SAPIENZA
UNIVERSITÀ DI ROMA

Beyond Brainwaves: Exploring Emotions, Identity, and Motor Imagery Through EEG-based BCI.

Faculty of Information Engineering, Informatics, and Statistics
Doctor of Philosophy in Computer Science (XXXVII cycle)

Angelo Di Mambro
ID number 1664996

Advisor
Prof. Danilo Avola

Co-Advisor
Dr. Daniele Pannone

Academic Year 2023/2024

Thesis defended on 15 January 2025
in front of a Board of Examiners composed by:

Simone Carderara (chairman)

Antonio Piccinno

Giovanna Varni

Beyond Brainwaves: Exploring Emotions, Identity, and Motor Imagery Through EEG-based BCI.

PhD thesis. Sapienza University of Rome

This thesis has been typeset by LaTeX and the Sapthesis class.

Author's email: angelo.dimambro@uniroma1.it

To that loving child who sought his dreams.

Acknowledgments

First and foremost, I sincerely thank my advisor, Danilo Avola, and my co-advisor, Daniele Pannone, for their guidance and personal support throughout these challenging years. It has been a long, tiring journey, but one worth doing. To my parents, my sister, and my entire family, I am deeply grateful to you for being my lighthouse throughout these years, especially when the ship seemed to drift off course, you never made me feel alone. To all my friends, my second family, who are always there to lend me a hand or brighten even my darkest days with a smile, you are my anchor. I would never have been able to navigate this journey without your support.

Abstract

An Electroencephalography (EEG)-based Brain-Computer Interface (BCI) is a system able to connect the human brain and external devices by analyzing EEG signals, translating the brain activity patterns into instructions for an interactive application. Initially, EEG-based BCI solutions were developed for medical purposes in clinical and rehabilitation applications, primarily to assist patients in regaining normal life functions. Beyond this original goal, these systems have also gained importance in non-medical fields such as cybersecurity and neuroscience applications. On this account, this thesis shows how the EEG signal can be directly exploited for solving person biometric identification, emotion recognition, and limbs activation classification tasks. Since most existing EEG-based biometric systems don't exploit the time-frequency information of EEG signals, this thesis introduces a novel identification system using graph representations, where nodes represent EEG channels signals and edges denote the Functional Connectivity (FC) measure between pairs of channels. The model, based on Graph Convolutional Neural Networks (GCNNs), integrates spatio-temporal and functional features, capturing local and global brain activity. Tested on PhysioNet and Multi-subject, Multi-session, and Multi-task Database for investigation of EEG Commonality and Variability (M3CV) datasets, the method demonstrated strong generalization across various human states (resting and active) and outperformed State-Of-the-Art (SOTA) EEG biometrics techniques in specific tasks. Regarding emotion recognition, has been proposed an innovative framework, namely Empátheia, able to encode EEG signals as compact images, preserving the original spatio-temporal information, and recognizing the associated emotion. Using the Processing and transfeR of Interaction States and Mappings through Image-based eNcoding (PRISMIN) framework, the original EEG signals are encoded as images, or atlases, following a spatio-temporal layout. Then, different deep learning models have been designed and tuned to classify the emotions captured in the produced atlases. Tests on the SJTU Emotion EEG Dataset (SEED) dataset showed high performance and efficient data representation, suggesting new possibilities for EEG-based emotion analysis. Instead, a novel multi-stream 1D Convolutional Neural Network (CNN) architecture is proposed for limbs activation classification. This method processes EEG signals through four convolutional streams with varying kernel sizes to capture information at different time scales. The resulting features are combined and fed to a dense classifier to determine limbs movement. Experiments on the PhysioNet EEG dataset showed that this model outperforms existing methods in both cross-subject and intra-subject settings.

Contents

Chapter 1

Introduction

In recent years, advancements in technology and machine learning algorithms have increased interest in EEG-based applications. EEG can facilitate continuous monitoring of fluctuations in human cognitive states under different tasks, finding developments in both the healthcare field and in research across various application domains areas. The promising future of EEG-based technologies has encouraged the scientific community to study brain activities, thus opening up new directions for EEG research. General researchers from a wide range of areas of expertise, including computer science, psychology, education, data analysis, engineering, gaming, and health, have explored this technology from different points of view, finding related content and research trends relevant to their respective areas. Since classical Human-Computer Interaction (HCI) problems are mostly restricted to vision-based or manual interfaces with the majority of other designs not being extensively adopted [1], exploiting EEG unique and peculiar characteristics, represent an alternative or support solution to these tasks. In fact, EEG is recognised to offer the potential for an approach that uses intuitive and natural human mechanisms of cognitive processing, by means of BCIs, to facilitate HCI or enabling new solutions in a variety of tasks such as healthcare, where EEG signal has been utilised to facilitate effective control of a robot for motor imagery classification [2], epileptic seizure detection [3] [4], depression recognition [5], vigilance estimation [6], sleep disorders [7], mental fatigue detection [8], and gaming interaction [1]. In biometric identification, which utilises EEG-based systems to decide whether to reject or accept the claiming identity of a subject [9], and in emotion recognition, understanding human emotions using specific brain signals evoked by certain emotional states [10]. The reason for choosing this signal is threefold. First, human privacy is naturally preserved because sensitive information is not collected through the EEG medium. Second, EEG is a biosignal and can't be affected by spoofing attacks, granting a higher level of security. Third, because EEG grants a very high temporal resolution and can take readings every millisecond, meaning it can record the brain's activity in real-time. Among the others, three of the most exciting fields still barely explored through EEG are person biometric identification, emotion recognition, and MI limb activation classification.

An individual's identity constitutes a distinctive combination of characteristics, experiences, physical traits, and personality attributes that differentiate them from all other individuals. This identity encompasses both objective elements, such

as name, age, gender, and physical attributes, and subjective elements, including beliefs, experiences, emotions, and personal relationships. In a digital society, the need to have a safe and reliable method of identification of the person is crucial in a multiplicity of fields: in cybersecurity, for granting a reliable authentication method for devices and network access, for protecting sensitive data and confidential information, for preventing online fraud and unauthorized access; in finance, for allowing identity verification for financial transactions, for avoiding credit card fraud and online transactions; in passport and visa control, for airport and transportation security and screening; and in national and public security for supporting the border control and immigration, improving detection of criminals and terrorists and identification of missing or unidentified people. Moreover, confirming human identity becomes even more crucial as many human activities shift into cyberspace, continuously exposed to dangers and fraud attempts that require new security standards to prevent identity theft or spoofing. As pointed out in [11], for a reliable biometric have been identified seven criteria to be used when assessing the suitability of any trait for use in biometric identification, namely *Universality, Uniqueness, Permanence, Collectability, Performance, Acceptability, Circumvention.* Based on the research of Jain et al. [11]. Currently, there is no gold standard for biometric identification matching these criteria, even though different methods have been proposed, including fingerprint recognition [12], iris recognition [13], palm vein recognition [14], face recognition [15]. A technology that has been garnering research interest in recent years in the security field is EEG. This technology allows for the measurement of electrical potentials generated by the human brain by applying electrodes integrated into a portable and non-invasive headset to the subject's scalp. Subsequently, with the use of a BCI, it is possible to record the brain waves and send the signals to a machine able to perform a specific task [16, 17], e.g., EEG signal classification [18], multimodal interaction [19–21], and biometric identification. This kind of signal offers unique advantages compared to other biometric modalities. Unlike the classic biometric methods such as fingerprints and palm recognition, EEGs are inherently linked to vital biological functions, making it considerably more challenging to circumvent this type of physiological-based biometrics. In addition, different from methods such as face and fingerprint recognition, in which the sample can be faked with photos or replicas, EEG requires the real-time and continuous brain activity of the individual. EEG can potentially provide continuous authentication [22], meaning it can continuously verify a user's identity while they are interacting with a system and can be recorded passively without the subject's active participation or awareness. This last aspect can be helpful in scenarios where user consent or cooperation might be challenging. Given its superior temporal resolution, usually in the order of milliseconds, and given its complex dynamic characteristics, it is regarded as a strong biological signal for biometric applications, alongside others like Electromyography (EMG) and Electrocardiography (ECG). Although EEG offers a number of advantages over traditional identification methods, research on EEG-BCI-based identification systems is still in its early stages. There are several approaches to utilize EEG as a biometric modality: some rely on the use of Event-Related Potentials (ERPs) [23], variations in electrical potential resulting from a visual, somesthetic, or auditory stimulus. These can be cumbersome as they require the user to be subjected to stimuli such as Visual-Evoked Potential (VEP) [24]

or Auditory-Evoked Potential (AEP) [25], also necessitating extra equipment like headphones and screens, in addition to BCIs. Another limitation of ERPs analysis concerns the fact that it ignores non-phase-locked signals [26]. Another line of research that has recently gained traction is the one that focuses on resting-state EEG [27], which examines the brain's neural activity in the absence of specific stimuli or tasks. While resting-state offers great flexibility, it is challenging for individuals to maintain. Although it does not require active involvement from the user, unlike ERPs, users may experience intrusive thoughts or emotions that make it difficult to remain in a true resting state during the identification process. To overcome this problem, current research is focusing on a system capable of operating independently of a specific state or stimulus. A significant limitation of current systems is that they usually represent EEG signals using a Fourier-based power analysis, which ignores temporal information, while a Time-frequency analysis can better characterize the oscillations contained in the EEG data [26].

Emotions are a key human trait that can be defined as a biological state associated with neurophysiological changes that are linked with thoughts, feelings, behavioural responses, pleasure, or displeasure sensations [28] and can affect almost every aspect of our existence, such as, among others, social interactions, relational life, work productivity, and even HCI. Although different communication channels can be used to express emotions, e.g., facial expressions, voice pitch, and posture, it can be not easy to understand these cues that convey additional information about a person [29]. As a consequence, emotion recognition through automatic approaches can be beneficial in diverse, relevant scenarios. A first example regards the diagnosis of depressive states or post-traumatic stress disorders (PTSD) [30] to identify if a patient is experiencing pain during a treatment [31], another interesting case study concerns the diagnosis of Parkinson's Disease (PD) [32] to understand if a patient exhibits emotional impairments when emotionally elicited, a last example is the detection of fake emotions during an interrogation in court [33]. Nowadays, this type of technology is increasingly being used in a wide range of application fields. As a matter of fact, many works in the current literature try to address the emotion recognition task by exploiting facial traits [34, 35], body movements [36, 37], speech [38], multimodal approaches [39–41], or even more complex data such as brain electrical activity [9, 42, 43]. Concerning the latter, it can be measured through EEG, which extracts brainwaves through the use of surface electrodes. As extensively described in Sec. 2.4, five different waves are retrieved with the EEG, i.e., delta, $\delta \in [1.5Hz - 4Hz]$ [44], theta, $\theta \in [5Hz - 8Hz]$ [45], alpha, $\alpha \in [9Hz - 14Hz]$ [46], beta, $\beta \in [15Hz - 40Hz]$ [47], and gamma, $\gamma \in [25Hz - 140Hz]$ [48]. These waves respond to specific activities, including daydreaming or active thought, via working memory and attention. What is more, brainwaves are also related to emotion processing, where specific patterns in high-frequency bands are associated with positive, neutral, and negative feelings through time-frequency analysis [49]. As can be observed in Figure 1.1, positive emotions show an increment in energy for beta and gamma frequency bands, whereas neutral and negative emotions present decreased beta and gamma energy. While the neural patterns of negative and neutral emotions have similar patterns in beta and gamma bands, the latter have higher energy of alpha oscillations. Brainwaves are commonly acquired through BCIs and, unlike other types of digital information, collecting EEG data can quickly become

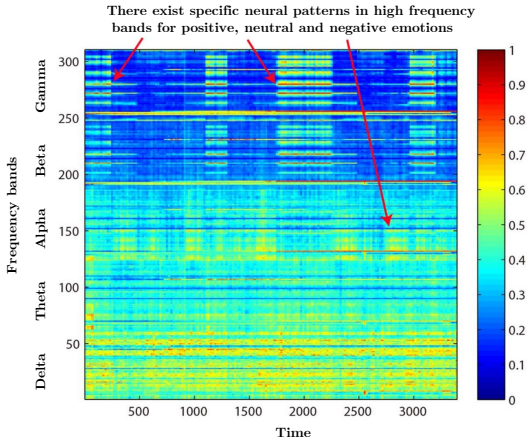


Figure 1.1. Emotions captured via EEG signals. The differential entropy energy feature [50] (scaled in [0,1]) is represented for each frequency band (in Hz) over time (in ms), highlighting specific patterns in higher frequency ranges corresponding to specific emotions. Adapted from [49].

complex and cumbersome, depending on the number of channels and the range of emotions that need to be monitored. As for other digital resources, compression techniques can support several critical aspects, including the reduction of storage space, which assumes particular importance in contexts where storage is expensive or limited, such as in mobile devices, drones, robots, and, in general, embedded systems; transmission efficiency, which becomes crucial in environments where bandwidth is bounded or communication channels are constrained, such as in long-distance monitoring; compatibility and scalability, where data compression allows us for easy scaling or conversion of data to different formats that can be compatible with various devices, applications, or protocols; among other advantages.

Since BCI systems allow people to send messages or commands to an electronic device only by means of brain activity instead of muscular activity, they can provide an alternative communication and control channel for people with limited motor function to improve the quality of their lives. One kind of EEG-based BCI system is based on the recording and classification of circumscribed and transient EEG changes in association with the imagination of different types of movements. MI can modify the neuronal activity in the primary sensorimotor areas in a very similar way as observable with a real executed movement, so as a result it can serve to generate self-induced variations of the EEG [51] [52], it may be seen as mental rehearsal of a motor act without any overt motor output. MI, it is undoubtedly one of the most fascinating and valuable fields for EEG-based interfaces. Thanks to EEG devices, it is possible to design low-cost, high-performance systems capable of significantly improving the quality of life for a wide range of individuals with motor dysfunctions [53–56]. It is well-known that methodologies addressing MI focus on interpreting predominant brain areas that are activated when a subject mentally simulates movements or actions without actual physical execution [57–59]. Figure 1.2 shows a clear Event-Related Desynchronization (ERD, reduction in power) pattern in alpha and beta frequency bands during hand MI in the contra-lateral motor areas of the cortex, while that of the resting state diagram does not. The nature of these

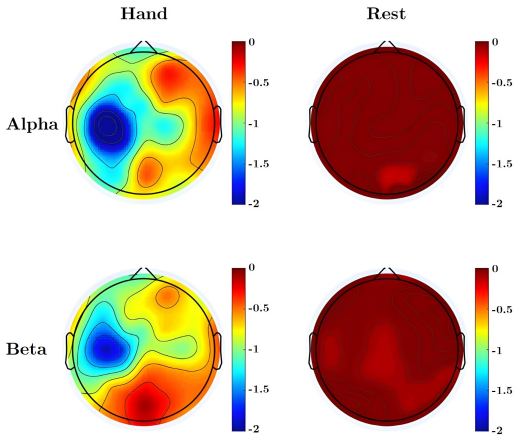


Figure 1.2. Averaged topographical distribution of power of 25 subjects in the upper alpha band (10–12 Hz) and upper beta band (23–25 Hz) during hand MI. Blue indicates ERD. Adapted from [63].

imagined activities depends on their intended goals, which can range from simple device interactions and performing rehabilitation exercises to piloting mechanical prostheses [60–62]. Unsurprisingly, recent years have seen remarkable results in complex tasks, thanks to the ever-increasing use of modern BCIs. These interfaces allow for the acquisition of EEG signals with increasingly accurate and coherent spatio-temporal content. Determining the motor intentions of an individual through the analysis of EEGs is a challenging task that concurrently holds considerable potential in aiding subjects with motor dysfunctions. Moreover, thanks to the recent advances in artificial intelligence models and EEG acquisition devices, such analysis can be carried out with ever higher accuracy. The latter aspect is of great importance, since the EEG analysis of subjects whose mental efforts are focused on moving limbs is frequently used for crucial tasks, including the control of interactive interfaces and prosthetic devices. Since 1996, the Graz BCI system has been proposed to discriminate between three simple limb motor imagery tasks (left hand, right hand, right foot), designing a BCI able to help patients with severe motor impairment (e.g., amyotrophic lateral sclerosis) to interact with their environment through voluntary modification of their EEG exploiting band power estimations from three bipolar EEG channels [64]. To confirm whether motor imagery could be available for patients with severe motor impairment, a tetraplegic patient had learned to operate an electrical-driven hand orthosis by discrimination of two mental states to restore the hand grasp function [65]. With respect to motor imagery of simple limb movement, several parts of limbs like the hand (forearm, postbrachium) and foot (shank, thigh) are involved in compound limb movement imagination, which may activate the neuron oscillation in multiple functional areas of cerebral cortex and, at the same time, also can satisfy the requirements of multiple instructions output of control information in MI-based BCI systems. The research on compound limb motor imagery has great significance for limb function rehabilitation for patients suffering from severe motor injury.

For person biometric identification, motivated by these observations and existing literature, this thesis introduces a novel identification system capable of discerning

a subject based on EEG signals across various human states. In detail, given a person multi-channel EEG signal, the proposed approach is able to represent the pre-processed signals as a complete graph where each node represents an EEG electrode, as the signal time-frequency spectrogram and an edge between two nodes represents the FC across that electrodes pair, expressed as the Phase-Locking Value (PLV). Finally, the provided graph that embodies local and global information of the subject's brain, both in the spatial and temporal domain, is used as input for a Graph Neural Network (GNN) classifying the subject identity. An extensive experimental study is conducted on the PhysioNet and M3CV public benchmark datasets, assessing the generalization capacity of the designed method across diverse human states, including resting states (eye-open and eye-closed) and active states during the execution of four distinct tasks, as well as a mixed protocol. The proposed system is compared with the most advanced EEG biometrics techniques, showing that the proposed method overcomes the state-of-the-art in specific tasks.

To address the emotion recognition problem, is introduced the original Empátheia [1] system, which encodes EEG data into images, referred to as atlases, before classifying the underlying emotion. In detail, the proposed approach pre-processes multichannel EEG signals and generates spatio-temporal atlases using an encoder based on Processing and transfeR of Interaction States and Mappings through Image-based eNcoding (PRISMIN) framework [66]. This framework compresses brain signals into a coarse visual representation, i.e., an image. Then, the system uses a deep learning-based pipeline as a classifier that recognizes emotions. Specifically, the architecture is composed of branches based on convolutional, recurrent, and transformer models designed and tuned to capture the spatial and temporal aspects of an emotion represented by the atlas. Extensive experiments were conducted for both the encoder and classifier. In particular, two encoding types, i.e., short-rainbow and grey-scale, and four different models, i.e., one based exclusively on a CNN, two based on mixtures of CNN and RNN, and a last based on a transformer, were tested to find an effective emotion recognition method. The Empátheia system was evaluated on the SEED dataset [49,50]. The proposed approach significantly reduced the dataset size, thus enabling the implementation of less computationally intensive models. In addition, it allows us to train faster while retaining high performance on the emotion recognition task.

Instead, for MI limb activation classification, is proposed an innovative architecture based on a modular multi-stream 1D CNN. The input EEG signal is processed by four convolutional streams, which differ in the size of convolutional kernels, thus allowing the extraction of information at different time scales. The resulting 1D feature maps are then fused together and provided to a dense classifier to identify which limb the subject intended to move. A comprehensive experimental study has been performed on the PhysioNet EEG motor movement/imagery dataset, the reference data collection for MI application field, for two reasons: to investigate if shallow models with more streams achieve better results with respect to deeper models with fewer streams with this specific type of data, and to show how the proposed model overcomes the key works of the current literature in both cross-subject and intra-subject analysis.

[1] from ancient Greek *en-*, "inside", and *pathos*, "sentiment, feeling". Today, it is translated as "Empathy".

1.1 State of the Art

In this section, the literature on the current state of the art regarding methods for EEG biometric identification, emotion recognition, and limb activation is discussed. In particular, the current literature is reviewed from three perspectives, namely, tasks and states (for biometric identification only), extracted features, and used classifiers.

1.1.1 EEG Person identification

In this section, the works in the current SOTA regarding methods for EEG biometric identification are discussed. In particular, the current literature is reviewed from three perspectives, namely, tasks and states, used features, and used classifiers.

Tasks and States

Authentication techniques relying on external stimuli-based protocols exploit the inherent uniqueness of human brain potentials generated in response to such stimuli, which are called ERPs. These stimuli are typically auditory (AEPs) or visually (VEPs), and as reported in [67], the latter are the most used. Different eliciting protocols were proposed for VEPs stimulation in literature. In [68] VEPs are elicited showing images of celebrities, food or words. Zhao et al. [69] uses three frame-based black-and-white alternating stimuli modulated by a flash signal, a periodic flicker signal, and a pseudorandom m-sequence. In [70] are used two different eliciting protocols: in the first, 8 images displaying geometric shapes followed by a black screen on an LCD screen are shown; in the other, 62 images displaying capital letter characters, lowercase letter characters, and digits ranging from 0 to 9 followed by a black screen are shown. Gui et al. [71] collected EEG signals triggering VEPs while the participants were asked to silently read an unconnected list of texts, which included 75 words. Steady-State Visual-Evoked Potentials (SSVEPs) are potentials generated in the brain in response to visual stimuli that flicker at a specific frequency, these potentials are often used because of their excellent signal-to-noise ratio and relative immunity to artifacts [72]. In [73], they used a device employed to elicit SSVEPs consisting of a square array of 9 green LEDs, whose flickering frequency can be manually tuned at four different elicitation frequencies, namely at 6, 12, 18, and 24Hz. In [74], subjects were exposed to flashing images on-screen at five stimulus frequencies of 6.67 Hz, 7.50 Hz, 8.57 Hz, 10 Hz, and 12 Hz. Min et al. [75] designed an eliciting protocol where SSVEPs signals were triggered by a flickering grid-shaped visual array based on Korean letters. However, users subject to this kind of visual stimuli could suffer serious fatigue that may induce discomfort, health hazards and deterioration of system performance [76]. For this reason, another VEP variant called Steady-State Motion Visual-Evoked Potentials (SSMVEPs), was proposed. SSMVEPs slightly differ from SSVEPs; in fact, these brain potentials can be triggered, showing movement patterns in contraction and expansion, resulting in less discomfort to subjects compared to flickering images. Gao et al. [74] proposes an elicit protocol in which black and white rings alternately contract and expand were shown on a screen at eight different stimulus frequencies of 6 Hz, 8 Hz, 10 Hz, 12

Hz, 14 Hz, 16 Hz, 18 Hz, and 20 Hz. Concerning AEPs, Seha et al. [77, 78] proposes a protocol where 4 auditory tones at 500, 1000, 2000, and 4000 Hz modulated by sinusoidal at 40, 44, 48, 54 Hz are presented simultaneously to the participant's ears to elicit the brain response. Since ERPs, VEPs, and AEPs signals are characterized by being dependent on a time-locked external stimulus that must be repeated in a controlled environment and require the use of additional equipment, are difficult to use in real scenarios. For this reason, has emerged another line of research that solely utilizes EEG signals acquired from subjects in resting-state, without the need for external stimuli. Various studies [79–83] have delved into resting-state biometric authentication. However, a fundamental issue arises from the fact that users cannot consciously choose to remain in a resting state. This renders it impractical in real-world scenarios where users may engage in mental or physical tasks that prevent them from maintaining the resting state, resulting in a general degradation of the identification system. Motor Imagery (MI) is another protocol that does not need external stimuli, which involves the user imagining the movement of a body part without physically performing it. Das et al. [84] protocol is designed to collect EEG signals corresponding to a series of right-arm, left-arm, right-leg and left-leg imaginary movements. Four different images are employed as stimuli for eliciting the desired responses and are sequentially shown in random order on an LCD monitor. Chuang et al. [85] uses a different protocol Motor Imagery (MI) consisting of different mental tasks in which subjects were asked to imagine performing different tasks like moving fingers, repeating a specific sport motion, singing a song, etc. Compared to resting protocols, Motor Imagery (MI) protocols tend to produce signals with less noise. On the downside, research works have reported that they can cause mental fatigue in users [85], which can affect the accuracy of the user authentication process.

The nature of EEG signals is to be sensitive to all the changes caused by imaginary mental tasks, physical activities, mood fluctuations, and sudden emotions. Therefore, an identification system should be robust to the entirety of these variables, independent of specific stimuli and/or states. Indeed, several studies [86,87], including the present one, focus on the development of identification systems that can generalize across diverse human states.

Features

The majority of works in the literature apply feature extraction techniques from the EEG signal recorded by electrodes. Two distinct categories of features can be identified: uni-variate and bi-variate. Uni-variate features are extracted from individual electrodes without establishing relationships between signals recorded from different electrodes. Several studies used uni-variate features as Power Spectral Density (PSD) [77, 82, 88], which represents the power distribution of EEG series in the frequency domain, Short-Time Fourier Transform (STFT) [72] and Wavelet (WT) [71, 89], used to convert the input signal from the time domain into the time-frequency domain, Mel-frequency Cepstral Coefficients (MFCC) [73, 77, 90], deriving a robust characterization of the temporal and spectral properties of the EEG signal, and entropy measures [91] (e.g., fuzzy entropy) which estimate signal complexity. Features based on Gaussian Filtering (GF) [77], Autoregressive (AR) [73, 90], and Chebyshev type I infinite impulse Response (IRR) [69] were also proposed for EEG

biometric.

A critical problem with uni-variate features lies in their susceptibility to alterations in EEG amplitudes, often stemming from inherent physiological or psychological factors like circadian rhythms [92, 93], substance assumption [94], or variances in signal acquisition methodologies [95]. In contrast, bi-variate features also capture information regarding the interconnection of signals extracted from distinct electrodes, establishing relationships between different parts of the brain, and according to La Rocca et al. [82], are less sensitive to the amplitude changes of EEG signals compared to uni-variate features. Such measures are called brain connectivity measures, and there are several (FC, Effective Connectivity (EC)) based on the specific measured correlated brain activities. According to Fingelkurts et al. [96], FC is the temporal correlation between spatially remote neurophysiological events, expressed as deviation from statistical independence (temporal correlation) across these events in distributed neuronal groups and areas. Friston et al. [97] defines FC as statistical dependencies among remote neurophysiological events and is an observable phenomenon that can be quantified with measures of statistical dependencies, such as correlations, coherence or transfer entropy. FC is an effective tool to estimate the coordinated activation over distinct brain regions in the EEG domain, and several connectivity metrics have been defined in the literature to measure it from the EEG signal. La Rocca et al. [82] investigate Spectral Coherence (COH), a measure of frequency-specific correlation between two EEG signals, providing information on the phase relationship between the signals in different frequency bands. A 100% Correct Recognition Rate (CRR) on 108 subjects in resting state was achieved using a match-score fusion approach of the COH features. Kim et al. [81] derive a Phase-Lag Index (PLI)-based FC matrix for different brain regions, concatenating in the eigenvectors extracted from each matrix single feature vector. The PLI estimates pair-wise phase synchrony between two time-series signals, captures the intrinsic phase coupling and reflects the asymmetry of the distribution of the phase difference between two signals ranging from 0 to 1. PLV is a measure of phase synchronization between two or more time series in the frequency domain tested in [91, 98]. The obtained results indicate that PLV is a stable, lightweight, promising metric of eeg biometric. Behrouzi et al. [98] compares PLV to Covariance (COV) in the same work, another FC metric that computes the covariance matrix to measure the linear correlation of two electrode signals computed in the time domain. Comparable results to PLV features are obtained in terms of accuracy, space and computational time. In [91, 99] Cross-Correlation (COR) is used to measure the linear relationship between EEG signals from different electrodes using a correlation coefficient, such as Pearson's correlation, to compute the strength and direction of the relationship. Riera et al. [99] applied Mutual Information (MI), a metric that can measure the nonlinear as well as linear dependence of two variables, to create a multi-feature set of uni-variate (Fourier Transform (FT), AR and bi-variate features (MI, COH, COR). Wang et al. [86] represented the EEG signals (electrodes) as a brain network graph, concatenating in a single feature vector connectivity (e.g., COR, PLV), node centrality (e.g., node degree, page rank) and global network (e.g., Transitivity, Modularity) features from each node. In addition to FC, effective connectivity corresponds to the intuitive notion of coupling or directed causal influence, and refers explicitly to the influence that one neural system exerts over another [97]. Min et al. [75] measures

effective connectivity using Granger Causality (GC), this metric provides information about directional interactions between brain regions, identifying causal relationships of two simultaneously measured signals with directional information. The main issues observed in previous works are related to the use of ERPs, unsuitable for real scenarios [75], the exclusive use of resting state [81, 82, 99] or single human states, ignoring the real-world scenario of mixed states [86, 98]. In the presented work, the feature extraction differs from the previously reported methods. Specifically, to maximize the utilization of information captured by the electrodes, local features are combined with global features into a graph structure that, for the first time, integrates both the local spatio-temporal information of the individual channel, represented by the nodes as the STFT spectrogram of each channel, and the global brain FC, represented by the edges as the PLV between each pair of channels.

