. 9, no. 2, 2020.

[128] J. Atkinson and D. Campos, “Improving BCI-based emotion recognition by combining EEG feature selection and kernel classifiers,” *Expert Syst. Appl.*, vol. 47, pp. 35–41, Apr. 2016.

[129] B. Kaur, D. Singh, and P. P. Roy, “EEG Based Emotion Classification Mechanism in BCI,” in *Procedia Computer Science*, 2018, vol. 132, pp. 752–758.

[130] Y. J. Liu, M. Yu, G. Zhao, J. Song, Y. Ge, and Y. Shi, “Real-time movie-induced discrete emotion recognition from EEG signals,” *IEEE Trans. Affect. Comput.*, vol. 9, no. 4, pp. 550–562, Oct. 2018.

[131] J. Yan, S. Chen, and S. Deng, “A EEG-based emotion recognition model with rhythm and time characteristics,” *Brain Informatics*, vol. 6, no. 1, 2019.

[132] Y. Li, W. Zheng, Z. Cui, T. Zhang, and Y. Zong, “A novel neural network model based on cerebral hemispheric asymmetry for EEG emotion recognition,” *IJCAI Int. Jt. Conf. Artif. Intell.*, vol. 2018-July, pp. 1561–1567, 2018.

[133] Z. M. Wang, S. Y. Hu, and H. Song, “Channel Selection Method for EEG Emotion Recognition Using Normalized Mutual Information,” *IEEE Access*, vol. 7, pp. 143303–143311, 2019.

[134] S. Parui, A. Kumar, R. Bajiya, D. Samanta, and N. Chakravorty, “Emotion recognition from EEG signal using XGBoost algorithm,” *2019 IEEE 16th India Counc. Int. Conf. INDICON 2019 - Symp. Proc.*, pp. 1–4, 2019.

[135] N. Kumar, K. Khaund, and S. M. Hazarika, “Bispectral Analysis of EEG for Emotion Recognition,” *Procedia Comput. Sci.*, vol. 84, pp. 31–35, 2016.

[136] Y. Liu and O. Sourina, “EEG-based subject-dependent emotion recognition algorithm using fractal dimension,” *Conf. Proc. - IEEE Int. Conf. Syst. Man Cybern.*, vol. 2014-Janua, no. January, pp. 3166–3171, 2014.

[137] N. Thammasan, K. Moriyama, K. ichi Fukui, and M. Numao, “Familiarity effects in EEG-based emotion recognition,” *Brain Informatics*, vol. 4, no. 1, pp. 39–50, 2017.

[138] I. Technology, “Yongbin Gaol, Hyo Jong Lee 1,2, Raja Majid Mehmood1,” *Deep Learn.*

Eeg Signals Emot. Recognit., vol. 2, pp. 1–5.

[139] M. S. Özerdem and H. Polat, “Emotion recognition based on EEG features in movie clips with channel selection,” *Brain Informatics*, vol. 4, no. 4, pp. 241–252, Dec. 2017.

[140] S. Alhagry, A. Aly, and R. A., “Emotion Recognition based on EEG using LSTM Recurrent Neural Network,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 10, pp. 8–11, 2017.

[141] E. S. Salama, R. A. El-Khoribi, M. E. Shoman, and M. A. Wahby Shalaby, “EEG-based emotion recognition using 3D convolutional neural networks,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 8, pp. 329–337, 2018.

[142] S.-E. Moon, S. Jang, and J.-S. Lee, “Convolutional Neural Network Approach for Eeg-Based Emotion Recognition Using Brain Connectivity and its Spatial Information,” in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 2556–2560.

[143] M. A. Asghar, M. J. Khan, M. Rizwan, M. Shorfuzzaman, and R. M. Mehmood, “AI inspired EEG-based spatial feature selection method using multivariate empirical mode decomposition for emotion classification,” *Multimed. Syst.*, Apr. 2021.

[144] K. H. Et.al, “Analysis of Emotion Recognition Model Using Electroencephalogram (EEG) Signals Based on Stimuli Text,” *Turkish J. Comput. Math. Educ.*, vol. 12, no. 3, pp. 1384–1393, 2021.

[145] J. Cheng *et al.*, “Emotion Recognition From Multi-Channel EEG via Deep Forest,” *IEEE J. Biomed. Heal. Informatics*, vol. 25, no. 2, pp. 453–464, Feb. 2021.

[146] A. Topic, M. Russo, M. Stella, and M. Saric, “Emotion Recognition Using a Reduced Set of EEG Channels Based on Holographic Feature Maps,” *Sensors*, vol. 22, no. 9, p. 3248, Apr. 2022.

[147] B. Pan and W. Zheng, “Emotion Recognition Based on EEG Using Generative Adversarial Nets and Convolutional Neural Network,” *Comput. Math. Methods Med.*, vol. 2021, pp. 1–11, Oct. 2021.

[148] C. Mühl *et al.*, “Bacteria Hunt: Evaluating multi-paradigm BCI interaction,” *J. Multimodal User Interfaces*, vol. 4, no. 1, pp. 11–25, Mar. 2010.

[149] G. Chanel, C. Rebetez, M. Bétrancourt, and T. Pun, “for Adaptation of Game Difficulty,”

Blood Press., vol. 41, no. 6, pp. 1052–1063, 2011.

[150] N. I. M. Razi, M. Othman, and H. Yaacob, “EEG-Based Emotion Recognition in the Investment Activities,” in *2016 6th International Conference on Information and Communication Technology for The Muslim World (ICT4M)*, 2016, pp. 325–329.