Classifiers

Most of the research on EEG employs on linear discriminators such as Support Vector Machine (SVM) [75, 79, 80, 90], K-Nearest Neighbors (KNN) [72, 79], Linear Discriminant Analysis (LDA) [77–79, 99] or other classifiers that measures features vectors distance, e.g., Mahalanobis [82, 86], Manhattan [73] and Cosine [70]. The main limitation for linear classifiers is that they tend to perform poorly when the type of data varies over time and is not stable, as is the case with EEG signals. For this reason, more complex non-linear classifiers that leverage deep learning (e.g., CNNs, Recurrent Neural Networks (RNNs)) have been proposed to address this issue. Regarding CNN, different works [84, 88, 90, 100] use a multiple layers CNN architecture followed by a dense layer to extract most discriminative features and classify them for people identification or authentication purposes. In particular, CNNs are used to learn a deep representation directly from the raw EEG time-series, without extracting any handcrafted feature. The main weakness for these work belongs to the classifiers training, in fact they rely on EEG data generated using specific resting or ERP protocols, making the proposed systems prone to changes in the human state. Wang et al. [91] uses a different kind of graph convolution, based on Chebyshev polynomials. In particular, they represent EEG signal as a graph based on within-frequency and cross-frequency FC estimates, and the use of GCNNs to automatically capture deep intrinsic structural representations from the EEG graphs for person identification. Concerning Deep Neural Networks (DNNs) architectures, they are used to classify the person's identity given hand-crafted features extracted from the raw signal. For example, the scheme used by Kim et al. [81] explores the use of Functional Network, a deep model used for person identification that exploits local and global concatenated features extracted from FC matrices computed on five brain quadrants. Similarly, the work from Gui et al. [71] focuses on a Feed-Forward Neural Network used to classify spatio-temporal features based on WT from VEPs signals. Concerning RNN-based methods, they tend to focus on spatial and temporal information, which can be naturally captured by recurrent architectures. Sun et al. [101] and Das et al. [83] propose similar systems consisting of a Conv-Long-Short Term Memory (LSTM) architecture composed by multiple convolutional 1D layers followed by two LSTM layers. Given a raw EEG, the proposed system firstly extracts deep features using CNNs, then captures the inherent spatio-temporal information

using the recurrent layers for the final user identification. In relation to Encoder-Decoder approaches, Du et al. [87] compares a proposed transformer-based approach for the EEG person identification task that extracts features in the temporal and spatial domains using a self-attention mechanism, against other Transformer-based approaches as Lite Transformer [102] and Evolving Attention (EA) Transformer [103]. Behrouzi et al. [98] integrates variational Auto Encoders (AEs) with GNN capturing dynamics from EEG time-series to learn graph embedding from EEG FC input both on resting and active states. In summary, current methods address the EEG identification task using deep learning approaches based on CNNs, RNNs, graph models, AEs, or a combination of these architectures. Building upon this foundation, this thesis introduces a model that strategically harnesses GCNN with a custom FC graph structure to capture both single-channel spatio-temporal information and cross-channel FC.

1.1.2 EEG Emotion Recognition

In this section, the current works in SOTA regarding methods for EEG emotion recognition are discussed. In particular, the current literature is reviewed from two perspectives, namely, the adopted features classifiers.

Features

An important aspect of EEG emotion recognition is the feature extraction from brain signals, which can affect the classification accuracy. After signal pre-processing, for instance, downsampling [18] and band-pass frequency filtering [104], EEG features can be divided into single-channel and multichannel features. The former class was generally the most common choice in previous work due to its proven effectiveness [105], and includes, among others, PSD [106], Differential Entropy (DE) [50], and Wavelet Features [107]. The second class has instead become the preferred option in recent years, especially with the evolution of deep learning approaches for EEG emotion recognition, resulting in various solutions available in the literature exploiting CNN, RNN, or graph-based architectures [108]. Miao et al. [109] employ a band-pass filter to process the EEG signal for each spatial channel across five frequency bands associated with Delta, Theta, Alpha, Beta, and Gamma rhythms. Subsequently, the DE feature is computed for the EEG signal on each spatial channel. Liu et al. [110] leverage a band-pass filter to isolate standard frequency bands from EEG signals, specifically including the ranges of (1-7 Hz), (8-13 Hz), (14-30 Hz), and (30-45 Hz). Subsequently, the data are transformed into two-dimensional features, from which Pearson Correlation Coefficient (PCC), Principal Component Analysis (PCA), and statistical descriptors such as variance, mean, kurtosis, and skewness are extracted for each channel to train a CNN. In their study, Li et al. [111] use two directional RNN modules capable of navigating the EEG spatial domains both horizontally and vertically, effectively forming a comprehensive linkage of the electrode locations. This method allows the extraction of high-level features for each EEG electrode, while concurrently aiding in the formation of features corresponding to various brain regions. In [49], a 256-point Short-Time Fourier Transform with a non-overlapped Hanning window of 1s is used to extract five frequency bands of EEG signals. Then

DE features are calculated for each frequency band. In [112], the original EEG is firstly decomposed into five constituent sub-bands, from which features were then extracted using time, frequency and nonlinear analysis.

Classifiers

Regarding CNN-based approaches, they tend to focus on spatial information derived from multiple EEG channels. Liu et al. [110], for example, utilize a Butterworth band-pass filter on the EEG channels and restructure the data for integration into a specialized deep neural network. This network includes a CNN, a sparse autoencoder (SAE), and a deep neural network (DNN), each trained independently to improve convergence speed. Liu et al. [113] incorporate spatial data into input signals via an attention mechanism and make use of a pre-trained convolutional capsule network for feature extraction alongside an additional double-layer capsule network. Li et al. [114] present an alternative example emphasizing the frontal lobe and employing the Papez circuit theory. They implement a frontal lobe double dueling DQN (FLD3QN) methodology, rooted in reinforcement learning, utilizing EEG channels combined with a bifrontal lobe residual CNN (BiFRCNN) to facilitate emotion recognition. Miao et al. [109] introduce a 3D deep residual learning model designed to examine EEG signals spanning various frequency bands. They apply group sparse regression to optimally select frequency bands and employ a 3D deep residual network to classify features. Finally, Hu et al. [115] present a scaling layer within their convolutional network designed to derive spectrogram-like features from EEG data. This layer employs diverse convolutional kernels to detect patterns at multiple scales, thereby removing the requirement for alternative feature extraction techniques such as DE.

Concerning RNN-based methods, they tend to focus on spatial and temporal information, which can be naturally captured by recurrent architectures. For instance, the study by Li et al. [116] explores the variations in emotional expression across the brain's left and right hemispheres. They employ four distinct RNNs within two specific brain areas to examine spatial correlations. Moreover, they develop a subnetwork to merge these hemispheres, improving the extraction of features for emotion recognition. In a similar study, Zhang et al. [117] employ a multidirectional RNN to examine long-range contextual signals in EEG data to detect spatial changes linked to human emotions. They apply projection matrices to both spatial and temporal states to locate areas rich in emotional content. The research conducted by Guo et al. [118] utilizes domain adaptation to diminish the variability in EEG signals across sessions by creating a spatio-temporal feature extractor. These features are subsequently aligned for emotion classification. Yang et al. [119] employ a blend of LSTM and CNN networks to study spatio-temporal features within raw EEG signals. The LSTM is used for capturing contextual information, whereas the CNN is tasked with detecting inter-channel correlations through a 2D signal format. Finally, In the work by Du et al. [120], they present an attention-based LSTM incorporating a domain discriminator (ATDD-LSTM) for extracting spatial features across EEG channels. The method is centered on capturing nonlinear connections between electrodes to enhance the selection of EEG channels and reduce variations in features between various subjects and sessions.

Graph-based methods make use of the inherent configuration of BCI devices to create graph-like structures. Liu et al. [121] propose a global-to-local feature aggregation network (GLFANet) that analyzes spatial and frequency domain features in EEG channels using topological graphs. This network incorporates both global learners, with graph convolutional blocks, and local learners, using convolutional blocks, for EEG signal feature extraction. Zhong et al. [122] emphasize left and right hemisphere coupling in emotion recognition through a regularized graph neural network (RGNN), examining both local and global EEG channel relationships. They highlight the informative nature of pre-frontal, parietal, and occipital regions. Song et al. [123] implement a dynamic graph convolution neural network (DGCNN) to capture multichannel EEG features, learning an adjacency matrix that represents EEG channel relationships for feature discrimination. In another approach, Zhou et al. [124] develop a progressive graph convolution network (PGCN) to identify both coarse and fine-grained emotional features, utilizing a dual-graph module to encapsulate dynamic functional connectivity and static spatial brain region data. Finally, Yin et al. [125] present a system combining graph convolutional neural network (GCNN) and LSTM network to analyze EEG signals. The GCNNs generate domain features from DE processed signal segments, and the LSTMs then extract temporal features and classify emotions by channel relationship.

1.1.3 EEG Limbs Activation

In this section, the works in the current SOTA regarding methods for EEG limbs activation are discussed. In particular, the current literature is reviewed from two perspectives, namely, used features, and used classifiers.

Features

In order to increase the Signal-to-Noise Ratio (SNR) and reduce the presence of artifacts, a labor-intensive pre-processing is commonly needed before interpreting the EEG signals, and this represents a significant obstacle to the broad use of MI EEG-based BCI. Effective feature extraction methods are crucial for better functioning of the system that will benefit from noise-free data containing only the information really relevant to the classification of the limb from MI EEG data. One of the most successful paradigms for data pre-processing is Independent Component Analysis (ICA) [126]. In the work proposed in [127], a channel optimization approach of EEG signals ICA is proposed to minimize the effect of artifacts on channel selection. Since ICA-based MI-BCI systems are sensitive to EEG channels and the quality of the training data, this work is mainly focused on a novel channel selection scheme able to pick only optimal channels resulting in improved performance of the overall system thanks to the proposed strategy. Authors in [128] and [129], on the other hand, focus their work on the dimensionality reduction of EEG signals through an iterative multi-objective optimization for channel selection algorithm and Granger causality [130], respectively. Another classical approach commonly used in literature is the Common Spatial Patterns (CSP). Gaur et al. [131], proposed a method able to examine multiple time segments within a trial rather than depending on a single time point. They found that inter-subject variability has been reduced to some extent, as

evidenced by the lower standard deviation. Moreover, their solution has performed significantly better than the existing state of the art both on stroke patients and on healthy individuals. Instead, the approach presented in [132] introduces an enhanced complex-valued common spatial pattern, which aims to analyze EEG signals by decomposing them into spatial patterns through covariance. The primary objective of their method is to maximize the variance between signals associated with distinct classes while minimizing the variance between signals within the same class. WT is an alternative feature for MI EEG aimed at extracting a time-frequency representation of a given EEG MI signal. Xu et al. [133] use WT to convert specific EEG channels, namely electrodes C3, Cz, and C4, of multichannel EEG signals into two-dimensional time-frequency images to derive the feature of motor imagery EEG signal and train a deep classifier. The authors in [134] propose a method that uses WT for the EEG feature extraction phase and a simple Multilayer Perceptron (MLP) with a single hidden layer for classification. They applied it to both mental and MI tasks to demonstrate the versatility of the proposed approach. Similar to the work reported in [134], the authors in [135] present an algorithm for cursor manipulation on a screen through a BCI. In this case, a WT is used for the feature extraction stage to train a hybrid system made up of a SVM and an MLP was developed. The research outlined in [136] centers around a pre-processing stage designed to decompose μ and β rhythms from EEG signals. The approach employs uncorrelated transform complex common spatial patterns to enhance the independence between the identified rhythms. Another approach involves the use of different aggregate features, for instance, the work reported in [137] differs from the previously mentioned approach because aggregates different time and frequency domain features extracted from the signal, namely mean, variance, skewness, kurtosis, zero crossings, absolute area under the signal, and peak-to-peak distance, to train a deep classifier. An alternative solution for analyzing MI tasks is reported in [138], where a PLV-based method, linked with the synchronization between EEG signals from different areas of the brain, is presented to discriminate tasks changes in cerebral rhythms, when a subject move or think left/right-hand movement. Zhang et al. [139] design a novel deep learning model called Graph-based Convolutional Recurrent Model (CGRAM) able to efficiently learn spatial information representing the input EEG as a graph, extracting attentional temporal dynamics by means of a recurrent attention network. Since in the standard recurrent module the temporal cues are usually accumulated to the last time step, eventually losing some critical information in early time steps, the proposed RNN model assigns weights to different temporal cues and aggregates all information for the final classification. Finally, some methods directly use a deep convolutional model to extract the most significant feature maps from the input EEG. Dose et al. [140] employ a CNN within an end-to-end framework to simultaneously extract features and classify EEG signals. This approach eliminates the necessity for ad-hoc features, resulting in a substantial enhancement in both intra-subject and cross-subject accuracies compared to prior non-deep methodologies. Instead, Huang et al. [141] aimed to propose a deep learning architecture introducing the local reparameterization trick into CNNs to directly process the EEG signals. In this way, the approach can generate a (statistically) effective gradient estimator, effectively reducing the variance and thereby making the classification accuracy higher and converging faster.

Classifiers

Regarding MI EEG different approaches have been proposed to classify the raw signal or the features extracted mentioned in previous section. In particular, concerning CNN methods, Huang et al. [141] deploys a simple 4-layer CNN architecture where each convolutional layer has 32, 16, 8, and 4 kernels, respectively. [140] proposes a CNN that consists of two convolutional layers with 40 kernels per layer. The former layer only performs convolution along the time axis, while the latter layer only performs convolution along the EEG channel axis. The classification is carried out by means of a Fully Connected (FC) layer. Karacsony et al. [142] employs a CNN in which the first convolutional layer performs temporal filtering, applying 100 filters along the time axis with a kernel size of 25 samples. The resulting temporally filtered feature maps are then passed to a second convolutional layer, which handles spatial filtering across EEG channels at each time step, producing 100 feature maps. This spatial filter also reduces the dimensionality across channels, resulting in 1D feature maps. These 1D maps, after temporal and spatial processing, are further processed through two additional convolutional layers, which refine the features and extract higher-level abstractions, each yielding 50 feature maps. Concerning LSTM methods, the paper from Zhang et al. [137] proposes an LSTM network with an attention mechanism for classifying left/right-hand movements using EEG signals. After EEG signals cleaning, a wide range of time and frequency domain features are extracted from each EEG segment, including mean, variance, and relative power in different frequency bands and given as input to a multi-layer LSTM network. Each layer contains LSTM cells that capture both long-term and short-term temporal dependencies in the EEG data. The attention mechanism is used to assign higher weights to time steps that contain more relevant information for classification. In contrast, the approach presented in [139] introduces a solution based on a CGRAM network, where a graph structure is initially constructed to represent the spatial arrangement of EEG electrodes. A convolutional recurrent model is then applied to learn EEG features across both spatial and temporal dimensions, emphasizing the most distinctive temporal intervals. On top of the RNN, a recurrent attention module is used to assign weights to different temporal cues, rather than relying on traditional accumulative temporal information.

1.1.4 Contributions and outline

This section reports in brief the contribution of this thesis work with respect to the state-of-the-art. Concerning EEG biometric identification an innovative graph representation of the EEG signal, able to integrate the time-frequency oscillations of the EEG and the brain functional connectivity has been presented. Then, a GNN able to extract and exploit distinctive features from the proposed graph structure is used in order to identify the subject, leading to better performance and enhanced capacity to generalize over diverse human states. A comprehensive investigation of the proposed system evaluated on EEG signals of diverse human states from the public PhysioNet and M3CV datasets, alongside a comparison with SOTA approaches.

Regarding EEG emotion recognition, an original emotion recognition framework,

namely Empátheia, has been deployed. The system provides a spatio-temporal atlas representation for EEG data is provided thanks to a custom encoder based on the PRISMIN framework. Then, to classify the emotion represented in the atlases, a multi-branch classifier, each one designed to capture spatial and temporal features of an image-encoded emotion, is used. The system, thanks to the compression capacity of the atlases, provides an alternative way to treat and manage large collections of EEG signals to support several critical issues, including storage space, embedded systems, transmission efficiency, and many others. An extensive experimental study has been performed to explore different encoding strategies and classifiers to retain a high performance on the SEED benchmark dataset despite the significant data quantization produced by the PRISMIN encoder.

Finally, with respect to EEG limb activation, a modular multi-stream 1D CNN architecture has been developed. The main novelty of the model is the use of a multi-stream strategy, where parallel kernels with different sizes on each stream are able to retrieve meaningful information from the same data at different time scales without the need for pre-processing methods. An extensive ablation study has been performed, and it has emerged that with this specific type of data, shallow models with more streams achieve better results than deeper models with fewer streams. Then, comparative experiments on the PhysioNet EEG motor movement/imagery dataset show how the proposed model overcomes the key works of the current literature in both cross-subject and intra-subject analysis.

Chapter 2

The EEG signal

Electroencephalography is a method to record the electrical voltages produced from the spontaneous electrical activity of the brain. The EEG signal enables the recording of fluctuations in bioelectric potential related to brain activity that can detected using electrodes placed on the scalp and appropriate instrumentation. In 1924, Hans Berger, through the invention of a rudimentary electroencephalograph, successfully recorded signals from his son's head, observing the first temporal patterns of brainwave activity. To fully understand how to exploit the EEG signal, it is essential to be familiar with the brain's areas and functions, as well as how these are linked to various brain functions or potential disorders. In this chapter, an anatomical description (Sec.2.1) and a functional overview (Sec.2.2) of the human brain are first provided. Then, Sec.2.3 presents the genesis of the EEG signal, while the associated rhythms are reported in Sec.2.4. Finally, the signal acquisition and processing systems are mentioned in Sec.2.5.

2.1 Brain anatomy

The brain is divided into three parts: a central part known as the diencephalon (whose primary structure is the thalamus) and two symmetrical parts, the cerebral hemispheres, which form the telencephalon. The two hemispheres are separated by the interhemispheric fissure, which extends to the white matter nucleus called the corpus callosum. Along with the Rolandic and Sylvian fissures, this fissure divides each hemisphere into four lobes: frontal, parietal, temporal, and occipital (Figure 2.1). Each lobe is further divided by grooves into convolutions, within which specific motor and sensory projection areas are located. The brain also contains four cavities (the lateral, third, and fourth ventricles) filled with cerebrospinal fluid. The majority of the diencephalon is occupied by numerous small nuclei forming the thalamus, which acts as a relay and processing station for sensory information en route to the cerebral cortex. The thalamus is involved in all sensory systems except the olfactory system and also plays a key role in voluntary movement, regulation of the sleep-wake cycle [144], maintaining the proper balance between excitation and inhibition, emotional responses [145], and certain aspects of memory capacity [146]. The walls of the hemispheres consist of grey matter (the cerebral cortex), rich in neurons, which covers the white matter formed by nerve fibers. The main white

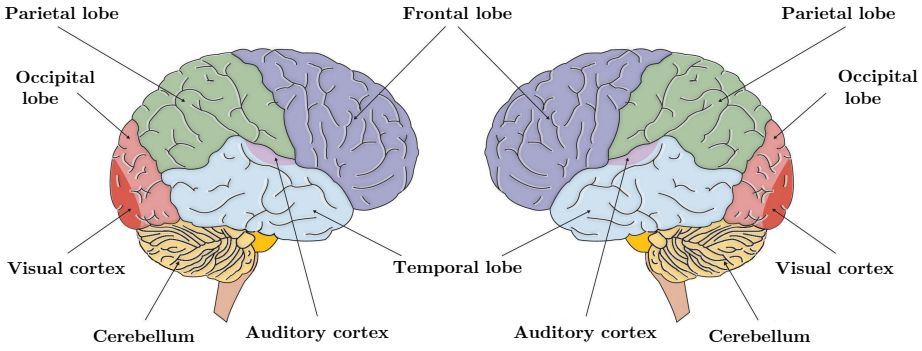


Figure 2.1. The Two Hemispheres. The brain is divided into two hemispheres (left and right), each of which has four lobes (temporal, frontal, occipital, and parietal). Furthermore, there are specific cortical areas that control different processes. Adapted from [143].

matter nuclei include the aforementioned corpus callosum, located beneath the interhemispheric fissure, and the fornix or cerebral trigone, a laminar structure located beneath the corpus callosum, separated from it by a transparent layer (the septum pellucidum). Embedded within the white matter are nuclei of grey matter, the most significant being: the hippocampus or Cornu Ammonis, which receives olfactory stimuli and participates in emotional mechanisms [147, 148]; the optic thalamus, important for regulating sensitivity and responses to painful or emotional stimuli; the striated nuclei (caudate and lentiform), which are centers for automatic activities associated with conscious action [149]; and the hypothalamus (including the optic chiasm, tuber cinereum, and the two mammillary bodies), which lies beneath the thalamus and houses various regulatory centers for specific functions (body temperature, appetite, urination, blood pressure, thirst, etc.) and is connected to the pituitary gland [150]. The cerebellum is located in the posterior-inferior region of the cranial cavity, between the brain and the medulla oblongata, and is divided into two cerebellar hemispheres, joined by the cerebellar vermis. The cerebellar cortex consists of three layers of grey matter: an outer molecular layer, an intermediate layer, and an inner granular layer. The white matter (arbor vitae) branches through the cerebellar cortex. The cerebellum (accounting for 10% of brain mass) coordinates muscular movements, governs the sense of the body's orientation in space, and maintains balance [151]. The brainstem, composed of the medulla oblongata (connected to the spinal cord), the pons (pons Varolii), and the midbrain, contains all the afferent and efferent nerve fibers that connect the higher brain centers to the spinal cord. The cranial nerves consist of twelve pairs (Figure 2.2), classified as sensory (I-II-VIII), motor (III-IV-VI-XII), and mixed (V-VII-IX-X and XI) [152]. The first and second pairs (olfactory and optic) originate above the brainstem and are considered part of the brain, while the third (oculomotor) and fourth

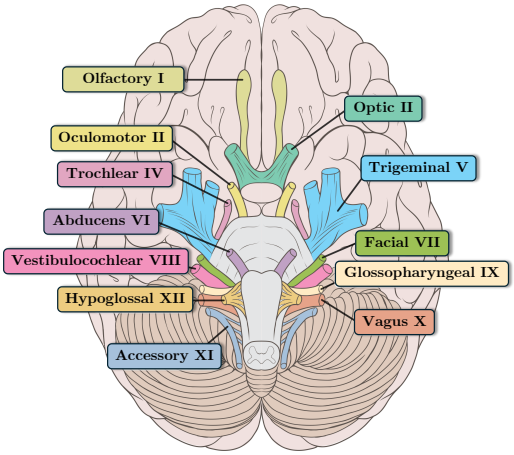


Figure 2.2. The disposition of cranial nerves. Adapted from [154].

pairs (trochlear) arise from the midbrain; the fifth (trigeminal, with ophthalmic, maxillary, and mandibular branches) and eighth pairs (auditory) originate in the medulla and pons; the sixth (abducens) and seventh pairs (facial) arise from the pons; and the ninth (glossopharyngeal), tenth (vagus), eleventh (accessory), and twelfth (hypoglossal) pairs arise from the medulla. Within the brainstem, between the cranial nerve nuclei and the major ascending and descending neural pathways, lies the reticular system, which coordinates complex functions such as wakefulness, swallowing, and respiration [153].

2.2 Brain function

The brain is the principal organ of the human nervous system and, from a physiological standpoint, the primary center of somatic, relational, and intellectual life. It is to the brain, and more precisely, to its cortex-that stimuli transmitted from the periphery via sensory nerves arrive, where they are transformed into conscious impulses and sensations. Voluntary responses originate from the cortex and are transmitted to muscles and peripheral organs via efferent nerves. Although the brain functions as a unified organ, certain regions are specialized for specific functions. For instance, the visual sensory area is located in the occipital lobe; the auditory sensory area is situated in the temporal lobes; the olfactory and gustatory sensory areas are at the level of the hippocampus; and higher-order functions are controlled by the frontal lobe. Thus, in general, the cerebral cortex can be divided into specialized functional areas (projection areas) for different cerebral functions. However, the concept of localization cannot be strictly defined, as each brain area integrates with others in an overall framework of coordination. The cerebral cortex is functionally divided into sensory, motor, and associative regions. The sensory cortex receives nerve pathways that carry stimuli from across the body. Motor impulses originate from the motor areas and are transmitted via the pyramidal pathway (typically, the cortical centers of one hemisphere are related to regions of the opposite side of the body). Associative areas integrate various sensations, store them, and contribute to the complex process of consciousness, including ideation, will, awareness, and

judgment. In both the sensory (left) and motor (right) cortices, there are areas corresponding to specific parts of the body. The size of each area is proportional to the sensory accuracy required for that specific body part. As regards intellectual activity, current knowledge suggests that there is no specific area responsible for the emergence of ideas or the localization of memory. Instead, these capacities are believed to be distributed across the entire cerebral cortex and realized through the association of different higher neural centers. Morover, the medulla regulates the heart beat and the breathing, in fact it contains cardiovascular and respiratory centers that are able to monitor and regulate heart rate, blood pressure, the frequency, and depth of breathing. The Varoli pons instead contains nuclei that collaborate with those in medulla to modulate the rhythm and respiratory pattern, influencing the transition between inspiration and exhalation.

2.2.1 Cortical neurons

The brain is made up of $10^{10} \setminus 10^{11}$ neurons [155]. Neurons consist of a central body, called the soma, from which two types of extensions emerge: dendrites and the axon. Dendrites are thick, short branches that extend near the soma and have the capacity to transmit electrical impulses received from external sources or other neurons to the cell body. The axon, a single long filament unique to each neuron, transmits the nerve impulse to other neurons or tissues, such as muscles, in the direction it extends. In fact, a bundle of several axons forms a nerve. The latter can transport information and signals from the central nervous system to peripheral organs/tissues (efferent nerves) or vice versa (afferent nerves), that is from the periphery to the Central Nervous System (CNS). Each nerve cell is enclosed by a cellular membrane that functions as a highly selective filter, extremely thin (0.005 μm thick), and consists of two layers of phospholipids. Specific receptors for certain chemical substances (both exogenous and endogenous) and specialized channels for the passage of ions with positive (Na^+, K^+, Ca^{2+}) or negative (Cl^-) electrical charges are located on the cell membrane. These channels can be opened or closed by specific biochemical modifications of the membrane itself. Neuronal cells constitute, together with glial cells, the nervous tissue. Glial cells outnumber neuronal cells and have a nourishing, supporting and protective function for neurons, also having a role in nerve impulse transmission. Communication between two neurons is made by synapses, also called synaptic junctions, which act as connecting points, allowing the transmission of information between neurons (inter-neuronal synapses) or with other cells (muscle, sensory or endocrine cells), in the form of electrical signals (electrical synapses) or chemical signals (chemical synapses). The cerebral cortex plays a primary role in bioelectrical activity. There are various types of cortical neurons, which are not randomly distributed across the surface of the cortex but are arranged in an organized fashion, forming what are known as cortical layers. The cortex, composed of 1.5–4 mm of grey matter, is generally divided into six sublayers (Figure 2.3) that contain neurons and fiber bundles [156]. Two main types of cortical neurons can be distinguished: pyramidal cells and non-pyramidal cells. Additionally, there are various horizontally oriented nerve fibers that extend between adjacent regions of the cortex, as well as vertically oriented bundles that extend from the cortex to more distant regions or down to the spinal cord. Pyramidal cells have a conical

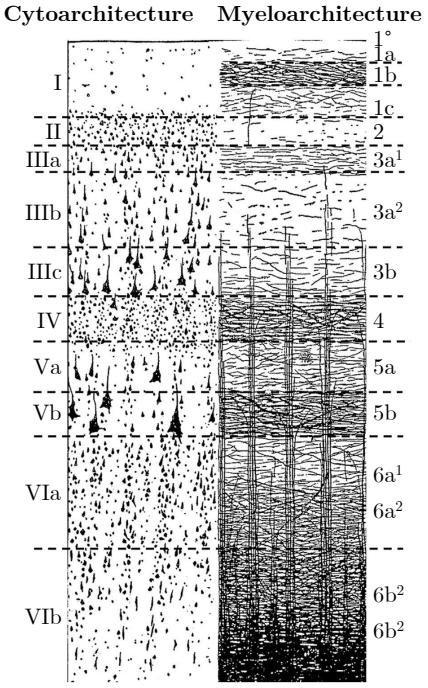


Figure 2.3. Generalized scheme of cortical layering. Adapted from [157].

triangular-shaped body, with the base pointing downward and the apex directed toward the cortical surface, resulting in a perpendicular arrangement relative to the cortical surface [158]. Pyramidal cells are among the largest neurons in the brain. Due to their branching structure, the overall length of their dendrites can reach several centimeters. These cells are heavily involved in voluntary movement guided by vision, playing a significant role in cognitive abilities, such as object recognition tasks. Pyramid cells (Figure 2.4) are composed of:

- a single long, branched axon that acts as an output for other neurons;
- a single long apical dendrite that descends from the apex of the cell body through the various cortical layers and finally branches into different endings. Pyramidal neurons respond to short depolarizing stimuli, generating action potentials lasting between 0.5 and 1 ms;
- basal dendrites that originate from the base of the cell body. The tree of basal dendrites is composed of a small number of dendrites. They do not reach the same length as the apical dendrite, but they also have numerous branches;
- dendritic spines, small protrusions on the dendrites that receive excitatory signals. The density of dendritic spines on a neuron is correlated with the number of connections it forms with afferent axonal terminals and can be considered an indicator of the complexity of its functions.

The non-pyramidal cells have a small cell body and their dendrites, practically without spines, branch in all directions in the immediate vicinity of the cell; with

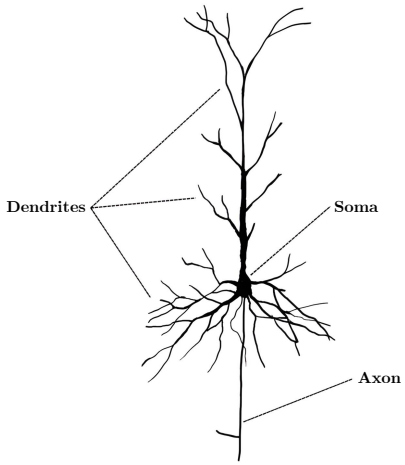


Figure 2.4. Basic components of a pyramidal cell. Adapted from [159].

very rare exceptions, the axons do not leave the grey matter, and project short axons to other local neurons. They do not have a preferential orientation with respect to the surface of the cortex and perform a function purely as inhibitory interneurons. In fact, these neurons release GABA [160], the main inhibitory neurotransmitter in the CNS, particularly at the level of synapses of the CNS, where it is able to regulate neuronal excitability and muscle tone.

2.2.2 Neural groups

As previously mentioned, the cerebral cortex is a continuous laminar layer of grey substance that covers the white substance of the cerebral hemispheres: it represents the outermost part of the brain in vertebrates. It consists of neurons, glia and nerve fibres and is about 2-4 mm thick. The human cerebral cortex plays a central role in complex cognitive mental mechanisms or functions such as thinking, awareness, memory, attention, and language. All brain functions are located in the cortex, which covers the cerebral hemispheres and in which all information is processed and integrated. The cortex is the result of the particular cerebral evolution of primates, and during this evolution, its volume has increased faster than the volume of the skull, forming a large number of cracks called furrows and gyrus, whose ridges are called turns. Some grooves are always present in the human brain, and they are the reference to divide the cortex of each hemisphere into four lobes: frontal, parietal, temporal and occipital (Figure 2.1). It has long been known that certain areas of the brain are engaged in specific tasks, both cognitive and motor. Each is assigned to one or more specific functions:

- The frontal lobe houses the motor cortex and premotor cortex, playing key roles in learning, memory, and, on the left side (Broca's area), in the formation and control of speech. It is essential for planning and executing purposeful, learned behaviors and is also the center for many inhibitory functions [161];
- The parietal lobe includes the primary somatosensory cortex. The left hemisphere of this lobe plays a dominant role in language comprehension, word

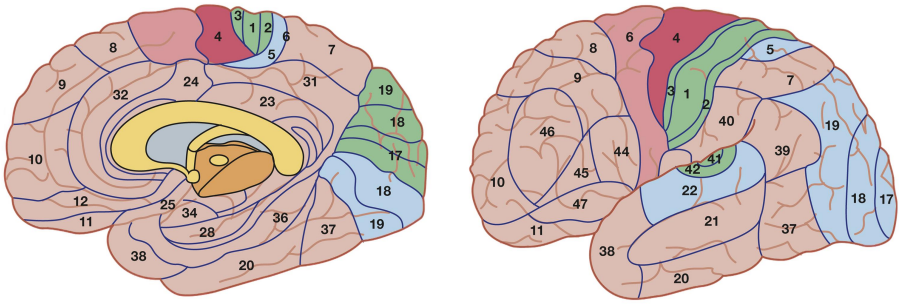


Figure 2.5. The Brodmann map. 4, primary motor cortex; medial 6, supplementary motor area; lateral 6, premotor cortex; 3, 1 and 2, (primary) somatic sensory cortex; 40, secondary somatic sensory cortex; 17, primary visual cortex; 18 and 19, visual association cortex; 41 and 42, primary auditory cortex; 22, auditory association cortex. Adapted from [166].

memory, and mathematical abilities, whereas the right hemisphere is primarily responsible for visuospatial functions [162, 163];

- The temporal lobe contains the auditory processing area and is involved in affective processing, emotional reactions, visual recognition, auditory perception, and memory. The left temporal lobe is crucial for speech and word selection, while the right temporal lobe aids in interpreting speech intonation and the sequence of sounds [164];
- The occipital lobe integrates all visual information, including input that influences posture and balance [165].

Each lobe is then divided in turn by grooves and folds into circumvolutions in which there are other distinct functional subregions, connected to each other by tracts of axons that start from one population of neurons and end in another. In the literature, a widely used subdivision of the cerebral cortex is that of Brodmann, who identified 47 precise areas (Figure 2.5) with distinct characteristics [166]. In particular, it is possible to distinguish between several areas that are related to certain specific tasks, such as the primary sensory area (areas 3-1-2) and secondary (areas 5 and 7), the primary visual area (area 17) and secondary (area 18, 19), the primary motor area (area 4), the premotor area (6). In the sensory cortex (left) and motor (right), there are areas that relate to specific body parts. The size of each of these areas is proportional to the sensory accuracy required for the specific body part. This gives rise to a representation known as the homunculus 2.6, a visual map of how different parts of the body are represented at the cortical level. The larger the areas are, the more important they are for sensory perception. Some sensory areas related to the lips and hands are much wider than the areas representing the central parts of the body and the eyes. These groups of cortical neurons undergo cycles of activity in which they are sequentially recruited, engaged in processing activities, and then released. During normal brain function, these networks are subjected to rhythmic activity occurring at frequencies ranging from 1 to over 100 Hz. The underlying neuronal activity proceeds at a frequency of thousands of hertz, but all

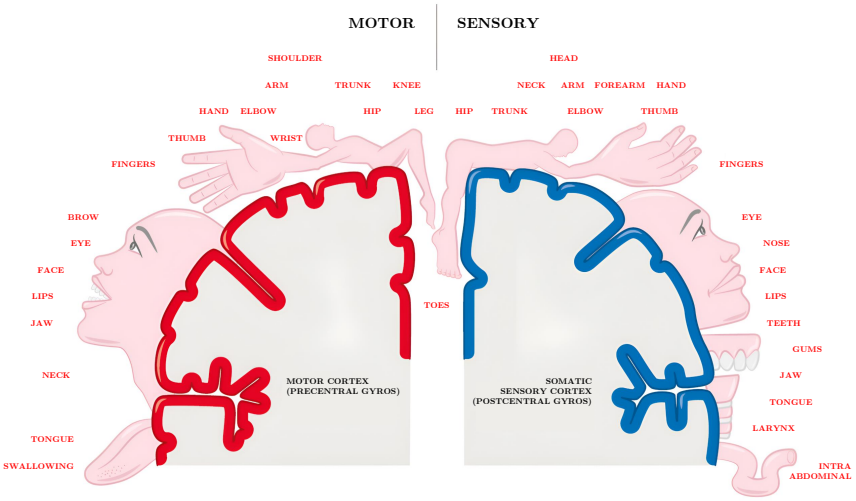


Figure 2.6. The somatosensory homunculus and the motor homunculus. Adapted from [168].

of the measurable external potentials are in the EEG range. Neuronal subgroups have the ability to isolate themselves from adjacent neurons through the mechanism of lateral inhibition. This mechanism consists of the inhibition, by the stimulated neurons, of the lateral zones to themselves through their own inhibitory cells. For this purpose, sensory receptors exploit specialized membrane zones, which are useful for transducing the signal and modifying the opening or closing of the ion channels with the generation or not of an action potential. This way the excited area of the cortex will be well-defined and corresponding to the specific part of the body source of the stimulus. By contrast, through lateral inhibition [167], the lateral neurons will be inhibited through stimulation of lower intensity received by the more stimulated neurons, thus distinguishing them from the former. This is the device that allows the CNS, for example, to discriminate between two stimuli occurring at two close points. The coordinated activity of different regions is evidenced by rhythmic waves, which are distinguished at particular positions according to specific brain activities or stimuli. This cyclical pattern of activity produces an identifiable rising and falling of measurable rhythms, which has a time course in the order of seconds and shows great variability. As a result, it is possible to identify the dominant rhythms present, each indicating the general state of activation or relaxation of a given region.

2.3 EEG signal genesis

The activity of the CNS is mainly connected to the current that develops in the synapses between dendrites and axons (asso-dendritic synapses), dendrites and dendrites (dendro-dentric synapses), and axons and axons (asso-axonic synapses). A negative polarity potential of the order of 60-70 μV can be measured between the inside and outside of the cell body membrane. The information transmitted by the nervous system is called Action Potential (AP). A stimulus must pass a firing threshold, about $-55\mu V$, to generate a AP, while a weak impulse does not generate

an action potential. Neurons have a resting potential of about $-70\mu V$, but it can change in case of depolarization or polarization of the membrane. If the opening of the ion channel, a pore-forming membrane protein that allows ions to pass through the channel pore, results in a net gain of positive charge across the membrane, the membrane is said to be depolarized as the potential comes closer to zero. This is referred to as an Excitatory Post-Synaptic Potential (EPSP), as it brings the neuron's potential closer to its firing threshold (about $-55\mu V$). If, on the other hand, the opening of the ion channel results in a net gain of negative charge, this moves the potential further from zero and is referred to as hyperpolarization. This is an Inhibitory Post-Synaptic Potential (IPSP), as it changes the charge across the membrane to be further from the firing threshold. These post-synaptic potentials have a duration ranging from ten to a hundred ms, higher than that of the action potential, and a maximum amplitude of 20 μV, which tends to attenuate in propagation. This amplitude is not sufficient to trigger, on its own, the action potential in another cell. Each neuron is made up of a multitude of dendrites and receives a large number of synapses from different neurons. These potentials are subject to summation, spatially and/or temporally. More specifically, if a cell is receiving input at two synapses that are near each other, their post-synaptic potentials add together. If the cell is receiving two excitatory post-synaptic potentials, they combine so that the membrane potential is depolarized by the sum of the two changes. If there are two inhibitory potentials, they also sum, and the membrane is hyperpolarized by that amount. If the cell is receiving both inhibitory and excitatory post-synaptic potentials, they can cancel out, or one can be stronger than the other, and the membrane potential will change by the difference between them. Regarding the temporal case, when a cell receives inputs that are close together in time, they are also added together, even if from the same synapse. Thus, if a neuron receives an EPSP, and then the pre-synaptic neuron fires again, creating another EPSP, the membrane of the post-synaptic cell is depolarized by the total of the EPSPs. The EEG tracing is the result of the sum of the synchronized post-synaptic potentials of a large number of neurons belonging to populations located below the scalp recorded over time. The closest structure to the scalp is the cerebral cortex, so the EEG reflects the activity of cortical neurons beneath the electrodes. Cortical neurons are pyramidal and non-pyramidal. Pyramidal neurons are among the largest in the brain and make up about $\frac{3}{4}$ of cortical neurons. As mentioned above, non-pyramidal neurons, also called inter-neurons, communicate at short distances due to short dendrites. For a potential change to be detected on the surface, it is necessary that the dendritic arrangement is ordered, otherwise, if the cell arrangement is random, a zero current would result. Conversely, the pyramidal cells are arranged vertically, perpendicular to the surface and parallel to each other. This geometric and orderly arrangement promotes sufficient synchronism in activation and makes these neurons the generators of the EEG. In fact, these cells, due to the different states of polarization of dendrites and soma, present a region of positive charge, referred to as a source, and a region of negative charge, referred to as a sink. The result is a potential difference between the apical and basal regions of the neuron. Thus, the effect of the post-synaptic potential can be represented by a dipole radially oriented to the cortex, the so-called "dipolar fields". In particular, the apical dendritic membrane, due to depolarization, is electronegative compared to soma and basal

dendrites, which are, therefore, electropositive. This phenomenon generates an ionic current flowing from the positive to the negative pole through the extracellular medium. The post-synaptic potential influences the signal detected on the surface through three parameters:

- the excitatory or inhibitory nature;
- the intensity;
- the location in relation to the measurement site.

The intensity of the post-synaptic potential depends on the ion variation of the membrane, due to the ionic currents passing through it, which in turn depend on the concentration of the neurotransmitter. Post-synaptic potentials are called 'local potentials' because they only transmit over short distances. This is due to two factors:

- the current leakage, because the membrane is not a good insulator;
- the cytoplasm is resistant to electrical flux.

For both these factors, the potential is reduced as it propagates through the membrane and moves away from the source site. For this reason, non-pyramidal cortical cells do not appear to make a significant contribution to the signal recorded on the skin because each cell produces a very small current flow in the immediate vicinity, but due to a phenomenon known as 'conduction volume', this is spread throughout the brain. Volume conduction (or "electrical spread" [169]) refers to the complex effects of measuring electrical potentials a distance from their source generators [170], in fact, there could be some biological tissue that fills the space between an electrical source and the sensor (e.g., EEG electrodes). These tissues conduct the electrical signals, causing them to spread and refract. This potentially alters the appearance of the signal once it reaches the electrodes. The whole body, including the head, consists of more than 80% of water and thus acts as a good conductor of electrical potentials. For this reason, other electrical signals, concerning eye movement, muscle and heart activity, can be transmitted through the head by volume conduction. When an EEG signal is recorded, small amounts of other signals from non-neuronal physiological sources will inevitably be included. Since several layers, like the scalp, skull, and many others, cover the brain, the signals produced are attenuated hundreds of times. For example, the skull attenuates the signal about 100 times more than soft tissues. Most of the noise is generated inside the brain (internal noise) or above the scalp (system noise or external noise). Therefore, only large populations of active neurons can generate enough potential to be recorded using electrodes positioned on the scalp. It can be concluded that an input synaptic stimulus generates dipoles and therefore current flows which, thanks to the strongly ordered arrangement, can be summed in time and space and measured, bearing in mind that these measurements will also show potentials arising from other physiological activities diffused due to the conduction volume.

2.4 Electroencephalography and EEG rhythms

The individual neuronal dipoles that add up change in amplitude and direction over time, giving rise to continuous fluctuations of the surface EEG signal. The EEG is, therefore, a complex waveform which includes several frequency components. The main EEG rhythms have been identified through clinical and research experience, and the associated frequency ranges were subsequently determined. In fact, it is preferable to define these EEG components as "component bands" rather than "frequencies", because what more accurately distinguishes them is their physiological significance and visual appearance rather than the use of specific frequencies. Indeed, a particular component may appear outside the typical frequency range. Follows a description of the EEG waves (Figure 2.7) that can be observed in normal background brain activity:

- Delta: has a frequency of 4.0 Hz or below. It tends to be the highest in amplitude and the slowest waves. It is normal as the dominant rhythm in infants up to one year, and in stages 3 and 4 of sleep in adults [171], it is also an attention indicator during mental tasks [44].
- Theta: has a frequency of 5 to 8 Hz and is classified as "slow" activity. Theta waves appear as consciousness slips towards drowsiness. Theta waves have been associated with access to unconscious material, attentional processing and in navigation an episodic memory [172]. Larger contingents of theta wave activity in the waking adult can be caused by various pathological problems like fibromyalgia and depression [173, 174] or mental disorders like Attention Deficit Hyperactivity Disorder (ADHD) [172];
- Alpha: has a frequency between 9 and 14 Hz. It is usually best seen in the posterior regions of the head on each side, being higher in amplitude on the dominant side. It appears when closing the eyes and relaxing and disappears when opening the eyes or alerting by any mechanism (thinking, calculating) [175]. It is the major rhythm seen in normal, relaxed adults. Over the life span, alpha activity becomes dominant during early adulthood [176] and is involved in encoding and retrieval of new information [46];
- Beta: is characterized by "fast" activity. It has a frequency of 15 to 40 Hz. Rhythmical beta activity is encountered chiefly over the frontal and central regions. A beta wave is the usual waking rhythm of the brain associated with active thinking, active attention, focus on the outside world, or solving concrete problems, and is found in normal adults [177]. It may be absent or reduced in areas of cortical damage [178]. It is generally regarded as a normal rhythm. It is the dominant rhythm in patients who are alert or anxious or have their eyes open;
- Gamma: has a frequency between 25 and 140 Hz. It is involved in various cerebral functions, such as perception, attention, memory, consciousness, and motor control [179], and can be increased in amplitude via meditation [180]. Altered gamma activity is a biomarker in many mood and cognitive disorders such as major depression [181] and schizophrenia [182].

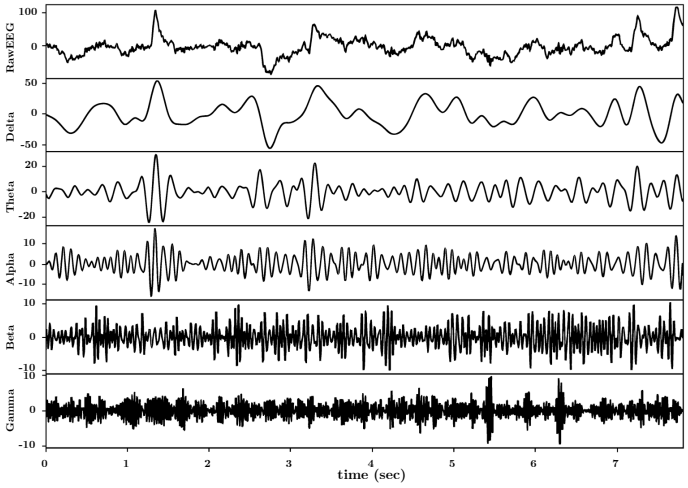


Figure 2.7. EEG raw signal and the relative decomposition into frequency bands.

2.5 Electrodes and positioning systems

An EEG system is therefore able to transform analog signal (i.e. brain potentials) into a digital equivalent that can be stored and analysed. Brain electrical activity is conventionally recorded by means of bio-potential electrodes placed on the head or in the ears. There are two main possible electrodes solutions to record these bio-potentials:

- wet electrodes: The silver/silver chloride (Ag/AgCl) electrode, or "wet electrode," is the most commonly used type in clinical practice for measuring electrical activity due to its ability to reduce skin-electrode impedance via an electrolyte gel. However, wet electrodes have several drawbacks: they require significant time to reach a low impedance level, and the gel can dry within hours, causing impedance to increase. They're unsuitable for long-term EEG monitoring, especially in high-density setups, as the gel may form bridges between close electrodes. Additionally, their setup is time-consuming, requires professional handling, and may be uncomfortable due to abrasive paste and gel;

- dry electrodes: are considered a promising alternative for long-term EEG monitoring, as they eliminate the need for scalp preparation and save setup time. They come in three types: contact, non-contact, and insulating electrodes. Contact dry electrodes touch the scalp directly with a spiked metal array, sometimes even penetrating the outer skin layer for stability and conductivity. Non-contact electrodes use capacitive coupling through an insulating layer, while insulating electrodes have a non-conductive base and also rely on capacitive coupling. Contact electrodes are preferred for EEG due to their lower impedance. However, dry electrodes present high impedance, leading to noise, interference, and movement artifacts. To address these issues, shielded EEG cables are essential to reduce interference, shielding the signal from ambient noise and preventing wire movements from generating interference.

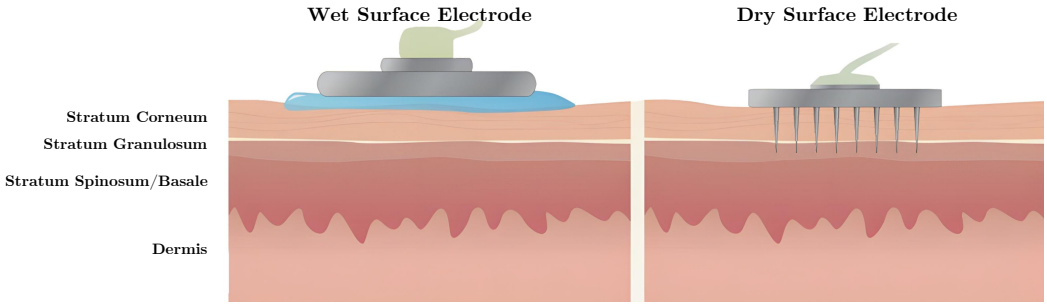


Figure 2.8. Comparison between wet and dry biopotential electrodes: (left) wet electrode that sits on the surface using electrolytic gel. Dry electrodes (right) do not require extensive preparation since they pierce the outer layers of the skin and achieve a low-impedance pathway to recording neural signals. Adapted from [183].

Modern EEG systems are equipped with delicate and accurate electrodes, differential amplifiers, special filters, analog-to-digital converters and instruments for recording and storing data. Once the analog signal is acquired it is amplified, so that it can be digitized accurately. Once amplified, the signal is converted from analog to digital, through an A/D converter. Depending on the purpose of the study, different sampling frequencies can be used, common uses are: 100, 200, 500, 1000, 2000 samples per second. In general, the raw EEG signal has distortions and alterations due to artifacts that can be internal or external. External artifacts are called system artifacts, as they are noise or interference generated by the instrumentation or its improper management. In particular, among the external artifacts are: interference at 50/60 Hz generated by the system power supply, impedance oscillation, faulty cables, unbalanced impedance of the electrodes and electrical noise generated by electronic components. Internal artifacts are associated with the subject and have a physiological nature. Among these, the most important ones to eliminate or attenuate are those related to the movement of the body of the subject, muscle movements, heart beat, eye movement, called blink, sweating and breathing of the subject. In order to remove these artifacts that may compromise the use of the signal, different filtering techniques can be adopted and will be subsequently discussed..