[151] E. P. Torres P, E. A. Torres H, M. Hernandez-Alvarez, and S. G. Yoo, “EEG-Based BCI Emotion Recognition : A Survey,” *Sensors (Switzerland)*, no. Emotion and Stress Recognition Related Sensors and Machine Learning Technologies, 2020.

[152] E. H. Houssein, *Human emotion recognition from EEG-based brain – computer interface using machine learning : a comprehensive review*, vol. 4. Springer London, 2022.

[153] M. W. Johnson and W. K. Bickel, “Within-Subject Comparison of Real and Hypothetical Money Rewards in Delay Discounting,” *J. Exp. Anal. Behav.*, vol. 77, no. 2, pp. 129–146, 2002.

[154] M. B. Mulder and B. A. Beheim, “Understanding the nature of wealth and its effects on human fitness,” *Philos. Trans. R. Soc. B Biol. Sci.*, vol. 366, no. 1563, pp. 344–356, 2011.

[155] D. J. Butler and S. L. Cheung, *Mind, Body, Bubble! Psychological and Biophysical Dimensions of Behavior in Experimental Asset Markets*. Elsevier Inc., 2019.

[156] M. Fenton-O’Creevy, J. T. Lins, S. Vohra, D. W. Richards, G. Davies, and K. Schaaff, “Emotion regulation and trader expertise: Heart rate variability on the trading floor,” *J. Neurosci. Psychol. Econ.*, vol. 5, no. 4, pp. 227–237, 2012.

[157] E. T. Rolls, “On the brain and emotion,” *Behav. Brain Sci.*, vol. 23, no. 2, pp. 219–228, 2000.

[158] C. Hodkinson, “‘Fear of Missing Out’ (FOMO) marketing appeals: A conceptual model,” *J. Mark. Commun.*, vol. 25, no. 1, pp. 65–88, 2019.

[159] A. Järvinen, “Games without Frontiers. Theories and methods for game studies and design.,” 2009.

[160] M. A. Strahilevitz, T. Odean, and B. M. Barber, “Once burned, twice shy: How naive learning, counterfactuals, and regret affect the repurchase of stocks previously sold,” *J. Mark. Res.*, vol. 48, no. SPEC. ISSUE, pp. 102–120, 2011.

[161] E. P. Torres, E. A. Torres, M. Hernández-Álvarez, and S. G. Yoo, "Machine Learning Analysis of EEG Measurements of Stock Trading Performance," in *Advances in Artificial Intelligence, Software and Systems Engineering*, Springer Nature, 2021, pp. 53–60.

[162] G. E. Chatrian, E. Lettich, and P. L. Nelson, "Ten Percent Electrode System for Topographic Studies of Spontaneous and Evoked EEG Activities," *Am. J. EEG Technol.*, vol. 25, no. 2, pp. 83–92, Jun. 1985.

[163] J. P. Borst, N. A. Taatgen, and H. van Rijn, "The problem state: A cognitive bottleneck in multitasking.," *J. Exp. Psychol. Learn. Mem. Cogn.*, vol. 36, no. 2, pp. 363–382, 2010.

[164] S. Koelstra *et al.*, "DEAP: A database for emotion analysis; Using physiological signals," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 18–31, 2012.

[165] N. V Chawla, K. W. Bowyer, and L. O. Hall, "SMOTE : Synthetic Minority Over-sampling Technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, 2002.

[166] V. M. Joshi and R. B.Ghongade, "Optimal Number of Electrode Selection for EEG Based Emotion Recognition using Linear Formulation of Differential Entropy," *Biomed. Pharmacol. J.*, vol. 13, no. 02, pp. 645–653, Jun. 2020.

[167] N. Sevdalis, K. V Petrides, and N. Harvey, "Trait emotional intelligence and decision-related emotions.," *Pers. Individ. Dif.*, vol. 42, no. 7, pp. 1347–1358, 2007.

[168] K. Takahashi, "Remarks on emotion recognition from multi-modal bio-potential signals," *Proc. IEEE Int. Conf. Ind. Technol.*, vol. 3, pp. 1138–1143, 2004.

[169] M. B. Er and H. Çig, "A new approach to recognition of human emotions using brain signals and music stimuli," vol. 175, 2021.

[170] F. Pedregosa, R. Weiss, and M. Brucher, "Scikit-learn : Machine Learning in Python," vol. 12, pp. 2825–2830, 2011.

[171] N. K. Manaswi, "Understanding and Working with Keras," in *Deep Learning with Applications Using Python*, Berkeley, CA: Apress, 2018, pp. 31–43.

[172] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, pp. 1–15, 2015.

[173] S. Suthaharan, *Machine Learning Models and Algorithms for Big Data Classification*, vol.

36. Boston, MA: Springer US, 2016.

[174] X. Jin, A. Xu, R. Bie, and P. Guo, “Machine learning techniques and chi-square feature selection for cancer classification using SAGE gene expression profiles,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 3916 LNBI, pp. 106–115, 2006.

[175] T. Costa, E. Rognoni, and D. Galati, “EEG phase synchronization during emotional response to positive and negative film stimuli,” *Neurosci. Lett.*, vol. 406, no. 3, pp. 159–164, 2006.

[176] N. Zhuang, Y. Zeng, L. Tong, C. Zhang, H. Zhang, and B. Yan, “Emotion Recognition from EEG Signals Using Multidimensional Information in EMD Domain,” *Biomed Res. Int.*, vol. 2017, 2017.

ANNEXES

A. Personal authorizations and information about participants.
B. A segment of the Stock-Emotion dataset.
C. A segment of the Poker game dataset.
D. Protocol for dataset obtention.
E. Features.
F. Periodograms.
G. Frequency bands.
H. Classification results