Standard positioning system 10/20

The conventional method for electrode placement, known as the 10-20 system [184], establishes the distance between two adjacent electrodes as a percentage of the total front-back and left-right distance of the skull, with these percentages being 10% and 20%. The electrodes are marked by a letter and a number. The letter identifies the lobe or area from which the electrode is acquiring the signal. In the conventional case, the letters F, T, C, P, O correspond to frontal, temporal, central, parietal, and occipital areas, respectively, with Fp referring to the pre-frontal lobe. Even numbers are assigned to the right side of the brain, while odd numbers are assigned to the left side. Electrodes on the midline are labelled with the letter z [185]. For EEG signal recording, a referential system is primarily used. In this method, one or two electrodes serve as references for the system, generally Cz or, for

example, a pair of electrodes near the ear or above the mastoid. Each electrode is then connected to one input of the differential amplifier, and the reference electrode to another. The EEG recording line generated in this way is commonly referred to as a "channel." Other techniques that do not use a single electrode as a reference employ the average signal across the electrodes as the reference for the system. The pairs of signals (derivations) fed into the amplifiers can be acquired in two different modes, depending on specific experimental requirements: in monopolar mode, one signal of each pair corresponds to an electrode at an active site, while the other serves as a reference signal. The reference can be obtained either through a physical reference electrode placed at an electrically neutral site (such as the tip of the nose, earlobe, or chin) or through a "virtual" electrode, where the reference is obtained as the average of all electrode signals. In this way, the potential of each electrode is measured relative to the neutral electrode or the average of all electrodes, highlighting the level of electrical activity underlying the active site. In bipolar mode, on the other hand, both electrodes are placed at active sites, and the signal detected corresponds to the difference between the activities at the two sites. Typically, in bipolar derivation, pairs of adjacent electrodes along the coronal or sagittal lines are connected to the amplifiers (e.g., F7-F3, F3-Fz, Fz-F4, etc.). Besides the reference electrode, another physically necessary electrode for EEG recording is the ground electrode. Therefore, for single-channel monopolar acquisition, three sensors are required (one active, one reference, and one ground). For two-channel monopolar acquisition, an additional active electrode is used, totalling four sensors. When using surface electrodes, the electrical signal output from the sensor is weak (typically ranging from 25 to 100 μV) and thus needs to be amplified before it can be transmitted, processed, and visualized. An EEG amplifier must, therefore, be capable of capturing very low-amplitude electrical signals from a high-impedance source and amplifying them by more than 1000 times without introducing distortions within variable frequency bands ranging from 0 Hz to 100 Hz or more. The fundamental property of an appropriate biological amplifier is that it is a differential amplifier, which is an electronic device that amplifies the difference between two sites and generates the difference signal as output. It features two input signals connected to its terminals, along with a ground connection; the two input signals are defined as the active signal and the reference signal, and the recording between the two sites represents a single EEG channel. The use of differential amplifiers is essential to separate the useful EEG signal from noise sources that generally appear in phase on both inputs, including offset, drift, and interference. A key characteristic of electroencephalographic amplifiers is the input impedance value, which must be 100 to 1000 times greater than the electrode's output impedance. The standard 10–20 system (Figure 2.9), which comprises a total of 19 sites, 8 on the left side, 8 on the right side and 3 in the middle, did not include electrodes in the inferior chain (at the level of the preauricular point). Thus, the inferior-basal and anterior part of the temporal lobe, which preferentially picks up activity originating or propagating from the mesial temporal structures, was not sampled [186] [187]. Given that several diseases (e.g., temporal lobe epilepsy due to hippocampal sclerosis, autoimmune epilepsy, Alzheimer's disease) are characterized mainly by mesial temporal pathology, this region needs to be targeted through additional scalp electrodes in standard recordings. For the reasons outlined above, [184] strongly recommend to use 25

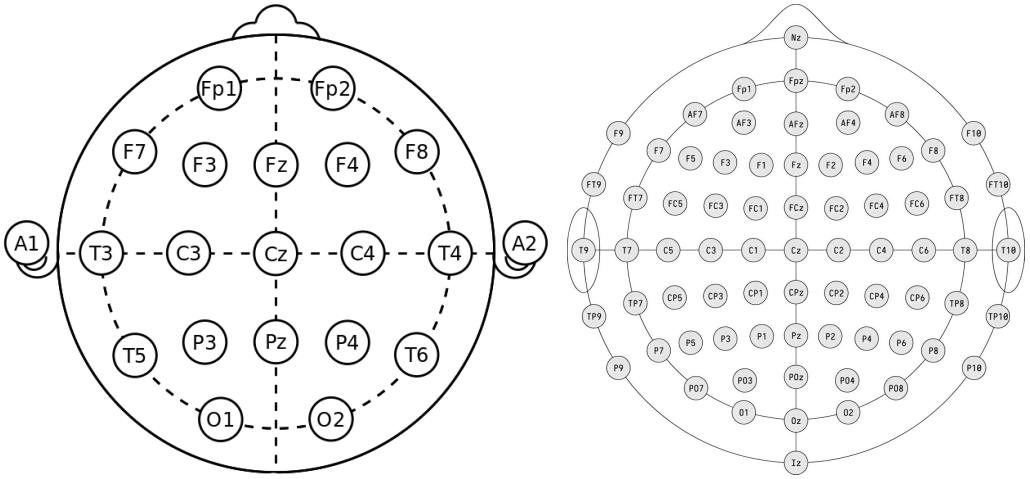


Figure 2.9. Electrode locations of International 10-20 system (left) and 10-10 system (right) for EEG recording. Adapted from [188, 189].

electrodes as a minimum for all standard recordings. The use of fewer electrodes, but no less than 19 electrodes, is acceptable if technical limitations do not permit the use of the full 25-array. Standard electrode set-ups provide an incomplete coverage of the patient's brain but represent a compromise between the everyday routine, given that additional electrodes require additional time and effort of technicians, and reliable detection of all epileptogenic activity. For these reason High Density EEG (HD-EEG), which refers to the use of 64-256 electrodes, has become an established tool over the past 10 years. Technical developments have made it easier to apply a large number of electrodes, which is particularly helpful in the clinical context. Geodesic electrode systems is a term used for equally distributed electrodes over a curved space, like the head. These systems provide dense and even sampling over the entire scalp, neck, and cheeks, allowing the detection of brain activity which could be otherwise missed. The development of digital EEG and the introduction of high-density EEG and source localization methods made it necessary to increase the electrode arrays. Therefore, a modification of 10-20 nomenclature with the definition of 10-10 combinatorial nomenclature (Figure 2.9) has been proposed and accepted by the American Clinical Neurophysiology Society (ACNS) and by the International Federation of Clinical Neurophysiology (IFCN). The modified combinatorial nomenclature is an extension of the 10-20 system, and it entails positioning more than 70 electrodes on the scalp, placed along 11 sagittal chains and 9 coronal chains. The modified 10-10 terminology replaces the inconsistent T4/T3 and T6/T5 terms with the consistent T8/T7 and P8/P7.

2.6 BCI

A BCI can be seen as a system able to translate the brain activity patterns of a person into messages or instructions for an interactive application, this task is measured and processed by the system. Brain signals are captured using invasive or non-invasive 2.10 set of electrodes from the signal acquisition module. Then, the signal processing module extracts only the information needed from the original raw signal. The signal acquisition module comprehends data pre-processing, feature extraction and classification algorithms sub-modules. Finally, the translated signals are passed on to the user through the application interface, and by exploiting the feedback mechanism, real-time results can be produced and understood by the user. A typical life cycle representation of a BCI system is shown in Figure 2.11. According to the task exist different BCIs specialized in capturing potentials related to specific brain tasks:

- P300 based BCIs: BCIs specialized in capturing P300 [190] (or P3) waves that are ERP. This kind of wave appears with a latency between 300 and 600 ms after an external stimulus, but it isn't linked to the physical attributes of the stimulus but reflects the processes involved in stimulus evaluation or categorization.

- SSVEP based BCIs: SSVEPs [191] are the neural responses to visual stimulation at specific frequencies that appear in the EEG signal. The frequency of the acquired EEG wave matched with the stimulus frequency or harmonics of the stimulus frequency. The visual cortex shows electric activity in the same frequency band (and harmonic frequencies) at which the retina is stimulated [192]. These potentials can be observed mainly in electrodes over the occipital and parietal lobes of the brain when a subject is looking at a light source flickering at a constant frequency. So, these potentials are the results of repeated stimulation of the visual cortex.

- Motor imagery based BCIs: with respect the two previous BCIs this one uses brain activity produced spontaneously by the user and not after an external stimulus. This kind of activity called motor imagery [58]is consciously generated by the person to trigger specific BCI commands.

2.6.1 Signal acquisition

To acquire the brain signals different approaches are possible:

- invasive techniques: in this approach, the electrodes are placed directly into the brain cortex of the person;

- semi-invasive techniques: in this approach, the electrodes are placed only on the exposed surface of the brain cortex, not in-depth;

- non-invasive techniques: in this approach the electrodes are placed only on the scalp of the person measuring electrical potentials produced by the brain.

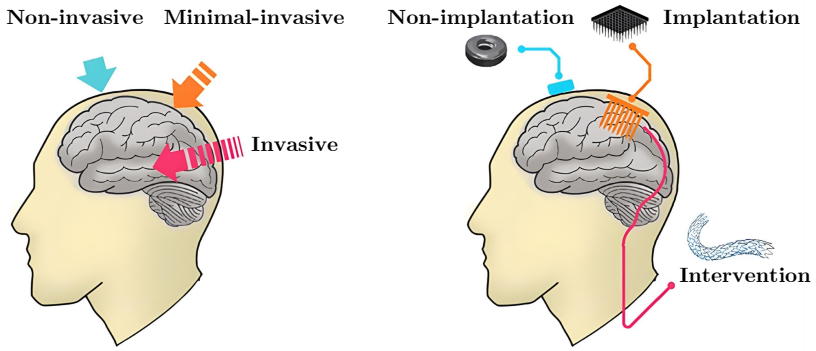


Figure 2.10. Classification of BCI signal acquisition technologies. Adapted from [193].

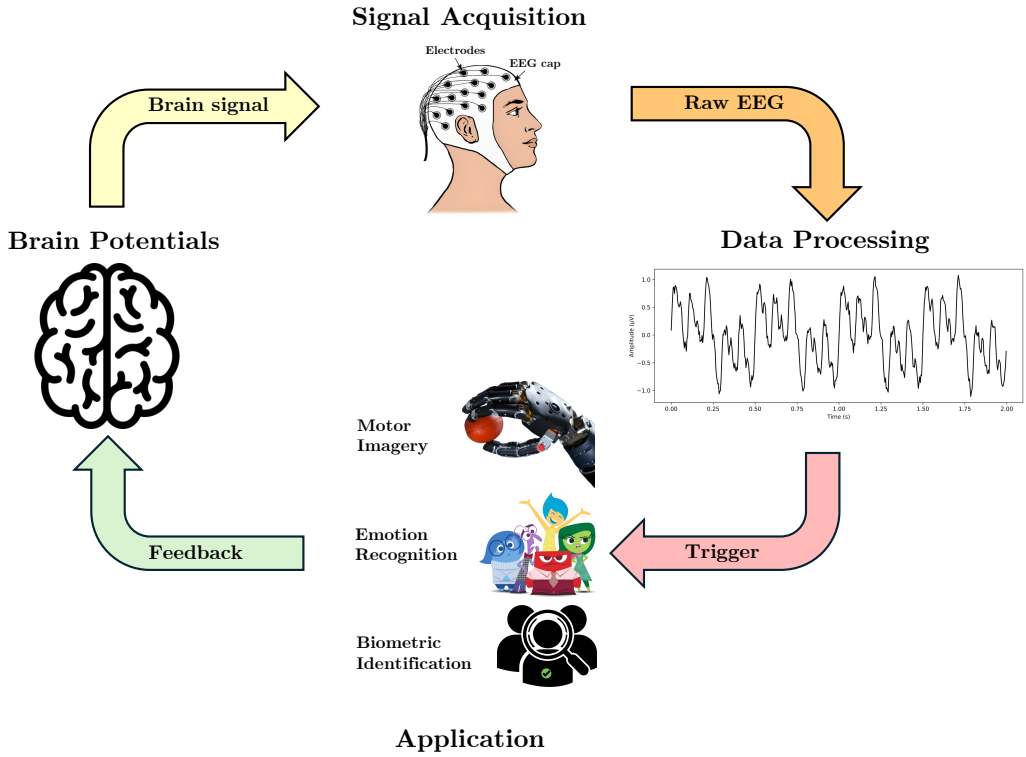


Figure 2.11. BCI system standard life cycle. The three main modules are reported, i.e., the signal acquisition, data processing, and application modules.

Compared to more invasive methods, EEG BCIs guarantee better usability but the extracted signal suffers from poor spatial resolution and considerable amount of noise because of same time background activity: during the recording of signal, various observations and conclusions can affect the target information signal. For these reasons, all the possible background activities that can add noise to the signal need to be mitigated in a pre-processing step.

2.6.2 Artifact removal

When EEG signals are collected through procurement apparatus then there are artifacts and noises such as heart beats, eye blinks and other disturbed factors that are commonly part of the signal. These activities, not related to the desired signal, are always present and should be filtered to obtain higher SNR. SNR is the ratio between the desired information or the power of a signal and the undesired signal or the power of the background noise. This measure is used to compare the level of a desired signal to the level of background noise and is defined as follows:

$$SNR = \frac{P_{signal}}{P_{noise}} \quad 0 \leq SNR \leq \inf \tag{2.1}$$

where P_{signal} is the signal power and P_{noise} is the noise power. To get a clean signal, ready for feature selection/extraction and classification phases, pre-processing can be crucial to maximize SNR. There are different possible approaches for this purpose.

Independent Component Analysis (ICA)

The concept of ICA was first proposed in the engineering field of signal processing [194]. In the simplest terms, ICA algorithms are a family of methods designed to unmix linearly mixed signals using only recorded information about time course, e.g., blind to detailed models of the signal sources as required by earlier signal processing approaches [195]. EEG sources project near-instantly to and sum linearly at the scalp electrodes. The relatively sparse interconnection of cortical EEG source areas suggests that their activities may, given sufficient data, indeed be near independent. Thus, EEG may be reasonably modelled as a linear mixture of the activities from multiple brain and non-brain sources with (near)independent time courses. The data sent to ICA are the recorded EEG channel data rearranged in a matrix of **n** channels (rows) by **t** time points (columns). During the analysis, channel location information is not used at all. ICA performs a blind separation of the data matrix (**X**), based only on the criterion that resulting source time courses (**U**) are maximally independent. Specifically, ICA finds a component unmixing matrix (**W**) that, when multiplied by the original data(**X**), returns the matrix (**U**) of independent component time courses:

$$\mathbf{U} = \mathbf{WX} \tag{2.2}$$

where **X** and **U** are $\mathbf{n} \times \mathbf{t}$ matrices and $\mathbf{n} \times \mathbf{n}$ matrices, respectively. ICA [196] considers EEG signals and artifacts as something independent, preserving data during artifacts elimination. EEG data is divided into spatial-fixed and temporal-independent components. It reaches the best performance, having a considerable amount of data available, and works particularly well at removing ocular signal artifacts.

Common Average Reference (CAR)

Common Average Reference (CAR) [197] is calculated by taking the average of all electrodes on the scalp. In principle, the average reference approximates an ideal

reference. The underlying principle is that the sum of potential fields (e.g., brain potentials) in a conductive sphere is exactly zero when measured over the sphere's surface (e.g., the human head). The approximation to a zero (i.e., inactive) reference is limited by the fact that EEG recordings can cover only $\frac{2}{3}$ of the head. In general, let R be the original reference electrode and let E_i be any other electrode. Any recording at a given site actually shows the difference in potential between the electrode in that site and the reference $(E_i - R)$. Mathematically, re-referencing to the average reference is computed by subtracting the average of all electrodes from each channel:

$$\frac{(E_i - R) - (E_1 - R) + (E_2 - R) + ... + (E_n - R)}{n} = \tag{2.3}$$

$$= \frac{(E_i - R) + R - (E_1 + E2 + \ldots + E_n)}{n} = \tag{2.4}$$

$$= \frac{E_i - (E_1 + E_2 + \ldots +_E n)}{n} = \tag{2.5}$$

$$= E_i - 0, \text{under the assumption that } (E_1 + \ldots + E_n) = 0. \tag{2.6}$$

Therefore, in the limiting case in which the assumptions of the average reference are met, any channel would represent ideal voltages, as the reference becomes zero. CAR emphasizes components that are present in the majority of electrodes and reduces such components. In this way, it functions as a high-pass spatial filter (accentuates components with highly focal distributions). On the other hand, components that are present in most of the electrode population but absent or minimal in the electrode of interest may appear as 'ghost potentials' in CAR recordings.

2.6.3 Signal Filtering

Noise sources in EEG signals can appear as oscillating frequencies. Specifically, low-frequency noise often results from factors like head movements, electrode wire shifts, and scalp perspiration, showing up as slow drifts in the EEG signal over several seconds. On the other hand, high-frequency noise arises from sources like electromagnetic interference and muscle contractions, particularly in facial and neck muscles, creating rapid up-and-down fluctuations in the EEG signal. The frequencies of high and low noise sources can sometimes overlap with the EEG frequency band of interest; however, they are generally lower or higher than typical EEG frequencies. By reducing the signal power at frequencies outside the desired range, it is possible to minimize the noise while preserving the signals of interest. This technique is known as filtering. Filtering in the EEG process occurs at two stages: first, during data collection, called online filtering, and then again during pre-processing, called offline filtering. When EEG data are recorded, the EEG amplifier typically includes a filter that limits frequencies above a set threshold. Known as the low-pass filter cutoff, this filter "passes" lower frequencies while reducing higher ones. A low-pass filter is essential in digital EEG recording (and for any digital signal) to prevent aliasing, a distortion that happens when high-frequency signals are sampled at a rate below their frequency, creating an artificial low-frequency artifact in the recorded signal. Offline filtering allows for more thorough noise removal compared to online filtering. Unlike online filtering, offline filtering enables the application of various

filters to observe their effects on the data, with the flexibility to revert to the original data and adjust filter settings as needed. One practice in offline filtering is to define a high filter and a low filter, going to isolate only a band of frequencies of interest, this is called band-pass filter.

Chapter 3

EEG Person Identification

This chapter describes the architecture designed to solve the person biometric identification task from the EEG signal. Initially, an overview of the proposed method is presented, and then the EEG data pre-processing, the graph representation, and the classifier are introduced in detail.

3.1 Proposed method

In this section, the proposed method is described in detail. The latter is comprised of three main components. The first step is the pre-processing, which deals with EEG signal cleaning from artifacts and noise. Given the EEG, the system produces a clean signal starting from the raw acquisition and restricts the frequencies to the ones that are most significant. The second consists of converting the pre-processed signal into a graph. In detail, a graph representation computing the time-frequency and FC features from the pre-processed EEG is derived. Finally, the last components consist of the GCNN trained on the EEG graphs to discriminate among different identities where the generated EEG graphs are fed into for graph feature extraction and classification. In particular, given a person multi-channel EEGs, the proposed approach is able to represent the pre-processed signals as a complete graph where each node is an EEG electrode, as the signal time-frequency spectrogram, and an edge between two nodes represents the FC across that electrodes pair, expressed as the PLV. Finally, the provided graph that embodies local and global information of the subject's brain in both the spatial and temporal domain is used as input for a GNN for classifying the subject identity. The architecture is based on two spline convolutional layers designed and tuned to capture the most discriminative features from the graphs. Extensive experiments were conducted to assess the generalization capabilities of the proposed architecture in different scenarios. A set of four experiments has been performed to assess the performance of the system under different conditions, training the model using a variety of EEG signals related to multiple brain states (resting, activity, mixed) using different segmentation length for the EEG signal to find the most suitable one. The proposed architecture is reported in Figure 3.1.

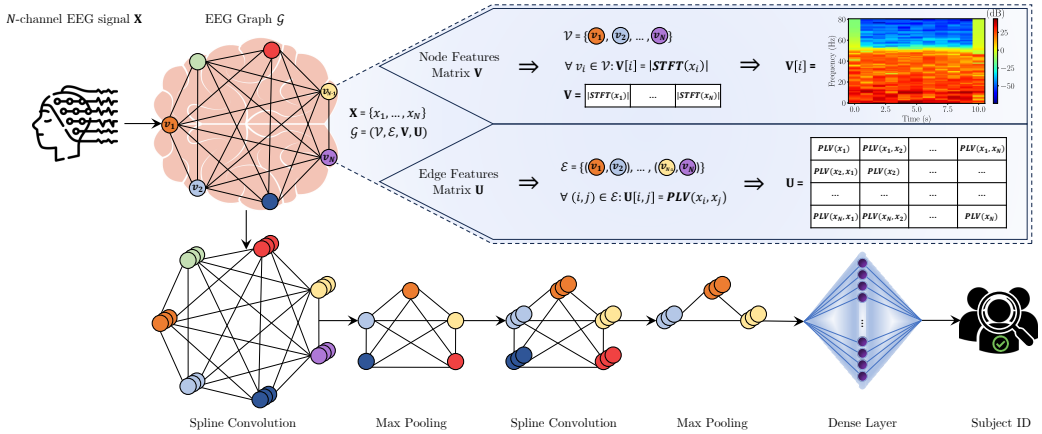


Figure 3.1. Architecture overview of the system.

3.1.1 Pre-processing

This section presents more in-depth the pre-processing operations applied to the raw EEG signal. As a first pre-processing step, for each time step, the CAR of the signals at all electrodes is computed and subtracted from the raw EEG signal. This allows the reduction, or in the best case, complete removal, of noise components that are common to all channels, such as environmental signals or noise that is relevant in a recording context. Then, a band-pass Finite Impulse Response (FIR) filter with Hamming window up to 42.5 Hz is implemented to restrict the available frequencies only to the most relevant range (0-42.5 Hz), and subsequently ICA is applied to remove muscle, Electrooculogram (EOG), and ECG artifacts from the signal. Finally, the filtered EEG recordings are divided into frames of length 1s with a 50% overlap.

3.1.2 EEG Graph

After cleaning the raw EEG signal from noise and artifacts, a graph representation of the clean EEG signal is created. For the sake of clarity, the terms channel and electrode are exchangeable, since the information contained in a single channel is provided by a single electrode. Given a N-channels (i.e., N-electrodes) EEG signal $X = \{x_1, ..., x_N\}$, is represented as a complete undirected graph, defined as $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{V}, \mathbf{U})$. Within this representation, we have that each EEG channel x_i is represented by a node v_i in the node set $\mathcal{V} = \{v_1, ..., v_N\}$, $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the edge set, $\mathbf{V}$ refers to a node features matrix, and $\mathbf{U}$ refers to a weighted adjacency matrix. Regarding the node features matrix $\mathbf{V}$, each entry $\mathbf{V}[i]$ contains the energy spectrogram of the channel x_i computed as the STFT magnitude:

$$\mathbf{V}[i] = |STFT[w, m]| = |\sum_{k=0}^{K-1} h(k) x_i[k + mH] e^{-i\frac{2\pi}{K}wk}|, \tag{3.1}$$

Where $\mathbf{h} \in \mathbb{R}^K : [h(0), h(1), ..., h(K-1)]$ is the Hann window function [198], K is the window length, $0 \leq w \leq K$, is the frequency, m is the index of the sliding

window, $x_i[k+mH]$ is an element of x_i, and $H = \lfloor \frac{K}{2} \rfloor$ is the hop length, the number of samples between the start of two consecutive windows in the signal windowing process. Thus, the node set $\mathcal{V}$ can be formulated as a node feature matrix $\mathbf{V} \in \mathbb{R}^{N \times K \times H}$.

Concerning the weighted matrix $\mathbf{U}$, the latter is formulated as an edge feature matrix $\mathbf{U} \in \mathbb{R}^{N \times N}$, where each entry $\mathbf{u}(i,j)$ represents the PLV measured between channels i and j. PLV analyzes responses to a repeated stimulus and identifies time points where the phase difference between signals remains consistent across trials, indicating phase-locking. Given two series of signals x_i, x_j, and a frequency of interest f, the process computes a measure of phase-locking between the components of x_i, x_j at frequency f for each latency [199]. In detail, the PLV for two signals x_i, x_j and their respective frequencies f_p, f_q is defined as:

$$\mathbf{U}[i,j] = PLV(x_i^{f_p}, x_j^{f_q}) = |\frac{1}{N}\sum_{n=1}^{N} \mathrm{e}^{i\Delta\phi_r(t_n)}| \tag{3.2}$$

where $\Delta\phi_r(t) = |\phi_{x_i}(t) - \phi_{x_j}(t)| \bmod 2\pi$, $\phi_x(t)$ is the instantaneous phase of the signal $x(t)$ obtained by applying the Hilbert transform, and N is the number of trials. The PLV is defined in the interval $[0,1]$ and captures the inter-trial variability of the phase over time. This measure is directly proportional to the phase coupling across the trials, being zero in the absence of phase coupling and increasing as the coupling becomes stronger.

3.1.3 Classifier

The primary challenge in extending classical convolutional neural networks to analyze graph inputs lies in the design of graph convolutional filters. The proposed classifier makes use of a convolution operator, called SplineCNN [200], which aggregates node features in local neighborhoods weighted by a trainable, continuous kernel function. Let $\mathbf{k} = \{k_1, ..., k_d\}$ be a non-decreasing sequence of real numbers, i.e., $k_i \leq k_{i+1}, i = 1, .., d$. The k_i are called knots, and $\mathbf{k}$ is the knot vector, defining a d-dimensional kernel size. Let $((N_{1,i}^m)_{1\leq i\leq k_1}, ..., (N_{d,i}^m)_{1\leq i\leq k_d})$ represent d-open B-Spline bases of degree m based on equidistant knot vectors [201]. where the i-th B-spline basis function of m-degree, denoted as $N_{i,m}(u)$, is defined as:

$$\begin{aligned} N_{i,0}(u) &= \begin{cases} 1 & \text{if } u_i \leq u < u_{i+1} \\ 0 & \text{otherwise} \end{cases} \\ N_{i,m}(u) &= \frac{u - u_i}{u_{i+m} - u_i} N_{i,m-1}(u) + \frac{u_{i+m+1} - u}{u_{i+m+1} - u_{i+1}} N_{i+1,m-1}(u) \end{aligned} \tag{3.3}$$

where $u \in [u_i, u_{i+1})$ is an half-open interval, called the $i-$th knot span; it can have zero length, since knots need not be distinct. Then, the proposed continuous convolution kernel is defined as functions $g_l : [a_1, b_1] \times ... \times [a_d, b_d] \to \mathbb{R}$, with:

$$g_l(u) = \sum_{\mathbf{p} \in \mathcal{P}} w_{\mathbf{p},l} \cdot B_{\mathbf{p}}(\mathbf{u}) \tag{3.4}$$

where $w_{\mathbf{p},l} \in \mathbf{W}$ is a trainable parameter of the weight matrix $\mathbf{W}$, with $1 \leq l \leq M_{im}$, for each of the M_{in} input feature maps and for each element $\mathbf{p}$ from the cartesian

product $\mathcal{P} = (N_{1,i}^m)_i \times ... \times (N_{d,i}^m)_i$ of the B-spline bases, while B_p is the product of the basis functions in $\mathbf{p}$:

$$B_{\mathbf{p}}(\mathbf{u}) = \prod_{i=1}^{d} N_{i,p_i}^m(u_i) \tag{3.5}$$

Given the kernel functions $\mathbf{b} = (g_1, ..., g_N)$ and the input node features matrix $\mathbf{V}$, the convolution operator for a node i is defined as:

$$(\mathbf{f} * \mathbf{g})(i) = \frac{1}{|\mathcal{N}(i)|} \sum_{l=1}^{M_{in}} \sum_{j \in \mathcal{N}(i)} \mathbf{V}[j, l] \cdot g_l(\mathbf{u}(i, j)) \tag{3.6}$$

where $\mathcal{N}(i)$ represents the neighborhood set of node $i \in \mathcal{V}$. The proposed classifier is composed of two Spline convolutional layers followed by a max pooling, to reduce dimensionality, and a dropout layer set to 0.8, to mitigate possible over-fitting during the training [202]. The two Spline CNN layers have a kernel size of 32 and 16, respectively. The last layer is a dense, fully connected layer that provides the subject ID as an output. Batch Normalization (BN) [203] is applied before Spline convolution and dense layers to accelerate the training process and make it more stable. The proposed GCNN architecture is summarized in Figure 3.1. Finally, the model is trained using the categorical cross-entropy loss, which is defined as follows:

$$\mathcal{L}_{CE} = -\sum_{b=1}^{B} \sum_{c=1}^{C} y_b^c \log \hat{y}_b^c, \tag{3.7}$$

Where y_b^c and $\hat{y}_b^c$ correspond to the ground truth and predicted class probability, respectively, C is the number of classes, and B is the batch size.

Chapter 4

EEG emotion recognition

This chapter describes the architecture designed to solve the emotion recognition task from the EEG signal. Initially, an overview of the proposed method is presented, then the spatio-temporal representation for EEG data, and the classifiers are introduced in detail.

4.1 Proposed method

In this section, the proposed method is described in detail. The latter is made up of two main components. Since the main goal of this system is to transform large EEG datasets into compact, time-driven images, the first component is a signal-to-image encoder. Given the EEG, the encoder is able to convert the original raw EEG signal into compressed images, called atlases, incorporating spatial and temporal information from the original input, according to different possible color space encodings. The second consists of a deep classifier trained on the produced compact images to discriminate among different emotions associated with the original signal and preserved in the encoded atlases. In particular, given a person multi-channel EEGs, the proposed approach is able to represent the raw EEG signals as a compact image following a time-driven layout where rows and columns of the encoded image represent, respectively, the channels and time frames of the EEG signals. Finally, the provided atlases that embody spatial and temporal information of the original EEG are used as input for a deep classifier to recognize the emotion associated with the original input. Extensive experiments were conducted for both the encoder and classifier to assess the generalization capabilities of the proposed architecture in different scenarios, exploring not only the performance in terms of emotion classification accuracy, but also evaluating the computational and compression aspects of the proposed approach. In particular, two encoding types, i.e., short-rainbow and grey-scale, and four different models, i.e., one based exclusively on a CNN, two based on mixtures of CNN and RNN, and a last based on a transformer. The proposed architecture is reported in Figure 4.1.

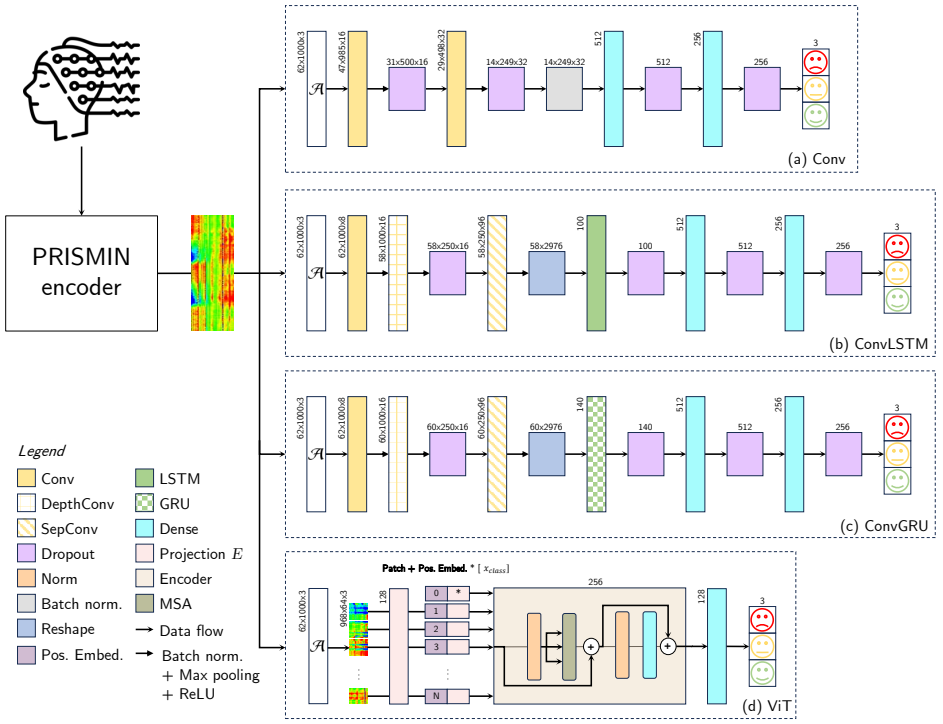


Figure 4.1. Architecture overview of the Empátheia system. The PRISMIN encoder transforms EEG signals associated with emotions into 2D atlases, which are then utilized by the Empátheia classifiers for emotion recognition.

4.1.1 PRISMIN encoder

A common initial step in many studies involves the direct pre-processing or feature extraction of EEG data to eliminate artifacts, such as ocular artifacts, or to isolate specific features, like particular frequency bands, from the original signal. The Empátheia system, however, adopts a distinct strategy, utilizing the PRISMIN framework in its pre-processing phase to convert raw EEG signals into coarse spatio-temporal image atlases that depict emotional states. Specifically, the open-source PRISMIN framework [66] encodes attributes and user states into compact, lightweight 2D images, facilitating easy manipulation and transmission. This framework enables the encoding of session data into image atlases and includes runtime tools for capturing and encoding specific attributes during interactive sessions. In terms of signal compression, as highlighted by Fanini et al. [66], the temporal arrangement coupled with lossless image formats, particularly in cases of gradual pixel variations [204], yields optimal compression ratios. Moreover, the framework is designed for computational efficiency, employing encoding and decoding routines that operate seamlessly on either CPUs or GPUs. In the presented work, the main focus is to design an encoder based on PRISMIN that can transform large EEG datasets into compact, time-driven image atlases [205]. To achieve this, PRISMIN allows the definition of the prism class, i.e., a custom data encoder. Specifically, a given prism $\mathcal{P}$ provides:

- A *refract* method to define how incoming data is encoded as well as the layout

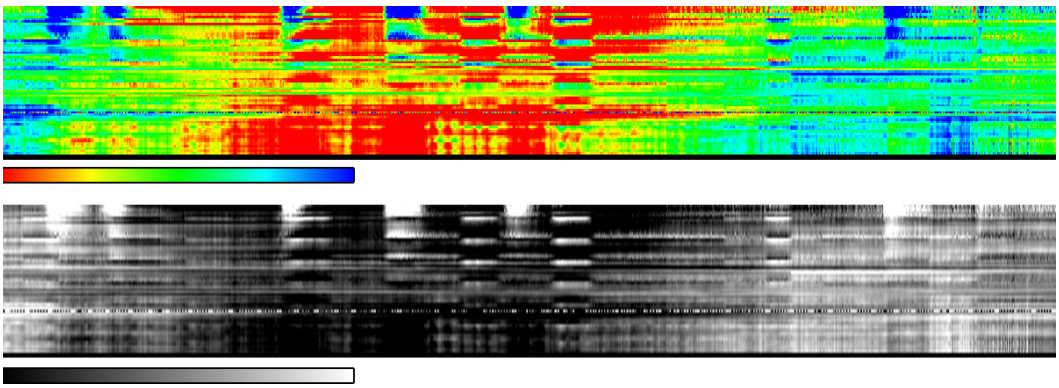


Figure 4.2. Examples of $\mathcal{P}_q$-generated atlases using (top) a short-rainbow color space and (bottom) grey-scale. Each row represents a different EEG channel, pixel color is the encoded voltage over time (x-axis).

adopted in the final image atlas. In this work, spatio-temporal atlases are used, although different methods can be used to implement the refract method.

- A *bake* method to write the actual atlas on disk, using specific image format and bit-depth. This can be implemented as an on-demand routine, for instance, to control write accesses on server-side storages within service-based deployments of the PRISMIN framework.

To adapt $\mathcal{P}$ to handle EEG signals, they must be described so that the *refract* method can accept them. Intuitively, EEG records are represented by streams of voltage values captured by the C channels of a BCI device over a given time period T for a specific EEG session. Formally, a session S containing the brain signals can be defined via the following matrix:

$$S(c,t) = v, \tag{4.1}$$

where $v \in \mathbb{R}$ corresponds to the voltage value of a given channel $c \in C$ at a given time instant $t \in T$. With this formalization, the session matrix S can be used to define a time-driven layout for the generated atlases, where rows and columns of the encoded image represent, respectively, the channels and time frames of the EEG signals.

A quantization error is indeed introduced by *refraction* of voltages (values v) on both short-rainbow and grey-scale. Quantization error in PRISMIN depends in this specific scenario on: 1) voltage ranges, 2) color space adopted, and 3) image bit-depth. An in-depth analysis is described in Fanini et al. [66, 205]. Specifically, given Δ_v as voltage range and bit-depth b to encode incoming values, the maximum quantization errors for grey-scale (ϵ_{gs}) and short-rainbow (ϵ_{sr}) are given by:

$$\epsilon_{gs} = \pm\frac{\Delta_v}{2^{b+1}}; \epsilon_{sr} = \pm\frac{\Delta_v}{2^{b+3}}. \tag{4.2}$$

After defining the *refract* and *bake* methods, a quantized voltage session prism $\mathcal{P}_q$ can be initialized to encode EEG signals into atlases. By applying $\mathcal{P}_q$ over the EEG

dataset, the atlases $\mathcal{A} = (\mathcal{A}_1, \mathcal{A}_2, ..., \mathcal{A}_n)$ can be computed, where n corresponds to the number of samples in the EEG collection. Examples of $\mathcal{P}_q$-generated atlases using two different color spaces are shown in Figure 4.2.

4.1.2 Empátheia classifier

To perform EEG emotion recognition on the encoded atlases $\mathcal{A}$, a classifier must be implemented. Due to the innovative 2D image representation of $\mathcal{A}$, a natural choice for classification would be a CNN model. However, since the atlases capture spatio-temporal features associated with specific emotions, alternative architectures, such as those based on RNNs or transformers, may also prove effective. The following sections detail several models that can be employed as classifiers within the Empátheia system.

Conv classifier

The first classifier, depicted in Figure 4.1(a), is a simple CNN architecture since the encoded EEG signals are effectively transformed into images. Specifically, it consists of two convolutional layers, each followed by batch normalization and max pooling operations, as well as the ReLU activation function, which is defined as follows:

$$f(x) = \begin{cases} x & \text{if } x > 0; \\ 0 & \text{otherwise.} \end{cases} \tag{4.3}$$

The uniqueness of these convolutions lies in their kernel size, which is sufficiently large, such as 16×16, to capture the temporal aspects embedded within the atlas. These layers are tasked with extracting feature maps from the input atlases, which, following a flattening operation, are subsequently classified using a series of three dense layers. These dense layers consist of fully connected layers interspersed with dropout layers to improve the model's ability to generalize and abstract features. Additionally, the first two dense layers utilize a ReLU activation function, while the final layer employs a softmax function to calculate the probability distribution across the set of available emotions as follows:

$$\sigma(x)_i = \frac{e^{x_i}}{\sum_{j=1}^{K} e^{x_j}}, \tag{4.4}$$

Where x_i indicates the i-th class score, K corresponds to the number of classes, i.e., the emotions, while x_j is used to normalize the obtained score over all the available classes. Finally, the model is trained using the categorical cross-entropy loss, which is defined as follows:

$$\mathcal{L}_{CE} = -\sum_{i}^{K} y_i \log \hat{y}_i, \tag{4.5}$$

Where y_i and $\hat{y}_i$ correspond to the ground truth and predicted class probability, respectively, while K is the number of classes.

ConvLSTM classifier

The second classifier, shown in Figure 4.1(b), takes inspiration from methods that combine CNN and RNN models to classify the spatio-temporal characteristics of EEG signals [118, 119], the ConvLSTM model takes a different approach from the initial model, employing three types of convolutions: standard, depthwise, and separable convolutions. These layers are designed to maximize the temporal information encoded in the atlases $\mathcal{A}$, with the resulting feature maps then processed by a bidirectional LSTM layer [206]. This LSTM layer includes both forward and backward layers, allowing it to examine the input from both temporal directions. Independent of the flow direction, the LSTM architecture consists of memory cells with input, forget, and output gates, along with cell and hidden states. Formally, the LSTM at a given time step t and the previous hidden state h_{t-1} is defined as:

$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] + b_i\right), \tag{4.6}$$
$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right), \tag{4.7}$$
$$o_t = \sigma\left(W_o \cdot [h_{t-1}, x_t] + b_o\right), \tag{4.8}$$
$$c_t = f_t \odot c_{t-1} + i_t \odot \hat{c}_t, \tag{4.9}$$
$$h_t = o_t \odot \tanh\left(c_t\right), \tag{4.10}$$

Where σ is the sigmoid activation function; W_i, W_f, W_o, and b_i, b_f, b_o indicate the weight matrices and bias terms for the input, forget, and output gates, respectively; $[h_{t-1}, x_t]$ is the concatenation of the previous hidden state h_{t-1} and the input at time step t; $\odot$ denotes element-wise multiplication; c_t and $\hat{c}_t$ correspond to the updated and candidate cell state at time step t, respectively; while tanh is the hyperbolic tangent activation function. Finally, the model is trained using the same categorical cross-entropy loss described in Eq.(4.5).

ConvGRU classifier

The third classifier, represented in Figure 4.1(c), follows the same structure as the ConvLSTM introduced in Sec. 4.1.2 and is trained using the categorical cross-entropy loss presented in Eq.(4.5). However, instead of employing a bidirectional LSTM layer, ConvGRU leverages the Gated Recurrent Unit (GRU), a different type of recurrent neural network capable of handling temporal data while using fewer parameters than an LSTM, often achieving comparable performance. Specifically, the GRU has a simpler memory cell structure, incorporating update and reset gates along with candidate and hidden states, which are formally defined as follows:

$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t] + b_z\right), \tag{4.11}$$
$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t] + b_r\right), \tag{4.12}$$
$$\hat{h}_t = \tanh\left(W_h \cdot [r_t \odot h_{t-1}, x_t]\right) + b_h, \tag{4.13}$$
$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \hat{h}_t, \tag{4.14}$$

Where σ is the sigmoid activation function; W_z, W_r, W_h, and b_z, b_r, b_h are the weight matrices and bias terms for the update gate, reset gate, and candidate cell state, respectively; $[h_{t-1}, x_t]$ is the concatenation of the previous hidden state h_{t-1}

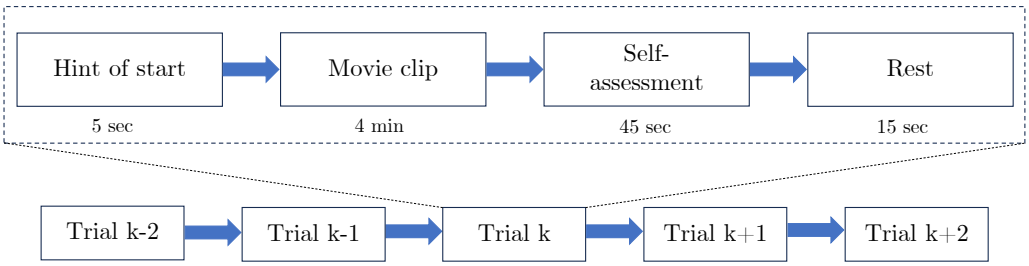


Figure 4.3. Detailed protocol used during experiments on SEED dataset.

and the input at time step t; $\odot$ denotes element-wise multiplication; while tanh represents the hyperbolic tangent activation function.

Vision Transformer

The last classifier, shown in Figure 4.1(d), is a fine-tuned version of the Vision Transformer (ViT) [207]. The ViT model, inspired by the NLP Transformer, splits the input images into patches to provide a sequence of linear embeddings given as input to a transformer, the same way as tokens in NLP applications. Starting from the encoded atlas $\mathcal{A}$ as a 2D image $\mathbf{x} \in \mathbb{R}^{H \times W \times C}$, it is handled into a sequence of flattened squared patches $\mathbf{x}_p \in \mathbb{R}^{N \times (P^2 \cdot C)}$, where H, W are the image height and width, respectively, C represents the number of channels, P is the dimension of each patch, and $N = \frac{HW}{P^2}$ is the total number of patches. To embed each patch into the model dimension D, a trainable linear projection E is applied, thus obtaining an embedding sequence $\mathbf{z}_0$, which is defined as:

$$\mathbf{z}_0 = [\mathbf{x}_{class}; \mathbf{x}_p^1\mathbf{E}; \mathbf{x}_p^2\mathbf{E}; ...; \mathbf{x}_p^N\mathbf{E}] + \mathbf{E}_{pos}, \tag{4.15}$$

where $E \in \mathbb{R}^{(P^2 \cdot C) \times D}$, and $E_{pos} \in \mathbb{R}^{(N+1) \times D}$. Furthermore, a learnable 1D position embedding is added to each patch embedding to retain positional information, then a learnable class embedding ($\mathbf{z}_0^0 = \mathbf{x}_{class}$) is prepended to the patches sequence, representing the image label $\mathbf{y}$ (Eq. 4.18) at the output ($\mathbf{z}_L^0$) of the L-layers transformer encoder. The resulting sequence of embedding vectors is the input of the transformer. The transformer uses a constant latent vector of size D and is composed of alternating layers of Multiheaded Self-Attention (MSA) [208] and a two-layer perceptron, both preceded by a Layernorm (LN) layer and followed by a residual connection. Finally, a classification head, implemented as a linear layer, is attached to $\mathbf{z}_L^0$. Formally, the transformer blocks are defined as follows:

$$\mathbf{z}_t = \text{MSA}\left(LN\left(\mathbf{z}_{\ell-1}\right)\right) + \mathbf{z}_{\ell-1}, \tag{4.16}$$

$$\mathbf{z}_\ell^{'} = \text{MLP}\left(LN\left(\mathbf{z}_\ell^{'}\right)\right) + \mathbf{z}_\ell, \tag{4.17}$$

$$\mathbf{y} = \text{LN}\left(\mathbf{z}_L^0\right), \tag{4.18}$$

where $\ell = 1 \ ... \ L$. Finally, the model is trained using the same categorical cross-entropy loss described in Eq.(4.5).

Chapter 5

EEG limbs activation

This chapter describes the architecture designed to solve the limb activation classification task from the EEG signal. Initially, an overview of the proposed method is presented, then the system architecture and the classifier are introduced in detail.

5.1 Proposed method

The proposed deep architecture is a CNN that approximates the function m: $\mathbb{R}^{b \times s \times c_{in}} \rightarrow R^{b \times |L|}$, which maps each EEG in b to a probability distribution. In detail, we have:

- $b \in N \geq 1$ represents the number of EEG trials in each batch given as input to the CNN during the training and inference phases. As pointed out in [209], this hyperparameter must be accurately chosen based on the problem being treated and the available hardware since it directly affects the performance, both in means of computational costs and of classification result;

- $s \in N \geq 1$is the number of samples for each EEG trial calculated as $s = \lfloor r \cdot t \rfloor$, where $r \in \mathbb{R} > 0$ indicates the sampling rate, expressed in Hertz (Hz), and $t \in \mathbb{R} > 0$ the recording time in seconds (s). Each EEG trial used in the training and testing phases for this work has a fixed length;

- $c_{in} \in \mathbb{N} \geq 1$ is the number of EEG channels recorded during the experiments. This parameter mainly depends on the device used for recording the data. In this work, are used all the available data for each trial;

- L is the set of classes (also called labels) of each trial in the used dataset, corresponding to $\{A_l, A_r\}$ for the PhysioNet dataset, where A_l and A_r mean left or right arm activation, respectively. Each trial is labelled either A_l or A_r.

The proposed model m is composed of two primary sub-modules: a set of convolutional streams, m_s, detailed in Sec. 5.1.1, which are designed to extract diverse features from the input, and the classifier, m_c, discussed in Sec. 5.1.2, that outputs a tensor of probabilities (in the form of logits) for L.

5.1.1 Convolutional Streams

A convolutional stream $m_{si} \in m_s$ consists of one or more convolutional blocks, each comprising 1D convolutional layers. All convolutional streams share the same architecture, differing only by the parameter k, which specifies the kernel size for each stream. The purpose of varying kernel sizes is to enable feature extraction from each channel at multiple temporal scales. Let the kernel size of the first stream be k_1. For each subsequent stream i, the kernel size is defined as $k_i = k_{i-1} + 2$. For instance, if $k_1 = 3$, then $k_2 = 5, k_3 = 7, k_4 = 9$, and so on, following $k_i = k_{i-1} + 2$. Each convolutional block contains two 1D convolutional layers with kernel size k, padding $\lfloor \frac{k}{2} \rfloor$, stride 1, and a ReLU activation function applied after each layer. Every convolutional block, except the final one, includes an additional 1D convolutional layer with stride 2 to reduce sample length. Following the last convolutional block, a 1D Adaptive Max Pooling layer standardizes the input length while maintaining the number of channels, and a flattening layer reshapes the resulting feature map for classifier compatibility. A convolutional stream approximates the function $m_{si} : \mathbb{R}^{b \times s \times c_{in}} \to \mathbb{R}^{b \times (c_{out} \cdot p)}$, where $c_{out} = 64$ is the number of output channels of the final convolution, and $p = 48$ is the final output dimension of the Adaptive Max Pooling layer. In this model, the optimal number of streams is 4, with kernel sizes k set to 7, 9, 11, and 13 for m_{s1}, m_{s2}, m_{s3}, and m_{s4}, respectively. Further parameter details are in Table 5.1. Figure 5.1 depicts the proposed multi-stream CNN architecture.

Layer	Input channels	Output channels	Stride	Padding
1D Conv.	c_{in}	256	1	$\lfloor \frac{k}{2} \rfloor$
ReLU	-	-	-	-
1D Conv.	256	256	1	$\lfloor \frac{k}{2} \rfloor$
ReLU	-	-	-	-
1D Conv.	256	256	2	$\lfloor \frac{k}{2} \rfloor$
1D Conv.	256	127	1	$\lfloor \frac{k}{2} \rfloor$
ReLU	-	-	-	-
1D Conv.	127	64	1	$\lfloor \frac{k}{2} \rfloor$
ReLU	-	-	-	-
1D Adap. Max Pooling	64	64	-	-

Table 5.1. Layers and parameters for each stream of the proposed model, where the output of each stream is in $\mathbb{R}^{b,64,48}$ and the kernel size is k, that is equal to 7, 9, 11 and 13 for m_{s_1}, m_{s_2}, m_{s_3}, and m_{s_4}, respectively.

5.1.2 Classifier

The classifier block m_c consists of a series of fully connected layers that take as input the concatenated, flattened feature maps generated by the convolutional blocks from each stream $m_{si} \in m_s$. Each fully connected layer is preceded by a dropout layer, which mitigates excessive neuron co-adaptation by randomly dropping each neuron and its connections with a fixed probability during training. Additionally, all linear layers except the final one are followed by a ReLU activation function to introduce

Layer	Input features	Output features	Dropout %
Dropout	-	-	50%
Linear	$64 \cdot 48 \cdot 4$	6,145	-
ReLU	-	-	-
Dropout	-	-	50%
Linear	6,145	$\|L\|$	-

Table 5.2. Layers and parameters for the classifier of the proposed model, where L is the set of labels.

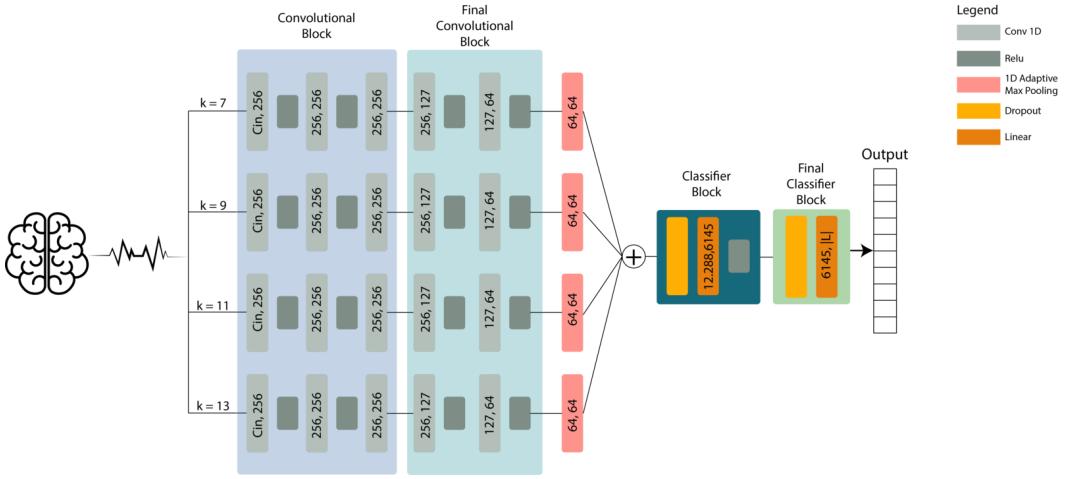


Figure 5.1. The architecture of the proposed model. The latter is composed of four CNN streams, which are defined in Table 5.1, and a set of fully connected layers, which are defined in Table 5.2

nonlinearity. The classifier block approximates the function $m_c : \mathbb{R}^{b\times(c_{out}\cdot p\cdot|m_s|)} \rightarrow \mathbb{R}^{b\times|L|}$. The parameters selected for the model in the results are listed in Table 5.2.

Chapter 6

Experimental results

This chapter reports on the implementation details and experimental results of all EEG applications presented and discussed. For person biometric identification the PhysioNet and M3CV datasets have been used. The proposed limb activation system makes use of the motor imagery subset of Physionet, while, for emotion recognition, a compressed version of the SEED dataset has been deployed to perform all experiments. For each task, an extensive set of experiments has been executed to assess the performance of each system against SOTA works.

6.1 EEG Biometric identification

6.1.1 Dataset

The proposed method has been evaluated on the public benchmarks PhysioNet EEG Motor Movement/Imagery Dataset [210] and M3CV EEG-based biometric competition dataset [211]. PhysioNet consists of over 1500 one- and two-minute EEG recordings, obtained from 109 subjects using a BCI2000 system equipped with 64 wet electrodes arranged according to the 10-10 system [212] at a sampling rate of 160Hz. Each subject was acquired for a total of 14 trials in four different human states: one trial in resting state with eyes-open (EO), one trial in resting state with eyes-closed (EC), six physical movements trials (PHY) in which the subjects are required to move their fists or feet, and six motor imagery movements trials (IMA) in which the subjects are required to imagine moving their fists or feet without any physical movements. Two one-minute baseline runs (one with eyes open, one with eyes closed), and three two-minute runs of each of the four following tasks:

- A target appears on either the left or the right side of the screen. The subject opens and closes the corresponding fist until the target disappears. Then, the subject relaxes.
- A target appears on either the left or the right side of the screen. The subject imagines opening and closing the corresponding fist until the target disappears. Then, the subject relaxes.
- A target appears on either the top or the bottom of the screen. The subject opens and closes either both fists (if the target is on top) or both feet (if the

target is on the bottom) until the target disappears. Then, the subject relaxes.

- A target appears on either the top or the bottom of the screen. The subject imagines opening and closing either both fists (if the target is on top) or both feet (if the target is on the bottom) until the target disappears. Then, the subject relaxes.

Out of the 109 subjects included in the dataset, 4 were excluded. Specifically, subjects 88, 92, and 100 were sampled at 128 Hz instead of 160 Hz, while subject 106 had only a partial recording in some trials. The pre-processing described in Section 3.1.1 has been adopted only to these 105 subjects, resulting in 1672 frames each for the experimental phase. M3CV consists of over 12000 epochs of EEG signals obtained from 106 subjects using a BrainAmp EEG amplifier and Easycap multichannel EEG caps equipped with 64 electrodes arranged according to the standard 10–20 system at a sampling rate of 250 Hz. Each subject was acquired for a total of 14 trials in different human states: two trials in resting state with eyes-open (EO), two trials in resting state with eyes-close (EC), four movements trials (PHY) in which the subjects are required to move their hands or ankle, and six evoked potentials trials. The whole set for this competition was divided into three parts: Enrollment Set, Calibration Set, and Testing set. The Enrollment Set provides the 1st session of EEG from 95 subjects, the Calibration set provides the EEG data of the 2nd session from 20 subjects, and the Testing set is hidden and used to evaluate the performance of the algorithm in the competition. In the present work, only the EO, EC and PHY states are used from the enrollment set, and the preprocessing described in Section 3.1.1 has been adopted only to these 95 subjects, resulting in 464 frames each for the experimental phase.

6.1.2 Implementation details

The proposed system was implemented using the PyTorch [213] and PyTorch Geometric (PyG) [214] frameworks. All experiments were performed using 80/10/10% splits for training, validation, and test sets, respectively. All models were trained using the Adam optimizer [215] for 100 epochs using the same hyper-parameters, i.e., 0.05 learning rate, 10^{-12} weight decay, and a batch size of 200. The experiments were executed using an AMD EPYC 7301 16-core Processor with 64 GB of RAM and an RTX QUADRO 6000 GPU with 24 GB of RAM.

6.1.3 Experimental protocol

For EEG-based person identification technology to be considered realistic and viable, it is crucial to ensure the stability and robustness of the system. This entails that the model should consistently and accurately recognize individuals based on their EEG signals, even when they are in different emotional states (hereinafter "states"), such as feeling happy, sad, or engaged in focused thinking. A series of different experiments were carried out to assess the effectiveness and practicality of the proposed system in EEG biometrics. PhysioNet and M3CV datasets provide EEG signals in the EO, EC, PHY, and IMA states and are used to assess the system's generalization capability. Three experiments have been devised, centered around these four states

to evaluate the performance of the system in diverse scenarios. The details of the experiments are outlined below.

Experiment 1 - Single human state: in this first experiment the system is trained and tested only on EEG signals related to a single fixed state among EO, EC, PHY, IMA. A five-fold cross-validation scheme with shuffling was applied for evaluation.

Experiment 2A/2B - Rest vs Activity: In these experiments, the system is trained solely on resting signals (EO/EC) and tested only on activity-related signals (PHY/IMA), and vice versa. These experiments were carried out to determine whether the system, trained only with resting-state EEG signals, is capable of generalizing over activity-related signals and vice versa. In these experiments, the dataset has been randomly divided into 4:1 as a training set and test set, respectively.

Experiment 3 - Mixed states: In these experiments, the system is trained and tested on a mixture of resting signals (EO/EC) and activity-related signals (PHY/IMA). This experiment has been designed to observe how the system performs in a realistic scenario where the signals to be classified are not specific to a particular human state but can vary, emulating what happens in the human brain where, even without external stimuli, the states might sudden alternate due to emotions or intrusive thoughts. A five-fold cross-validation scheme with shuffling was applied for evaluation.

Experiment 4 - Segmentation length: In EEG identification methods, a consensus regarding the optimal segmentation length has not yet been reached. Wang et al. [91] and Du et al. [87], for instance, adopted a segmentation length of 1 second with a 50% overlap in their works. In contrast, Thiago Schons et al. [100] employed a 12-second segmentation length with a 1% overlap. The variance in sample segmentation lengths among different methodologies can result in substantial differences. To address this, are conducted evaluations on distinct dataset splits, each with lengths of 1s, 5s, and 12s, and overlaps of 50% and 80%, respectively. This experiment aims to explore how much the system performance is affected by the chosen segmentation length.

Table 6.1. Experiments results - Training and testing within each human state. Results are reported as test stage accuracy (average ± standard deviation of 5-fold cross-validation)

State	PhysioNet	M3CV
EO → EO	99.90 ± 0.09%	99.82 ± 0.22%
EC → EC	99.98 ± 0.03%	99.59 ± 0.32%
PHY → PHY	99.98 ± 0.01%	99.97 ± 0.01%
IMA → IMA	99.97 ± 0.01%	-
MIX → MIX	99.98 ± 0.02%	99.05 ± 0.06%
PHY → REST	99.54 ± 0.01%	97.96 ± 0.00%
IMA → REST	99.50 ± 0.02%	-
REST → PHY	96.70 ± 0.00%	97.40 ± 0.01%
REST → IMA	97.50 ± 0.00%	-

Table 6.2. SOTA comparison on PhysioNet Dataset on single states. Results are reported as test stage accuracy (average ± standard deviation of 5-fold cross-validation)

Method	EO → EO	EC → EC	PHY → PHY	IMA → IMA
FuzzEn + SVM [91]	84.14 ± 0.83%	83.73 ± 0.71 %	77.93 ± 0.59%	80.84 ± 0.18%
Raw + CNN [91]	96.89 ± 0.77%	67.43 ± 47.36 %	97.96 ± 1.55%	97.42 ± 0.83%
Graph + Mahalobis [86]	99.07 ± 0.19%	97.56 ± 0.24%	99.74 ± 0.13%	99.61 ± 0.11%
ConvLSTM [83]	98.00%	99.95%	-	-
COH + Mahalobis [82]	96.26%	97.50%	-	-
Graph VAE [98]	99.78 ± 0.04%	99.04 ± 1.69%	99.65 ± 0.08%* 99.19 ± 0.23%*	99.62 ± 0.05%* 99.53 ± 0.11%*
PLV + ChebConv [91]	99.97 ± 0.03%	99.88 ± 0.03%	**99.99 ± 0.02%**	**100.00 ± 0.00%**
Lite Transformer [87]	83.77 ± 17.39%	85.57 ± 22.24 %	99.76 ± 0.01%	99.65 ± 0.08
EA-Transformer [87]	98.45 ± 1.17%	98.19 ± 3.14%	99.93 ± 0.00%	99.90 ± 0.01
TST [87]	**100.00 ± 0.00%**	99.96 ± 0.06%	99.97 ± 0.01%	**100.00 ± 0.00%**
Our	99.90 ± 0.09%	**99.98 ± 0.03%**	99.98 ± 0.01%	99.97 ± 0.01%

* Behrouzi et al. divide the IMA and PHY datasets into two subsets: IMA/PHY-LR (left or right fist only) and IMA/PHY-B (both fists).

Table 6.3. SOTA comparison on PhysioNet Dataset Training on resting states and testing on activity states. Results are reported as test stage accuracy (average ± standard deviation of 5-fold cross-validation)

Method	REST→ PHY	REST→ IMA
FuzzEn + SVM [91]	16.16 ± 0.01%	15.61 ± 0.00%
Raw + CNN [91]	49.26 ± 3.85%	52.51 ± 2.26%
Graph + Mahalanobis [86]	69.98 ± 0.38%	69.47 ± 0.64%
ConvLSTM [83]	-	-
COH + Mahalanobis [82]	-	-
Graph VAE [98]	-	-
PLV + ChebConv [91]	85.40 ± 1.62%	87.03 ± 2.53%
Lite Transformer [87]	87.37 ± 1.10%	89.03 ± 0.73%
EA-Transformer [87]	89.47 ± 0.34%	90.66 ± 0.39%
TST [87]	**97.29 ± 0.03%**	97.45 ± 0.13%
Our	96.70 ± 0.00%	**97.50 ± 0.00%**

6.1.4 Results

Each experiment described in 6.1.3 was performed on both PhysioNet and M3CV datasets in order to test the robustness of the proposed model on signals acquired with different protocols and both on dry and wet headsets. Subsequently, in order to compare the performance of the proposed system also with other state-of-the-art models, an extensive comparison with the above was performed for each experiment. At present, methods on EEG person identification exploit different features and classifiers, varying from classical linear models to most advanced deep architecture. In

Table 6.4. SOTA comparison on PhysioNet Dataset training on mixed states and testing on mixed states. Results are reported as test stage accuracy (average ± standard deviation of 5-fold cross-validation)

Method	MIX → MIX
FuzzEn + SVM [91]	73.45 ± 0.10%
Raw + CNN [91]	99.85 ± 0.06%
Graph + Mahalanobis distance [86]	96.22 ± 0.23%
ConvLSTM [83]	-
COH + Mahalanobis [82]	-
Graph VAE [98]	-
PLV + ChebConv [91]	99.98 ± 0.02%
Lite Transformer [87]	98.16 ± 0.65%
EA-Transformer [87]	99.90 ± 0.01%
TST [87]	99.90 ± 0.03%
Our	**99.98 ± 0.02%**

the experiments, are used as baselines a set of classical approaches, e.g., Convolutional-LSTM, denoted as 'ConvLSTM', SVM with fuzzy entropy features (FuzzEn), denoted as 'FuzzEn+SVM', CNN with raw EEG time-series inputs, denoted as 'Raw+CNN', and Mahalanobis discriminator with COH features, denoted as 'COH+Mahalanobis', in conjunction with State-of-the-art methods, such as graph-based approaches, e.g., Graph Variational Autoencoder (VAE), denoted as 'Graph VAE', Chebyshev Convolutional GNN with PLV features denoted as 'PLV+ChebConv', and different recent transformer-based approaches which combine CNN and attention mechanism, e.g., EA-Transformer, denoted respectively as 'EA-Transformer', 'Lite Transformer', and 'TST'. Finally, the proposed method performance is compared against the baseline results for each experiment reported in Section 6.1.3.

6.1.5 Evaluation and State-Of-The-Art comparison

The results of each method for experiments 1, 2a, 2b, and 3 are summarized in Table 6.1, while experiment 4 results are reported in Figure 6.1. In the first experiment, the model was trained and tested on the same human state, achieving an average accuracy above 99% in each state on PhysioNet, while on the M3CV dataset, an accuracy of 99.82% is reached for state EO, 99.59% for state EC, and 99.90% for state PHY. In this experiment, the presented model had a very high identification performance for each single state, both on PhysioNet and M3CV datasets. In experiment 2a, since the training and testing were performed using signals in different states, the performances on PhysioNet and M3CV are slightly degraded compared to other experiments. However, the system was able to generalize over diverse states using abstracts learned from resting states to classify PHY and IMA state samples. In particular, on the M3CV dataset, the model trained on resting state data and tested

on PHY data is able to achieve an average accuracy of 98.30%. While on PhysioNet, the model tested on PHY and IMA data is able to reach an average accuracy of 96.70% and 97.50%, respectively. Experiment 2b is introduced for the first time in this work, in fact none of the SOTA work have ever trained their models on activity states and tested on resting state. This experiment highlights that the proposed model has a stable and reliable performance also in this scenario, achieving excellent performance both on PhysioNet and M3CV. In particular, on the M3CV dataset, the model trained on PHY data and tested on resting data is able to achieve an average accuracy of 98.20%. While on PhysioNet, the model trained on PHY and IMA data is able to reach an average accuracy of 99.54% and 99.50%, respectively. This experiment can provide another method of comparison for future works in order to understand how the proposed model can generalize on unseen data. In experiment 3, where signals in diverse states were mixed, the proposed model achieved an average accuracy of 99.98% on PhysioNet and 99.05% on M3CV, confirming the robustness already emerged from other experiments. Finally, experiment 4 compares different segmentation lengths and overlaps percentages to find the best window. As reported in 6.1, six different configurations have been involved in this experiment, three window lengths of size 1s, 5s, and 12s, respectively, and two overlap ratios for the moving window of size 50% and 80%. From the results, it is clear that a 1s or 5s moving window with a 50% overlap ratio is sufficient to provide enough training data for the model to generalize. Moving toward a higher overlap ratio doesn't lead to better performances but rather implies a longer training time because of the increased amount of data. On the other side, increasing the window length to 12s leads to huge performance degradation, not allowing a proper generalization for the model caused by the reduced amount of training data. Indeed, 1s and 5s options yields similar performance, the experimental results show that the best one is the 1s 50% overlap option. In experiment 1, the performance of the proposed model is investigated in the same single state. The system has been trained and tested on a single-state subset of PhysioNet and M3CV. The results are shown in Table 6.2. The experimental outcomes demonstrate that the presented method exhibits state-of-the-art performance on PhysioNet in each state compared to all other methods in the current literature when the data remains within the same state. Experiment 2a was performed in order to observe how robust the proposed approach to state changes is and to assess its generalization capability. Therefore, the model has been trained only on EO and EC data and tested only on PHY or IMA data. The obtained results are compared with SOTA in Table 6.3. Differently from experiment 1, experiment 2 entails training and testing the models across diverse states, using data acquired in different states under different conditions. This leads to a degradation of most methods in the SOTA, especially those relying on univariate features. In general, all the other methods degraded their performance, Graph-based methods drop their accuracy by around 15-30%, Lite and EA-Transformer by around 10 to 15%, CNN by up to $\sim$ 50%, and SVM by up to $\sim$ 65%. The proposed method has a much lower degradation compared to almost all the other, dropping only 3.28% and 2.47% in PHY and IMA, respectively. Compared to [91], the proposed model has an improvement of 11.3% and 10.47%, tested on PHY and IMA, respectively. Compared to Transformer-based approaches, the introduced system outperforms Lite-Transformer and EA-Transformer both on PHY and IMA, finally the TST is

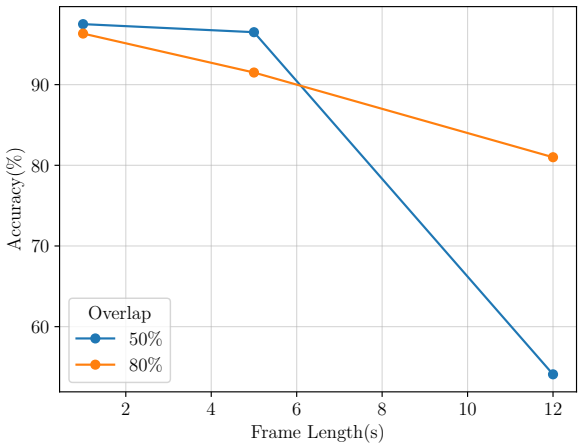


Figure 6.1. Results of the proposed model on different frame length and overlap obtained on PhysioNet following experiment 2a protocol on IMA data.

outperformed on IMA identification of 0.05%, while it presents a slight increase in accuracy on PHY identification of 0.59%. Finally in experiment 3, where signals in diverse states were mixed, the proposed model achieved an averaged accuracy of 99.98%, achieving SOTA performance also in this context. It suggests that the system is robust and effective also in scenarios where multiple states are mixed in the training set and is able to learn to extract features universal to all states, confirming its flexibility and generalization capability. Results from experiment 3 are reported in Table 6.4. Notably, unlike all the other SOTA methods, the proposed system has the same excellent performance on two distinct public datasets that have been acquired with two different acquisition protocols using both dry and wet electrodes and different headsets, a further indication of the robustness and reliability of the proposed model.

6.2 EEG emotion recognition

This section assesses the effectiveness of the proposed approach in EEG emotion recognition. In detail, Sec. 6.2.1 introduces the public dataset used to evaluate the Empátheia system. Sec. 6.2.2 reports the implementation details required to reproduce the experiments. Sec. 6.2.3 examines the performance of the PRISMIN atlas encoder. Finally, Sec. 6.2.4 evaluates the Empátheia classifier through ablation studies and a state-of-the-art comparison.

6.2.1 Dataset

The dataset used to train and test the Empátheia system is the SEED [49], a public collection focusing on the EEG emotion recognition task. The dataset is composed of 15 subjects, 7 males and 8 females, all right-handed students of Shanghai Jiao Tong University. A total of 6 clips were shown to the participants of the experiment, and each clip was associated with negative, neutral, and positive emotions. Finally, each emotion has 5 corresponding emotional clips. Each trial comprehends a 5-second

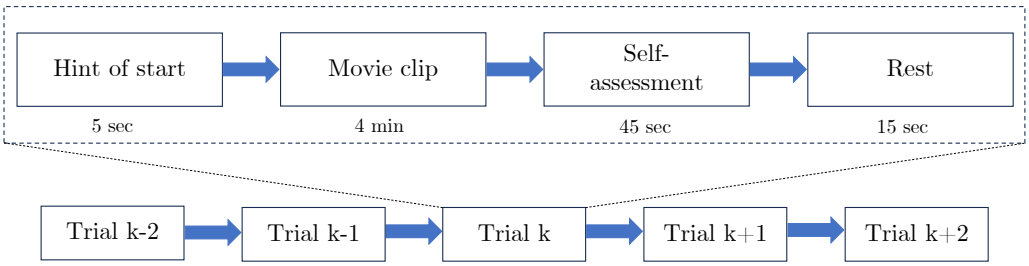


Figure 6.2. Detailed protocol used during experiments on SEED dataset.

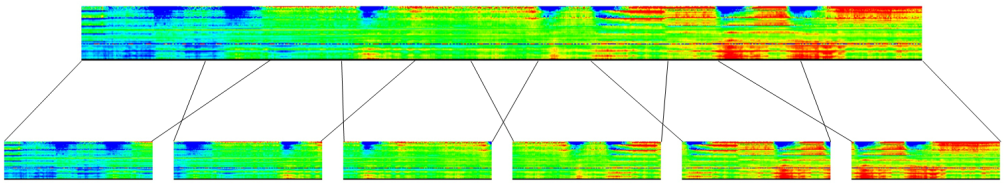


Figure 6.3. Data augmentation strategy example.

hint before each clip, a roughly four-minute-long clip, followed by 45 seconds for self-assessment, and is concluded with 15 seconds of rest. The dataset is provided already pre-processed by the authors. In detail, the original EEG data was downsampled to a sampling rate of 200 Hz. Then, visual inspection of the data was performed on the new EEG signals, and recordings significantly affected by EMG and EOG interferences were manually excluded. EOG data, recorded during the experiments, was also utilized to identify blink artifacts within the recorded EEG data. To mitigate noise and remove artifacts, the EEG data has been processed using a bandpass filter with a range of 0.3 to 50 Hz. Subsequently, to the preprocessing process, EEG segments corresponding to the duration of each movie were extracted. Each channel of the EEG data was then divided into non-overlapping epochs of equal length, i.e., 1 second. For a single experiment, approximately 3300 clean epochs were obtained. The detailed acquisition protocol for a single trial is summarized in Figure 4.3.

Table 6.5. Dataset reduction using PRISMIN encoding.

$\mathcal{D}$	Baseline	C-Rate	Augmented	C-Rate
$\mathcal{D}_{GS}$	998 MB	6.9x	744 MB	9.2x
$\mathcal{D}_{SR}$	670 MB	10.3x	907 MB	7.6x

Regarding the acquisitions, EEG signals were collected using the 62-channel ESI NeuroScan System at a sampling rate of 1000 Hz, according to the international 10-20 system for 62 channels. Furthermore, the authors down-sampled the signals to 200 Hz and applied a 0-75 Hz bandpass frequency filter. The resulting EEG signal segments correspond to the duration of each clip; therefore, their length can slightly differ when considering distinct segments.

6.2.2 Implementation details

The Empátheia system was implemented using the PyTorch framework[1]. All experiments were performed using 80/10/10% splits for training, validation, and test sets, respectively. All models were trained using the AdamW optimizer for 100 epochs using the same hyper-parameters, i.e., 1e-03 learning rate and a batch size of 64. Standard classification metrics, i.e., accuracy, precision, recall, and F1-score, were employed to assess the system. The experiments were executed using an AMD EPYC 7301 16-Core Processor with 64 GB of RAM and an RTX QUADRO 6000 with 24 GB of RAM.

6.2.3 PRISMIN encoder evaluation

The first component to be evaluated is the PRISMIN encoder tasked with the atlas generation. In particular, since the primary focus of the encoder is to reduce the SEED dataset size, its assessment revolves around the compression rate (C-Rate) of the input dataset. To achieve this goal, this thesis explores the effectiveness of two different encodings, i.e., linear gray-scale (GS) and short-rainbow (SR) mappings, during the implementation of the *refract* and *bake* methods discussed in Sec. 4.1.1. The resulting compressed datasets are reported in Table 6.5. In detail, for each sample trial in the SEED collection, the devised PRISMIN encoder generates a PNG atlas containing either its GS or SR encoding. The resulting atlas represents the entire EEG trial and has a final shape of 62×10800, corresponding to the number of channels and recording length of the captured EEG signals. Note that the width of the generated atlases depends on the corresponding session length. Therefore, to ensure that all image-encoded signals have the aforementioned shape, the atlases are cropped on the left and right sides by up to 10% of their width. These portions are generally less relevant for emotion recognition as they are associated with the start and end of an EEG acquisition. From this procedure, the PRISMIN encoder effectively constructs two new datasets, i.e., $\mathcal{D}_{GS}$ and $\mathcal{D}_{SR}$, corresponding, respectively, to the grey-scale and short-rainbow encodings, that both maintain the number of samples of the original SEED dataset, i.e., 675. However, due to the performed compression, the resulting collections manage to significantly reduce the original dataset size by a factor of 10.3x and 6.9x, attesting to a disk space of 670 MB and 998 MB for the GS and SR encoding, respectively.

The direct conversion of the SEED dataset using the described PRISMIN encoder effectively reduces the collection size. Despite that, the generated atlases depict a coarse representation of the original EEG signals and do not allow deep learning models to learn the emotion recognition task. What is more, even works analyzing EEG signals from the SEED dataset tend to suffer from this issue, which is generally addressed by applying a data augmentation strategy through slicing of the original recordings [110, 112]. Thus, following this rationale, starting from the beginning of the generated atlases, they are sequentially split into smaller ones by extracting sub-frames of shape 62×1000, as depicted in Figure 6.3, with a 200 pixels overlap on the x axis among subsequent sub-frames. This approach, already used in different literature works [109, 110, 112], is adopted to partially preserve middle information

[1]Source code is available at: `https://github.com/Prometheus-Laboratory/2024_prismin`

among subsequent frames, limiting emotions cut-off. With this configuration, the resulting $\mathcal{D}_{GS}$ and $\mathcal{D}_{SR}$ datasets, used to evaluate the Empátheia classifiers, contain 9450 samples instead of 675 and enable the implemented models to perform emotion recognition, as reported in Sec. 6.2.4.

Even after the reported data augmentation strategy, $\mathcal{D}_{GS}$ and $\mathcal{D}_{SR}$ still considerably reduce the original SEED dataset size by factors of 9.2x and 7.6x for the GS and SR encodings, respectively. Interestingly, for the GS encoding, when considering smaller image portions, the PRISMIN encoder further compresses the atlases, resulting in an even smaller collection. This outcome is possibly due to the reduced amount of noise in the grey-scale sub-frames and suggests that additional encoding strategies can be explored in the future to improve the atlas generation. Summarizing, the PRISMIN encoder generates coarse atlases containing emotions from EEG signals and significantly reduces the SEED dataset size with both GS and SR encodings even when applying a data augmentation strategy, thus satisfying the need for a smaller collection to train different types of deep learning models.

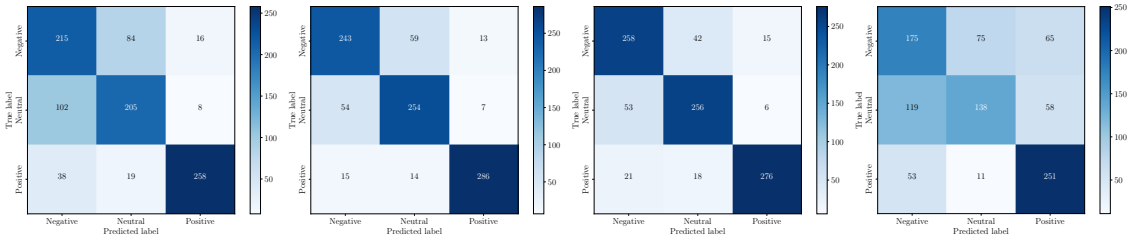
6.2.4 Empátheia classifier evaluation

This section presents ablation studies and a state-of-the-art comparison to demonstrate the effectiveness of the Empátheia system. Specifically, the former is discussed in Sec. 6.2.4, which examines various aspects of the implemented architectures. The literature comparison is presented in Sec. 6.2.4.

Ablation study

A comprehensive set of experiments was conducted to rigorously assess the performance of the Empátheia classifiers when applied to an innovative input source, specifically the PRISMIN-generated atlases. Ablation studies were meticulously carried out on both gray-scale and short-rainbow encodings, namely datasets $\mathcal{D}GS$ and $\mathcal{D}SR$, using various kernel sizes ($K \in \{3, 5, 16\}$), learning rates ($LR \in \{$1e-3, 2e-3, 5e-4$\}$), and, for recurrent models, a range of hidden unit counts ($U \in \{30, 50, 70\}$). In the context of the ViT model, evaluations were executed under differing learning rates ($LR \in \{$1e-4, 1e-5, 1e-6$\}$) and numbers of attention heads ($\mathcal{H} \in \{8, 16, 32\}$), with a patch size of 16×16 and a transformer encoder of 12 layers. The results obtained on the development set are reported in Table 6.13 for the Conv classifier, Table 6.6, Table 6.7, Table 6.8 for the ConvLSTM classifier, Table 6.9, Table 6.10, Table 6.11 for the ConvGRU classifier, and Table 6.12 for the ViT classifier. The best model in each table for the grey-scale encoding is highlighted in green, while the best for the short rainbow is in blue. The decision to utilize a convolutional approach for feature extraction from the atlas was informed by the benchmarking results in Avola et al. [216]. In that benchmark, vanilla CNN, LSTM, and GRU models were evaluated on raw EEG signals, with CNNs demonstrating superior accuracy for this type of data. As seen in the tables, all Empátheia classifiers show strong performance on the emotion recognition task with the generated atlases, achieving accuracies up to 83.5%. Within convolutional-based classifiers, the Conv model displays slightly lower performance compared to the ConvLSTM and ConvGRU architectures, showing accuracy differences of 13.8% and 8.6% on the $\mathcal{D}GS$ and

$\mathcal{DSR}$ datasets, respectively. These findings suggest that recurrent models effectively capture the temporal evolution of emotions within the atlases through their recurrent layers. Moreover, the kernel size in convolutional layers appears to critically impact overall performance. Specifically, wider kernels (i.e., $K = 16$) enable the ConvLSTM and ConvGRU models to achieve superior performance, indicating that temporal information is also effectively captured within their convolutional receptive fields. Additional key hyperparameters influencing model performance include the learning rate and, for recurrent models, the number of hidden units. Learning rate impacts model convergence and exhibits improved metrics when smaller LR values are paired with larger kernels and vice versa. This outcome indicates that a slower progression along the error surface during training, using smaller LR values (e.g., 5e-4), allows the model to thoroughly analyze broad receptive fields in conjunction with temporal information. Conversely, a faster progression, with higher LR values (e.g., 2e-3), enables the model to avoid over-focusing on noise-capturing smaller kernels, given the coarse representation of EEG signals in the atlases. For recurrent architectures, such as ConvLSTM and ConvGRU, hidden unit counts of $U = 50$ or 70 appear optimal for capturing emotion-related details within an atlas, particularly when combined with lower LR values, as illustrated in Table 6.11. Regarding ViT, it underperforms relative to Conv-based architectures, with accuracy reductions of 23.5% and 20.3% compared to ConvGRU on datasets $\mathcal{DGS}$ and $\mathcal{DSR}$, respectively. Nevertheless, the number of attention heads significantly impacts performance. As depicted in Table 6.12, ViT achieves improved results with an increased number of attention heads (i.e., $\mathcal{H} = 32$), suggesting that higher model complexity is beneficial for the proposed encodings. The encoding type further impacts model performance, with Conv classifiers demonstrating better results using short-rainbow encoded atlases from $\mathcal{DSR}$, while recurrent classifiers like ConvLSTM and ConvGRU achieve enhanced metrics on gray-scale mappings, or $\mathcal{DGS}$. This suggests that temporal dynamics of emotions, which recurrent architectures capture more effectively, are better represented through grey-scale encoding. Thus, it can be inferred that employing different encodings could yield significantly varied results depending on the architecture used.



(a) **(b)** **(c)** **(d)**

Figure 6.4. Test set confusion matrices results of Conv (a), ConvLSTM (b), ConvGRU (c), and ViT (d) models.

Regardless of the underlying hyper-parameter configurations, all Empátheia classifiers exhibit stable performance across all classification metrics. Specifically, accuracy, precision, recall, and F1 scores are consistently aligned, suggesting that the models provide balanced performance without bias towards any particular class,

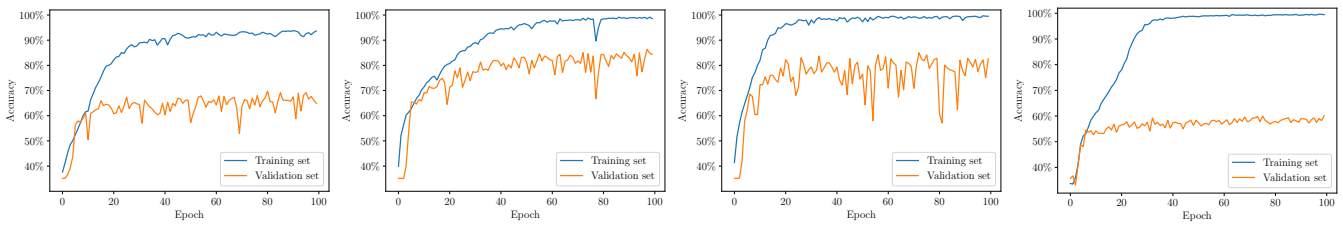
Table 6.6. ConvLSTM LR=1e-3.

$\mathcal{D}$	K	U	Acc	Prec	Recall	F1
$\mathcal{D}_{GS}$	3	30	77.3%	77.5%	77.3%	77.4%
$\mathcal{D}_{SR}$	3	30	73.7%	73.4%	73.8%	73.6%
$\mathcal{D}_{GS}$	3	50	79.1%	79.7%	79.1%	79.4%
$\mathcal{D}_{SR}$	3	50	75.4%	75.5%	75.1%	75.3%
$\mathcal{D}_{GS}$	3	70	80.1%	80.1%	80.1%	80.1%
$\mathcal{D}_{SR}$	3	70	76.4%	75.8%	76.1%	76.0%
$\mathcal{D}_{GS}$	5	30	77.7%	78.5%	77.8%	78.2%
$\mathcal{D}_{SR}$	5	30	74.5%	74.3%	73.9%	74.2%
$\mathcal{D}_{GS}$	5	50	77.5%	78.0%	77.6%	77.8%
$\mathcal{D}_{SR}$	5	50	74.5%	73.8%	73.7%	73.8%
$\mathcal{D}_{GS}$	5	70	74.8%	74.2%	73.9%	74.0%
$\mathcal{D}_{SR}$	5	70	71.6%	70.2%	70.2%	70.2%
$\mathcal{D}_{GS}$	16	30	80.0%	80.6%	80.0%	80.3%
$\mathcal{D}_{SR}$	16	30	76.4%	76.3%	76.0%	76.2%
$\mathcal{D}_{GS}$	16	50	79.7%	79.5%	79.8%	79.6%
$\mathcal{D}_{SR}$	16	50	76.4%	75.3%	75.8%	75.5%
$\mathcal{D}_{GS}$	16	70	82.1%	82.0%	82.1%	82.1%
$\mathcal{D}_{SR}$	16	70	77.6%	77.8%	77.7%	77.8%

Table 6.7. ConvLSTM using LR=2e-3.

$\mathcal{D}$	K	U	Acc	Prec	Recall	F1
$\mathcal{D}_{GS}$	3	30	75.9%	76.4%	76.0%	76.2%
$\mathcal{D}_{SR}$	3	30	62.1%	62.2%	62.1%	62.2%
$\mathcal{D}_{GS}$	3	50	78.6%	78.6%	78.6%	78.6%
$\mathcal{D}_{SR}$	3	50	69.5%	69.5%	69.4%	69.5%
$\mathcal{D}_{GS}$	3	70	77.0%	77.2%	77.0%	77.1%
$\mathcal{D}_{SR}$	3	70	68.1%	68.3%	68.0%	68.2%
$\mathcal{D}_{GS}$	5	30	75.6%	75.5%	75.7%	75.6%
$\mathcal{D}_{SR}$	5	30	66.8%	66.8%	66.9%	66.8%
$\mathcal{D}_{GS}$	5	50	75.0%	75.0%	75.0%	75.0%
$\mathcal{D}_{SR}$	5	50	66.3%	66.3%	66.3%	66.3%
$\mathcal{D}_{GS}$	5	70	77.8%	78.2%	77.9%	78.0%
$\mathcal{D}_{SR}$	5	70	68.8%	69.2%	68.8%	69.0%
$\mathcal{D}_{GS}$	16	30	74.0%	74.3%	74.1%	74.2%
$\mathcal{D}_{SR}$	16	30	65.4%	65.7%	65.5%	65.6%
$\mathcal{D}_{GS}$	16	50	77.2%	77.8%	77.2%	77.5%
$\mathcal{D}_{SR}$	16	50	68.3%	68.8%	68.2%	68.5%
$\mathcal{D}_{GS}$	16	70	76.8%	76.6%	76.8%	76.7%
$\mathcal{D}_{SR}$	16	70	73.0%	73.2%	73.0%	73.1%

effectively minimizing false positives and accurately identifying positive instances for each emotion. This indicates that the learned weights are robust and can accurately abstract the emotion represented in the PRISMIN-extracted atlas. This outcome is also observable in Figure 6.5 and Figure 6.4, which depict, respectively, accuracy curves during training using the best hyper-parameters for each model and the confusion matrices computed on the test set. As shown in Figure 6.5, recurrent models maintain high performance on the development set throughout their training. Conversely, the ViT and Conv classifiers exhibit more pronounced overfitting, leading to reduced performance metrics compared to the other models. Specifically regarding CNNs, this behavior becomes more apparent in the confusion matrices related to the



(a) **(b)** **(c)** **(d)**

Figure 6.5. Training set accuracies results of Conv (a), ConvLSTM (b), ConvGRU (c), and ViT (d) models.

Table 6.8. ConvLSTM using LR=5e-4.

$\mathcal{D}$	K	U	Acc	Prec	Recall	F1
$\mathcal{D}_{GS}$	3	30	80.6%	80.7%	80.6%	80.7%
$\mathcal{D}_{SR}$	3	30	73.9%	74.2%	74.0%	74.1%
$\mathcal{D}_{GS}$	3	50	80.6%	80.5%	80.6%	80.6%
$\mathcal{D}_{SR}$	3	50	76.3%	74.4%	74.4%	74.4%
$\mathcal{D}_{GS}$	3	70	80.5%	80.9%	80.7%	80.8%
$\mathcal{D}_{SR}$	3	70	75.5%	74.8%	74.5%	74.6%
$\mathcal{D}_{GS}$	5	30	78.7%	78.4%	78.7%	78.6%
$\mathcal{D}_{SR}$	5	30	74.4%	72.5%	72.7%	72.6%
$\mathcal{D}_{GS}$	5	50	82.8%	83.0%	82.9%	82.9%
$\mathcal{D}_{SR}$	5	50	76.4%	76.7%	76.6%	76.6%
$\mathcal{D}_{GS}$	5	70	76.1%	76.2%	76.2%	76.2%
$\mathcal{D}_{SR}$	5	70	71.6%	70.4%	70.4%	70.4%
$\mathcal{D}_{GS}$	16	30	79.6%	80.0%	79.7%	79.8%
$\mathcal{D}_{SR}$	16	30	75.4%	73.9%	73.6%	73.7%
$\mathcal{D}_{GS}$	16	50	80.0%	80.7%	80.0%	80.3%
$\mathcal{D}_{SR}$	16	50	75.4%	74.6%	73.9%	74.2%
$\mathcal{D}_{GS}$	16	70	81.7%	81.8%	81.8%	81.8%
$\mathcal{D}_{SR}$	16	70	76.0%	76.1%	76.1%	76.1%

Table 6.9. ConvGRU using LR=1e-3.

$\mathcal{D}$	K	U	Acc	Prec	Recall	F1
$\mathcal{D}_{GS}$	3	30	71.7%	75.3%	71.7%	73.5%
$\mathcal{D}_{SR}$	3	30	67.8%	68.1%	67.8%	68.0%
$\mathcal{D}_{GS}$	3	50	76.5%	77.7%	76.5%	77.1%
$\mathcal{D}_{SR}$	3	50	73.3%	73.6%	74.3%	73.4%
$\mathcal{D}_{GS}$	3	70	80.3%	80.7%	77.8%	79.2%
$\mathcal{D}_{SR}$	3	70	77.5%	76.5%	75.5%	75.4%
$\mathcal{D}_{GS}$	5	30	79.2%	79.5%	79.3%	79.4%
$\mathcal{D}_{SR}$	5	30	76.3%	75.3%	77.0%	75.6%
$\mathcal{D}_{GS}$	5	50	79.7%	79.7%	79.8%	79.8%
$\mathcal{D}_{SR}$	5	50	77.0%	75.5%	77.5%	76.0%
$\mathcal{D}_{GS}$	5	70	77.1%	78.0%	77.1%	77.6%
$\mathcal{D}_{SR}$	5	70	74.3%	73.9%	74.9%	73.9%
$\mathcal{D}_{GS}$	16	30	79.3%	79.6%	79.4%	79.5%
$\mathcal{D}_{SR}$	16	30	76.3%	75.4%	77.1%	75.5%
$\mathcal{D}_{GS}$	16	50	77.7%	77.8%	77.8%	77.8%
$\mathcal{D}_{SR}$	16	50	75.3%	73.7%	75.5%	74.1%
$\mathcal{D}_{GS}$	16	70	75.5%	76.6%	75.6%	76.1%
$\mathcal{D}_{SR}$	16	70	75.3%	75.9%	75.3%	75.6%

test set, where the Conv classifier commits more errors. Concerning ViT, as pointed out by Khan et al. [217], transformer architectures lack inherent inductive biases (prior knowledge) for processing visual data. They typically require large amounts of training data to discern the underlying modality-specific rules. This increased complexity results from the larger number of parameters, as shown in Table 6.14. For this reason, the ViT model presents the highest overfitting value (approximately 40%), requiring significantly more data to generalize effectively. In fact, unlike CNN-based models equipped with built-in translation invariance, weight sharing, and partial scale invariance, transformer networks must autonomously infer these image-specific concepts from the provided training examples. Based on the experimental results, it is evident that the ViT architecture is unsuitable for the objectives of this work. Specifically, compared to other models, ViT not only exhibits the lowest classification accuracy but also requires substantial computational resources for parameter and weight management, as reported in 6.14. Interestingly, all models tend to misclassify samples predominantly associated with negative and neutral emotions. This observation suggests that these categories share common patterns in the generated atlases, indicating that further exploration of the atlas generation by the PRISMIN encoder might enhance the final classification performance. Finally, Table 6.14 compares the best configurations among the reported ablation studies to underline the effectiveness of the devised solutions in classifying the GS and SR encodings. As can be observed, recurrent models achieve the highest performance,

Table 6.10. ConvGRU LR=2e-3.

$\mathcal{D}$	K	U	Acc	Prec	Recall	F1
$\mathcal{D}_{GS}$	3	30	76.7%	76.5%	76.7%	76.6%
$\mathcal{D}_{SR}$	3	30	71.9%	71.4%	72.0%	71.7%
$\mathcal{D}_{GS}$	3	50	73.9%	74.6%	74.0%	74.3%
$\mathcal{D}_{SR}$	3	50	67.1%	67.8%	67.0%	67.4%
$\mathcal{D}_{GS}$	3	70	66.1%	69.7%	66.1%	67.9%
$\mathcal{D}_{SR}$	3	70	59.8%	63.4%	59.8%	61.6%
$\mathcal{D}_{GS}$	5	30	73.0%	72.8%	73.0%	72.9%
$\mathcal{D}_{SR}$	5	30	66.2%	66.2%	66.1%	66.1%
$\mathcal{D}_{GS}$	5	50	73.8%	73.7%	73.9%	73.8%
$\mathcal{D}_{SR}$	5	50	67.1%	67.0%	66.9%	66.9%
$\mathcal{D}_{GS}$	5	70	73.0%	74.9%	73.0%	73.9%
$\mathcal{D}_{SR}$	5	70	66.2%	68.1%	66.1%	67.0%
$\mathcal{D}_{GS}$	16	30	81.9%	82.5%	81.9%	82.2%
$\mathcal{D}_{SR}$	16	30	74.3%	75.0%	74.2%	74.5%
$\mathcal{D}_{GS}$	16	50	82.8%	83.0%	82.9%	82.9%
$\mathcal{D}_{SR}$	16	50	75.1%	75.5%	75.1%	75.2%
$\mathcal{D}_{GS}$	16	70	81.7%	82.1%	81.8%	82.0%
$\mathcal{D}_{SR}$	16	70	71.6%	72.9%	71.6%	72.2%

Table 6.11. ConvGRU LR=5e-4.

$\mathcal{D}$	K	U	Acc	Prec	Recall	F1
$\mathcal{D}_{GS}$	3	30	82.2%	82.5%	82.2%	82.4%
$\mathcal{D}_{SR}$	3	30	73.4%	73.7%	73.4%	73.6%
$\mathcal{D}_{GS}$	3	50	79.6%	80.9%	79.7%	80.3%
$\mathcal{D}_{SR}$	3	50	72.4%	79.4%	79.6%	79.5%
$\mathcal{D}_{GS}$	3	70	83.5%	83.9%	83.6%	83.7%
$\mathcal{D}_{SR}$	3	70	80.3%	78.1%	76.6%	77.3%
$\mathcal{D}_{GS}$	5	30	77.8%	78.0%	76.8%	77.4%
$\mathcal{D}_{SR}$	5	30	69.6%	72.6%	70.4%	71.5%
$\mathcal{D}_{GS}$	5	50	71.8%	73.5%	71.9%	72.7%
$\mathcal{D}_{SR}$	5	50	65.1%	68.4%	65.9%	67.2%
$\mathcal{D}_{GS}$	5	70	77.0%	77.3%	77.0%	77.2%
$\mathcal{D}_{SR}$	5	70	69.6%	72.0%	70.6%	71.3%
$\mathcal{D}_{GS}$	16	30	75.5%	75.3%	75.6%	75.4%
$\mathcal{D}_{SR}$	16	30	68.7%	70.1%	69.3%	69.7%
$\mathcal{D}_{GS}$	16	50	72.2%	74.2%	72.3%	73.2%
$\mathcal{D}_{SR}$	16	50	65.2%	69.1%	66.3%	67.6%
$\mathcal{D}_{GS}$	16	70	81.0%	81.5%	81.0%	81.2%
$\mathcal{D}_{SR}$	16	70	69.8%	75.1%	69.8%	72.4%

with ConvGRU being the best model on both $\mathcal{D}_{SR}$ and $\mathcal{D}_{GS}$ datasets. This demonstrates that temporal information is preserved in the coarse representation of EEG signals transformed into atlases, indicating that architectures with recurrent elements can exhibit varying performance depending on internal design but should be the preferred choice, especially when applied to the proposed input representation.

State-of-the-art comparison

To conclude the evaluation of the Empátheia classifier, a state-of-the-art comparison was conducted using the SEED dataset. The results are presented in Table 6.15. As can be observed, the Empátheia system achieves performance comparable to many existing works in the literature. This finding is noteworthy, considering that the generated atlases represent a coarse transformation of the EEG signals utilized by other approaches. This suggests potential room for improvement in the presented PRISMIN encoder. Furthermore, existing solutions, even those with the highest performance, employ advanced yet complex models to handle the fine-grained details of EEG signals. However, this complexity comes at the cost of increased computational demands and the necessity for larger datasets. In contrast, as demonstrated in Table 6.5, the Empátheia classifiers achieve the reported performance with a dataset of lower quality that requires less disk space to store, enabling the use of lightweight models. For instance, Table 6.16 compares the best-performing Empátheia classifier, ConvGRU, with existing architectures from a computational

Table 6.12. ViT classifier ablation.

$\mathcal{D}$	$\mathcal{H}$	LR	Acc	Prec	Recall	F1
$\mathcal{D}_{GS}$	8	1e-4	56.4%	51.5%	53.8%	52.6%
$\mathcal{D}_{SR}$	8	1e-4	57.1%	52.3%	54.3%	53.3%
$\mathcal{D}_{GS}$	8	1e-5	55.7%	52.4%	52.1%	52.7%
$\mathcal{D}_{SR}$	8	1e-5	60.0%	56.7%	57.5%	57.1%
$\mathcal{D}_{GS}$	8	1e-6	56.1%	53.3%	53.5%	53.4%
$\mathcal{D}_{SR}$	8	1e-6	53.7%	52.4%	52.0%	52.2%
$\mathcal{D}_{GS}$	16	1e-4	58.2%	55.2%	55.4%	55.3%
$\mathcal{D}_{SR}$	16	1e-4	59.7%	57.7%	58.4%	58.1%
$\mathcal{D}_{GS}$	16	1e-5	56.2%	53.8%	54.4%	54.1%
$\mathcal{D}_{SR}$	16	1e-5	58.8%	55.7%	56.6%	56.1%
$\mathcal{D}_{GS}$	16	1e-6	54.7%	52.8%	53.2%	53.0%
$\mathcal{D}_{SR}$	16	1e-6	52.7%	50.5%	51.3%	50.9%
$\mathcal{D}_{GS}$	32	1e-4	57.3%	57.7%	56.9%	57.3%
$\mathcal{D}_{SR}$	32	1e-4	60.0%	59.0%	59.6%	59.4%
$\mathcal{D}_{GS}$	32	1e-5	55.0%	56.3%	54.6%	55.4%
$\mathcal{D}_{SR}$	32	1e-5	55.5%	56.0%	55.3%	55.6%
$\mathcal{D}_{GS}$	32	1e-6	60.0%	59.7%	59.6%	59.6%
$\mathcal{D}_{SR}$	32	1e-6	53.9%	53.0%	53.6%	53.3%

Table 6.13. Conv classifier ablation.

$\mathcal{D}$	K	LR	Acc	Prec	Recall	F1
$\mathcal{D}_{GS}$	3	1e-3	69.7%	70.5%	69.7%	70.1%
$\mathcal{D}_{SR}$	3	1e-3	66.6%	65.9%	66.6%	66.3%
$\mathcal{D}_{GS}$	3	2e-3	58.1%	57.0%	58.0%	57.5%
$\mathcal{D}_{SR}$	3	2e-3	71.1%	71.5%	71.1%	71.5%
$\mathcal{D}_{GS}$	3	5e-4	63.5%	67.7%	63.5%	65.6%
$\mathcal{D}_{SR}$	3	5e-4	71.6%	71.3%	71.6%	71.5%
$\mathcal{D}_{GS}$	5	1e-3	66.8%	66.9%	66.8%	66.9%
$\mathcal{D}_{SR}$	5	1e-3	70.6%	72.6%	70.7%	71.6%
$\mathcal{D}_{GS}$	5	2e-3	57.1%	71.8%	57.1%	63.6%
$\mathcal{D}_{SR}$	5	2e-3	62.2%	72.4%	62.8%	67.3%
$\mathcal{D}_{GS}$	5	5e-4	68.9%	69.6%	68.9%	69.3%
$\mathcal{D}_{SR}$	5	5e-4	71.1%	71.5%	71.1%	71.3%
$\mathcal{D}_{GS}$	16	1e-3	63.4%	64.7%	63.4%	64.1%
$\mathcal{D}_{SR}$	16	1e-3	71.7%	72.8%	71.7%	71.3%
$\mathcal{D}_{GS}$	16	2e-3	54.8%	60.4%	54.8%	57.4%
$\mathcal{D}_{SR}$	16	2e-3	64.6%	65.7%	64.6%	65.1%
$\mathcal{D}_{GS}$	16	5e-4	65.1%	65.9%	65.1%	65.5%
$\mathcal{D}_{SR}$	16	5e-4	67.1%	69.5%	67.1%	68.3%

standpoint. Not only is the proposed model noticeably smaller than existing solutions, but it also processes individual trials significantly faster, approximately 45 times faster, using similar hardware. This underscores the rationale behind the Empátheia system and highlights the advantage of its PRISMIN encoder, which reduces the dataset size and, consequently, improves computational efficiency in terms of model size and training speed, thereby demonstrating the effectiveness of the proposed approach.

6.3 EEG limbs activation

In this section, the dataset, experimental setup, and experimental protocol are presented first. Subsequently, the results obtained from the intra-subject and cross-subject experiments, along with a comparison to the current literature, are discussed.

6.3.1 Dataset

The PhysioNet EEG motor movement/imagery dataset has been used for the experimental phase also in this work, in fact it is the reference dataset for motor imagery experiments. In particular, only the subset of the data related to left and right limb activation has been used here. The description of the dataset has already been done in Section 6.2.1. No pre-processing steps were used on the dataset samples. This

Table 6.14. Best ablation configurations comparison.

Model	$\mathcal{D}$	Acc	Prec	Recall	F1	FLOPS*	Params
Conv	$\mathcal{D}_{GS}$	69.7%	70.5%	69.7%	70.1%	0.1G	28.6M
Conv	$\mathcal{D}_{SR}$	71.7%	72.8%	71.7%	71.3%		
ViT	$\mathcal{D}_{GS}$	75.5%	75.4%	75.5%	75.4%	40.16G	86M
ViT	$\mathcal{D}_{SR}$	72.6%	72.1%	72.2%	72.2%		
ConvLSTM	$\mathcal{D}_{GS}$	82.8%	83.0%	82.9%	82.9%	0.037G	0.13M
ConvLSTM	$\mathcal{D}_{SR}$	77.6%	77.8%	77.7%	77.8%		
ConvGRU	$\mathcal{D}_{GS}$	83.5%	83.9%	83.6%	83.7%	0.035G	0.14M
ConvGRU	$\mathcal{D}_{SR}$	80.3%	78.1%	76.6%	77.3%		

* Number of floating point operations per input.

Table 6.15. State-of-the-art performance comparison.

Model	Acc	Prec	Recall	F1
MFBPST-3D [109]	96.79%	-	-	-
DNN-SAE [110]	96.77%	-	-	-
TANN [111]	93.34%	-	-	-
SVM [49]	86.65%	-	-	-
DBN [49]	86.08%	-	-	-
ConvGRU (our)	83.50%	83.90%	83.60%	83.70%
DTCW-SRU [112]	83.13%	82.24%	81.53%	81.24%
ConvLSTM (our)	82.80%	83.00%	82.90%	82.90%
LRM. [49]	82.70%	-	-	-
KNN [49]	72.60%	-	-	-
Conv (our)	71.70%	72.80%	71.70%	71.30%
ViT (our)	60.00%	59.70%	59.60%	59.60%

choice is due to the robustness of the CNN models in extracting features also in contexts in which the data present some noise.

6.3.2 Implementation details

The final model architecture employed in the experiments was determined by considering the optimal configuration that emerged from the ablation study delineated in Section 6.2.4. Consequently, the model comprises four distinct streams in m_s, each containing two convolutional blocks. The kernel size for the first stream is set to $k = 7$ to analyze a sufficiently wide temporal window. In the experiments, the best results are obtained with one convolutional block, one final convolutional block, and one pooling block. The structural delineation of each of these blocks is explained in detail in Table 5.1. The classifier, mc, comprises two linear layers, as depicted in Table 5.2. The model has been implemented by using the Pytorch framework, while the EEG data has been handled with the MNE framework [218]. The model has

Table 6.16. State-of-the-art computational comparison.

Model	Params	Time*	GPU	FLOPS
MFBPST-3D [109]	9M	33s	RTX3090	-
DTCW-SRU [112]	2.45M	27s	RTX3090	-
ConvGRU (our)	0.3M	0.66s	Quadro RTX6000	0.6G

* training time required to analyze a single trial.

been trained and tested on a machine with a 2-core Intel Xeon CPU at 2.2 GHz, 13 GiB of RAM and an nVidia Tesla T4 GPU bundled with 16 GiB of GDDR6 RAM.

6.3.3 Experimental protocol

This work is focused on two types of experiments: cross-subject (Section 6.3.3) and intra-subject (Section 6.3.3) classification of MI-based EEGs.

Cross-subject

The following is the formal definition of the cross-subject classification task. Given a pool $E = \bigcup_{1 \leq i \leq N}(x_i, y_i)$ of N labelled EEGs belonging to all the subjects, a model is trained on a subset $E_t \subset E$ to identify patterns in EEG signals to predict the thought action, and its performance is evaluated on a subset $E_v \subset E$, where $E_t \cap E_v = \emptyset$ and $E_t \bigcup E_v = E$. Each sample $(x, y) \in E$ is formed by an EEG record $x \in \mathbb{R}^{s \times c_i n}$of a fixed duration of s samples and $c_i n$ channels, and a label $y \in L$ that indicates the action being thought by the subject in that trial.

Intra-subject

Intra-subject classification has been structured as follows. For each subject $S_i \in S$ in a pool S of subjects, where $S_i = \bigcup_{1 \leq i \leq N}(x_i, y_i)$ is a collection of N labelled EEGs of a single subject, a model is trained on a subset $E_t \subset S_i$ to identify patterns in EEG signals and predict the thought action. Its performance is evaluated on a subset $E_v \subset S_i$ where $E_t \cap E_v = \emptyset$ and $E_t \cup E_v = S_i$. As in section 6.3.3, each sample $(x, y) \in E$ is formed by an EEG record $x \in \mathbb{R}^{s \times c_{in}}$ and a label $y \in L$. At the end of the process, a model is produced for each subject, and the final intra-subject performances are calculated as the mean of the performances of each model.

6.3.4 Results

In this section, the results obtained in the performed intra-subject and cross-subject experiments, together with the comparison with the current literature, are presented.

Performance metrics

Given the intricate nature of EEG data, which is characterized by high dimensionality and susceptibility to noise, it is imperative to employ a variety of metrics to comprehensively assess the model performance on the validation sets. In the following paragraphs, the metrics employed for measuring the goodness of the proposed model

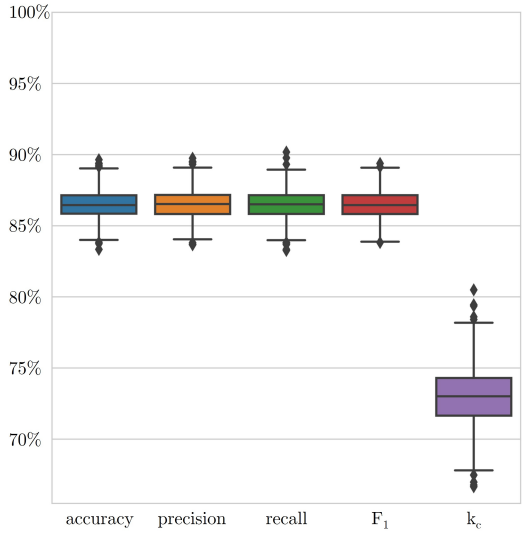


Figure 6.6. The four metrics computed for the cross-subject experiments. For each of the metrics, a box plot is provided, which allows a better overview of the metrics distribution and their variability in the 10 folds.

are described [219].
Accuracy: Represents the ratio between correctly classified samples and their total amount within the dataset. Hence, the accuracy gives an overall measure of model performance.
Precision: Represents the ratio between correctly classified positive samples and the number of positive predictions. Precision quantifies the number of true positive predictions among the total number of positive predictions made, reflecting the model's reliability in predicting positive instances.
Recall: Represents the ratio between correctly classified positive samples and the number of positive samples. The recall represents the proportion of actual positive instances that the model correctly identified, providing insight into the model's sensitivity to positive instances.
F1 Score: Represents the relation between precision and recall using their harmonic mean. F1 offers a balanced measure of the model's performance, especially when the class distribution is imbalanced. Accuracy, precision, recall, and F1 score are critical for assessing the model's ability to make correct predictions and minimize errors, providing insights into the model's reliability in real-world applications.

Cohen's K Represents the agreement between two annotators that classify the items, and also considers the agreement obtained by chance, ranging in [-1, 1] [220]:

$$k_C = \frac{accuracy - p_e}{1 - p_e} \tag{6.1}$$

where p_e represents the expected agreement, assuming both annotators assign labels at random. Cohen's k thus quantifies the alignment between the model's classifications and expert annotations, accounting for chance agreement, and thereby adds a critical validation layer, which is particularly important in medical contexts

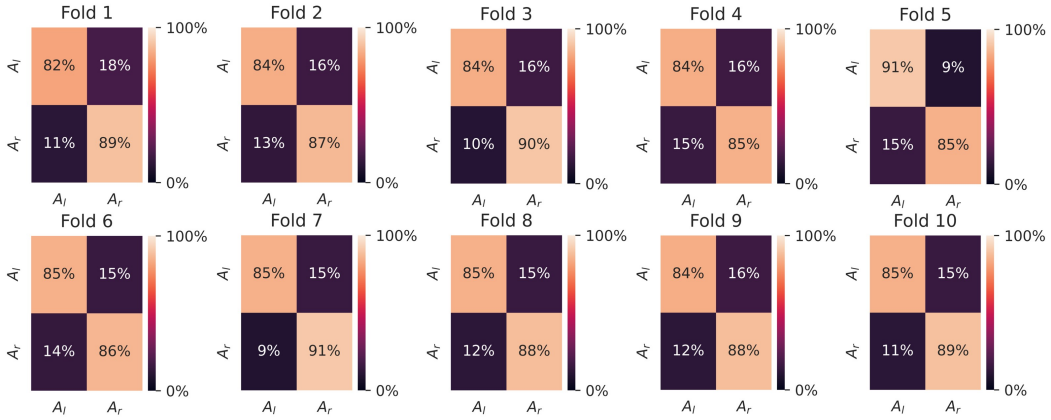


Figure 6.7. Confusion matrix for the cross-subject experiment on the PhysioNet dataset. A_l and A_r are, respectively, the labels for left and right arm activation.

where reliability is paramount. In cross-subject experiments, applying diverse metrics is essential, given the inherent variability and complexity stemming from differences in subject characteristics, behaviors, and responses. This comprehensive metric approach enables a detailed assessment of the model's generalization, reliability, and robustness across subjects. In contrast, intra-subject experiments generally present reduced variability, allowing for a streamlined evaluation with fewer metrics, which remains sufficient to capture the model's performance within individual subjects. The selected metric set offers a rigorous method for assessing model performance, uncovering insights and nuances that might otherwise be overlooked with a narrower evaluative scope, as often seen in prior studies.

6.3.5 Evaluation and State-of-the-art comparison

It is important to note that many studies [134] [135] [142] primarily showcase the accuracy metric due to its simplicity, which provides a quick overview of model performance. However, this approach can overlook important subtleties and imbalances in the evaluation process. In contrast, the introduced methodology, similar to the approaches in [129] and [137], employs multiple complementary metrics, as detailed in Section 6.3.4, to ensure a nuanced understanding of model performance and to provide a more comprehensive insight into its reliability and effectiveness. Furthermore, including both cross-subject and intra-subject evaluations distinguishes the proposed work from the norm, offering a level of depth found in only a few studies, such as [128], [129], [140], and [137]. This rigorous approach enables a thorough assessment of model performance and adaptability across various subjects as well as within individual subject contexts. By focusing solely on one type of evaluation, critical insights and nuances integral to a comprehensive understanding of model behavior and efficacy may be overlooked.

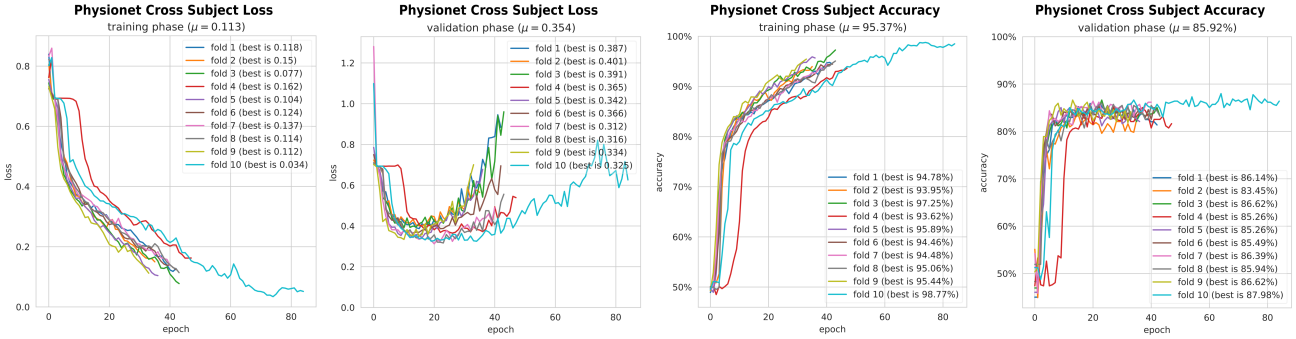
Cross-Subject results

In cross-subject experiments, the proposed model achieves state-of-the-art results compared to related work. As shown in Figure 6.8, the validation loss reaches

Table 6.17. Comparison with the related work on the PhysioNet dataset. Bold values are the best ones in their respective columns.

Work	Year	Subjects	Within-subject accuracy	Cross-subject accuracy
[134]	2013	109	68.2%	-
[135]	2014	100	75%	-
[136]	2016	24	80.1%	-
[141]	2022	50	92.4%	-
[138]	2014	103	-	71.6%
[132]	2014	56	-	72.4%
[139]	2020	105	-	74.7%
[142]	2019	105	-	80.5%
[128]	2016	85	63.6%	60.1%
[140]	2018	105	86.5%	80.4%
[137]	2020	103	98.3%	83.2%
[129]	2021	105	98.3%	83.6%
Ours	2023	105	**99.4%**	**85.9%**

its minimum around the twentieth epoch; thereafter, the model begins to overfit, attaining a mean accuracy score of $\mu = 85.9$. This accuracy is calculated by averaging the maximum accuracy of each fold, consistently surpassing the previous state-of-the-art score reported in [129] by approximately 2.3%. Notably, the latter study, along with the work in [128], employs a SVM classifier, necessitating a data pre-processing step to remove noise and artifacts. This pre-processing can significantly impact classification results, as there is a risk of discarding relevant information critical to the task. In contrast, the use of CNN-based approaches, as proposed in this thesis and in studies such as [140] and [142], allows for the direct input of raw EEG data into the CNN model. This is possible because CNNs can automatically extract relevant features from the input data while effectively capturing local patterns and filtering out irrelevant information, making the model robust for real-world applications. The normalized confusion matrix for the experiments, presented in Figure 6.7, indicates that left-arm samples are classified with higher accuracy. The boxplot in Figure 6.6 provides a detailed examination and comparison of the metrics discussed in Section 6.3.4 for the cross-subject experiments. It is observed that accuracy, precision, recall, and F1 scores predominantly fall within the 84% to 90% range, demonstrating the model's strong performance and effectiveness in correctly identifying and classifying instances across subjects. In contrast, Cohen's k values range from 68% to a maximum of 81%. The relative decrease in Cohen's k compared to the other metrics highlights potential inconsistencies in instance classification, hinting at the presence of false positives/negatives that may impact inter-annotator reliability. This variation among metrics underscores the complex dynamics of model performance and reinforces the importance of employing multiple metrics to ensure the reliability and validity of the model's predictive capabilities across diverse experimental conditions. It is also noteworthy that the feature extraction component of the pipeline is fully managed



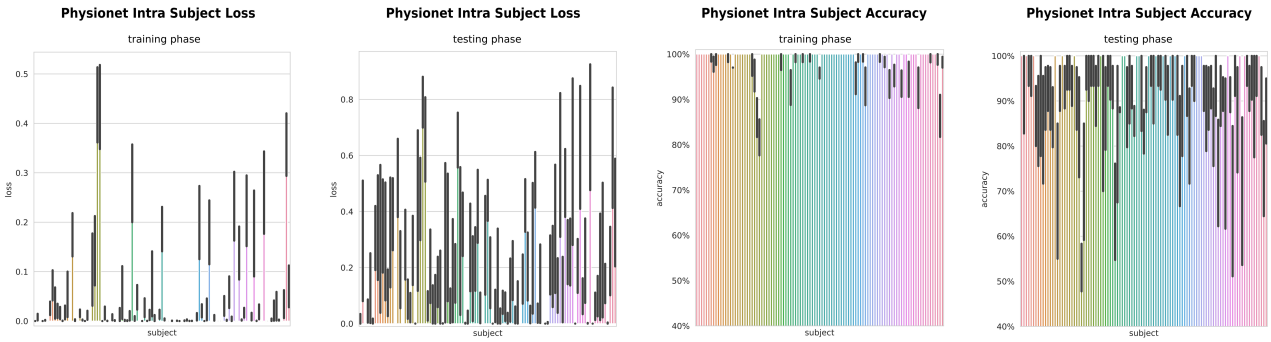
(a) **(b)** **(c)** **(d)**

Figure 6.8. Loss values for training (a) and testing (b) phases on the PhysioNet dataset using 10-fold cross validation, and accuracy values for training (c) and testing (d) phases on the PhysioNet dataset using 10-fold cross validation. Both accuracy and loss refer to the cross-subject experiment. Each line in the plots represents the i-th fold, while the value μ represents the average value obtained among the several folds. Formally, let $M \in \mathbb{R}^{k_f \times e}$ be the accuracy matrix, where k_f is the number of folds in the cross-validation and e the maximum number of epochs, $\mu = \frac{1}{k_f} \sum_{1 \leq i \leq k_f} \max(M_i)$ is equal to the mean of the maximum accuracy reached for each fold. In the experiments, $k_f = 10$ and $e = 100$.

by the streams through learned convolutions, which automatically filter and extract meaningful information from the data. This approach allows researchers to avoid the need for manually defining algorithms for noise reduction or channel selection, enabling a greater focus on the neural network's architecture.

Intra-Subject results

The results of the intra-subject experiments are highly encouraging, yielding a μ value of 0.996, where μ is calculated as the mean of the maximum accuracy scores per subject for each fold. This indicates that nearly all subjects were correctly classified, as illustrated in Figure 6.9. In contrast to Figure 6.8, which uses a line plot to depict cross-subject variations from a single training and validation process, Figure 6.9 utilizes a bar plot to convey intra-subject variability across multiple training sessions relative to the number of subjects. The bar plot effectively represents and contrasts discrete, categorical data points, facilitating clear comparisons between individual subjects. The choice of different plot types in Figures 6.8 and 6.9 is crucial for accurately conveying the distinct aspects of the neural network's performance and variability in both cross-subject and intra-subject contexts. Furthermore, the previous state-of-the-art accuracy of 0.983, as reported by [129] and [137], falls short by approximately 1.3% compared to the accuracy achieved by the proposed method. Despite both studies achieving equal intra-subject accuracy, it is noteworthy that [129] conducts experiments with 105 subjects, while [137] involves 103 subjects. Furthermore, among the works utilizing deep learning-based approaches, the highest results are consistently observed, with the exception of the study by [129], which employs an SVM. The strong performance of the latter can be attributed to the automatic selection of channels based on their relevance during MI tasks.



(a) **(b)** **(c)** **(d)**

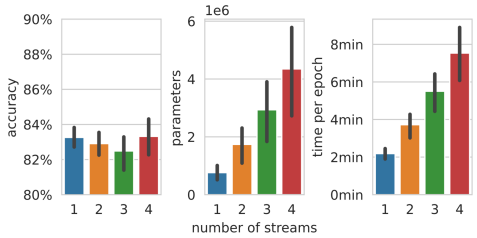
Figure 6.9. Bar plots of loss (a,b) and accuracy (c,d) per subject on the PhysioNet dataset using 10-fold cross-validation for intra-subject experiments. Each bar represents a subject of the dataset, and the grey part of the bar represents the variability among the several folds of the considered measure.

Ablation study

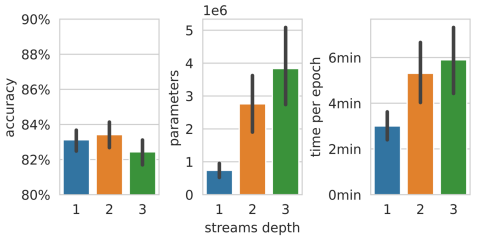
In this section, the choices made during the selection of the architecture of the final model are discussed. The several experiments have been made using a split At of the PhysioNet dataset D as the training set, with $|A_t| \approx 0.7 \cdot |D|$, and a test set A_v such that $A_t \cup A_v = D$ and $A_t \cap A_v = \emptyset$. This ablation study consisted of several experiments that focused on different parts of the network: the number of streams and their depth, the structure of the convolutional block, and the depth of the final classifier. Figure 6.10 provides a quantitative representation of the conducted experiments. Each experimental configuration is systematically illustrated with three distinct bar plots representing accuracy, the number of parameters, and the average time taken for each training epoch on the benchmark machine, as described in Section 6.3.2. The y-axis delineates the scale of the applied metrics, with percentage, quantity, and time (in minutes) clearly represented. The x-axis portrays the varying values of the parameter being examined. Since more than one run was executed for each unique configuration, the bar plots include representations of fluctuations among the runs. These variations are illustrated through a black line superimposed on each box, offering a comprehensive view of the inherent variability in each configuration. This design enables a nuanced understanding of the relative stability and reliability of the tested parameter values.

Number of streams

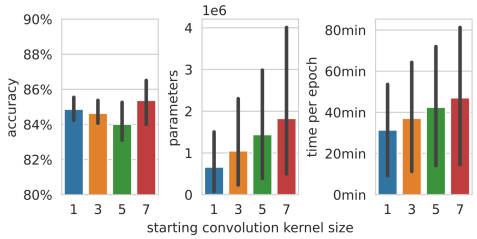
This experimental phase consisted of changing the number of streams of the model, and each stream follows the structure described in Section 5.1.1. According to the results shown in Figure 6.10a, the number of parameters and the time required for the training proportionally increases as the number of streams increases. On the contrary, the accuracy value does not follow this trend. In fact, while with 1, 2, and 3 streams, the maximum value is comparable, with 4 streams, the accuracy increases significantly. This behavior is due to the bigger kernel size in the latest streams, thus allowing the extraction of information from a bigger temporal interval.



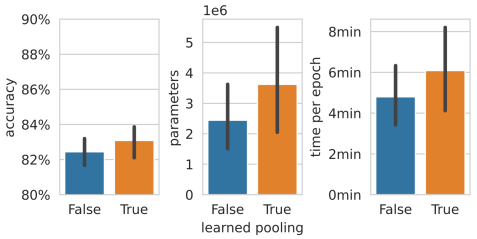
(a) Results about the number of streams in m_s.



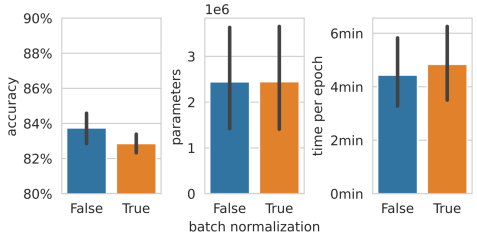
(b) Results about the number of convolution blocks in each stream.



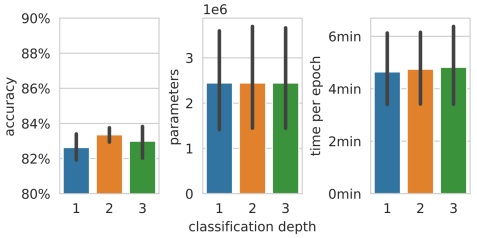
(c) Results about the starting kernel size (*i.e.* the value of k in m_{s_1}, which is progressively increased by 2 in the following streams).



(d) Results about the replacement of Max Pooling layers with 1D Convolutions at the end of the convolution blocks.



(e) Results about the presence of Batch Normalization between convolutions in convolutional blocks.



(f) Results about the number of linear layers in m_c.

Figure 6.10. Results of ablation studies on the structure of the proposed architecture. The features plotted are the accuracy on the PhysioNet dataset using 10-fold cross-validation, the number of parameters of the model, and the average minutes per training epoch.

Number of streams

In this experimental phase, several convolutional blocks per stream have been tested. The number of blocks ranged from one to three. The structure of convolutional streams and blocks follows the description provided in Section 5.1.1. According to the results shown in Figure 6.10a, the highest accuracy value is achieved by using two convolutional blocks. This is amenable to the fact that a single block is not able to extract fine-grained features, while using three blocks, the signal results to be too much processed, thus leading to losing part of the information from the original signal. It is worth noting that the number of parameters and average training time are not proportional, as seen in Figure 6.10a. This is due to the fact that when a single convolutional block is used, the final convolution of the block having stride 2 is removed, where instead b inner convolutional blocks, i.e., all the convolutional

blocks except the last one, have $b \cdot 3$ convolutions.

Kernel size

In this experimental phase, different kernel sizes for several streams have been tested. In detail, the kernel size of the first stream has been set, respectively, to a size of $[3, 5, 7]$. Then, the size of subsequent streams has been computed by following the process defined in Section 5.1.1. As depicted in Figure 6.10c, this parameter does not have a significant impact on the final accuracy.

Learned pooling

In this experimental phase, tests were conducted to determine which layer, between convolution and max pooling, must be placed at the end of the convolutional block to achieve the best results. As shown in Figure 6.10d, the learned convolution allows us to get better accuracy results at a slightly higher computational cost compared to the max pooling alternative, which is not a learnable layer. These findings are in line with the current state-of-the-art, as the trend in CNNs is to completely replace pooling operations [221].

Batch normalization

In this experimental phase, the effectiveness of batch normalization has been tested. In detail, the batch normalization layer is tested after the first two convolutional layers of each convolutional block. Batch normalization is a technique adopted to reduce the internal covariate shift of batches, thus speeding up the convergence and solving various initialization problems [203]. From the results shown in Figure 6.10e, normalizing batches reduces accuracy and slightly increases the training time. The decrease in accuracy is due to the smoothing effect of batch normalization on signals within the same batch, which results in the loss of several peaks in the signals and makes the samples too similar. Regarding the increasing training time, instead, it is simply due to the additional operation performed.

Depth of classification layers

In this last experimental phase, the depth of the linear layer has been tested. Such depth consisted of a number of layers ranging from 1 to 3. As shown in Figure 6.10f, the highest accuracy value was achieved using two classification blocks, although the average training time and the number of parameters remained similar across all configurations.

Chapter 7

Conclusions

7.1 EEG identification

This thesis presented a novel model for EEG-based person identification. The model consists of three components: a pre-processing component in charge of signal cleaning from artifacts and noise using established procedures. In this way, the presence of background noise such as muscle, EOG, and ECG artifacts is limited, which could lead to a degradation of the system's ability to recognize identity. A graph representation component based on Phase-Locking Value and EEG spectrogram to represent both temporal and spatial information from EEG signals, focusing on the most relevant features for the identification task. This innovative representation allows integrate in a single data structure of both time-frequency and brain functional connectivity information of the original EEG signal; And a classification component based on GNNs and Spline convolution to classify the extracted features. This component is able to extract and exploit the most distinctive features from the graph input, leading to better performance being able to generalize also over different human states. An extensive set of experiments was conducted to evaluate the generalization capability of the system under different conditions and against the best SOTA works already present in literature. The system has been trained, tested, and evaluated on different combinations of resting and activity state signals, including, for the first time in literature, a novel experiment which provides for the training of the model only on activity states and testing on resting state signals. Furthermore, to make the model evaluation more comprehensive and exhaustive, another experiment has been proposed to explore the impact of different EEG frames length on the system performance. The experiments suggest that the use of longer EEG frames leads to a smaller training dataset, slightly decreasing the identification accuracy. Besides, the application of GNN requires sufficient data to ensure its performance. The outcomes demonstrated that the proposed method is effective and robust against different human states and on signals acquired with both on dry and wet headsets, on two public benchmarks reaching State-of-the-art results. As future work, since the biometric identification system proposed in this thesis has been evaluated on all publicly available EEG datasets composed solely of single-session data, a more complex dataset will be acquired. This dataset will incorporate multi-session data, where each subject's EEG sessions are recorded at varying time intervals from the

initial session, and will include a larger pool of distinct subjects. Such an approach could enhance the robustness of the system in more challenging scenarios, extending beyond single-session acquisitions, while providing a larger volume of data to further validate and improve the model's generalizability across diverse conditions and sessions. Additionally, although the system has been trained and tested to solve the proposed task reaching acceptable time also for real-time application, e.g. in sensitive environments requiring high-security authentication, its performances are expected to improve significantly with the adoption of high-end hardware. In this thesis, commonly available mid-range hardware was used, which suggests that leveraging on more advanced computing resources could further optimize efficiency, accuracy, and general performance. Moreover, the applicability of this system to real-world use cases is closely tied to technological advancements and the acceptability of EEG devices. At present, EEG acquisition systems tend to be invasive, bulky, and susceptible to noise, which limits their widespread adoption. In the future, miniaturization of EEG devices and improvements in their design could make them more user-friendly, less prone to noise, and better suited for practical, everyday applications.

7.2 EEG emotion recognition

This thesis presented the Empátheia system, which performs emotion classification from EEG signals using a reduced amount of data. The proposed approach consists of two main components: the PRISMIN encoder and the Empátheia classifier. The PRISMIN encoder generates coarse atlases—2D images encoded using grey-scale or short-rainbow mapping—to represent the captured emotion within EEG signals. This representation significantly reduces the input dataset, enabling the implementation of lightweight models and faster training times. The second component, the Empátheia classifier, utilizes deep neural networks tailored to capture spatio-temporal characteristics present in the atlases to perform emotion classification. Multiple and different tests on public reference datasets, i.e., SEED, have been performed to define the most accurate network that could classify emotions. The PRISMIN encoder reduced the dataset size by up to 10.3 times. The Empátheia classifiers achieved competitive performance in line with several existing literature works. However, due to the coarse nature of the atlas representation, they fell short of the best-performing approaches. Nevertheless, these results suggest ample room for improvement in both the PRISMIN encoder and Empátheia classifier. Furthermore, ablation studies on various Empátheia classifier hyper-parameters revealed stable performances across all experiments. The mixed models exhibited higher metrics due to their internal configurations, demonstrating the atlas ability to accurately represent emotion from EEG signals. The performed experiments showed that, among the tested models, the ConvGRU achieved the best results. As a potential avenue for future research, exploring additional encoding strategies is warranted, given their varying performance depending on the underlying architecture. This indicates that new strategies for the PRISMIN encoder could yield improved results. Additionally, implementing more advanced models as Empátheia classifiers could potentially compensate for the coarse representation derived from the atlases, for

example, integrating the capability of recurrent architectures to capture temporal information with the attention mechanism [222, 223]. Moreover, the adoption of different datasets will be explored in order to have a deeper analysis of the system performance under different conditions to verify also the scalability for larger amounts of data. Also, some recent techniques, e.g., Neural Dynamic Classification (NDC) algorithms [224], Dynamic Ensemble Learning (DEL) approaches [225], Finite Element Machine (FEM) [226], Domain Adaptation (DA) strategies [227], FC [228], and self-supervised learning [229] could provide a guideline about possible strategies to apply to compressed signals.

7.3 EEG limbs activation

Finally, this thesis proposed a novel multi-stream 1D CNN model for classifying EEG signals related to the MI application field. The main novelty of the model is the use of the multi-stream strategy, where parallel kernels with different sizes on each stream are able to retrieve meaningful information from the same data at different time scales. Comparative experiments on the reference dataset, i.e., PhysioNet EEG motor movement/imagery data collection, for this type of application context, i.e., MI, show that the proposed method outperforms the current state-of-the-art in cross-subject and intra-subject experiments. Furthermore, the introduced approach demonstrates a marked advancement over related works, primarily attributed to the utilization of a more extensive and diverse array of metrics. This comprehensive evaluation methodology enables a more nuanced and thorough understanding of model performance, thus ensuring the validity and reliability of derived findings across varied contexts and applications. Finally, a consistent ablation study has been conducted to assess the performances of multiple feature extraction streams and their structure, which led to the best model configurations surpassing the previous results in the literature. The biggest limitation of the proposed method is the optimization, by having different streams, it is more difficult to optimize the proposed architecture for different reasons. Firstly, the higher number of hyper-parameters that must be correctly tuned. In addition, the definition of the architecture itself can be non-trivial, e.g., number of streams, depth of the streams, streams of the same length or not, and more. In future research, to further evaluate the performance of the proposed method, the system will be tested with additional state-of-the-art datasets, and a more complex dataset will be collected, integrating multi-session data, where each subject's EEG session is acquired at different time intervals from the first, and a greater number of distinct subjects. While the current results demonstrate the effectiveness of the approach, expanding the evaluation to include other datasets will provide a more comprehensive understanding of its capabilities and robustness, assessing the generalizability of the proposed method across different conditions. Additionally, will be explored potential modifications and enhancements of the method, e.g., by considering attention mechanisms, to verify if it is possible to reduce the number of streams or the number of convolutional layers.

7.4 Ethical aspects

Wearable devices capable of measuring and recording physiological signals are becoming increasingly accessible to the general public, offering improved affordability and enhanced signal quality. However, the data generated by these devices raise significant ethical concerns that remain largely overlooked. Users often lack a clear understanding of how this data can be utilized to uncover sensitive information about them, and developers may not fully anticipate how the physiological data collected today could be repurposed in the future for entirely different applications. The potential widespread adoption of EEG wearables introduces a critical set of challenges related to research ethics and legal compliance. As new technologies emerge to interface directly with the brain, they bring with them heightened risks of privacy violations. According to the General Data Protection Regulation (GDPR), which took effect in 2018, brain data is classified as sensitive data, requiring stricter protections compared to standard personal data. This classification imposes more rigorous standards for processing and securing such data, reflecting the significant impact that misuse could have on individuals' lives. EEG devices have the capacity to reveal a wide range of information about individuals, including health conditions, mental disorders, psychological traits, and cognitive states, potentially leading to issues such as discrimination based on neural data [230]. Some researchers argue that due to the sensitive nature of EEG data, individuals may not fully realize the breadth of information being extracted from their brains [230]. Consequently, scholars like Farisco et al. [231] emphasize that informed consent in biomedical applications must adhere to three key principles, namely: the disclosure of all necessary information, the subject's ability to comprehend this information, and the voluntariness of their participation [231]. In the context of commercial EEG applications, companies and manufacturers are required to secure valid consumer consent by providing specific, clear, and unambiguous information regarding the processing of sensitive data. Under GDPR, this includes disclosing the identity of the data controllers, the purpose of data processing, and the activities involved. Considering these factors, privacy and data protection stand out as essential elements for ensuring a smooth and ethical transition to a society increasingly shaped by BCI technologies.

Publications and collaborations

Published articles

Avola, D., Cinque, L., Di Mambro, A., Diko A., Fagioli, A., Foresti, G. L., Marini M.R., Mecca A., Pannone D. **Low-altitude aerial video surveillance via one-class SVM anomaly detection from textural features in UAV images.** In: Information. vol. 13, no.1, pp. 1. MDPI, 2021.
DOI: 10.3390/info13010002

Avola, D., Cinque, L., De Marsico, M., Di Mambro, A., Fagioli, A., Foresti, G. L., Lanzino R., Scarcello, F. **LieToMe: An LSTM-Based Method for Deception Detection by Hand Movements.** In: International Conference on Image Analysis and Processing (ICIAP). Lecture Notes in Computer Science, vol. 14233, pp. 387-398. Springer, 2023.
DOI: 10.1007/978-3-031-43148-7_33

Avola, D., Cinque, L., Di Mambro, A., D., Fagioli, A., Marini, M. R., Pannone, D., Fanini B., Foresti, G. L. **Spatio-Temporal Image-Based Encoded Atlases for EEG Emotion Recognition.** In: International Journal of Neural Systems, vol. 34, no 5, pp. 2450024, 2024.
DOI: 10.1142/S0129065724500242

Avola D., Cinque L., Di Mambro A., Lanzino R., Pannone D., Scarcello F. **Multi-Stream 1D CNN for EEG Motor Imagery Classification of Limbs Activation.** in: IEEE Access, vol. 12, pp. 83940-83951, 2024.
DOI: 10.1109/ACCESS.2024.3412710

Under review articles

Avola D., Cannistraci I., Cascio M., Cinque L., Diko A., Di Mambro A., Fagioli A., Foresti G. L., Mecca A., Pannone D., **The Evolution of UAV-based Remarkable Points Understanding: a Survey**. Submitted to: ACM Computing Surveys.

Di Mambro A., Avola D., Pannone D., **A Graph-based EEG Person Identification System Integrating Spatio-Temporal and Functional Connectivity Features**. Submitted to: IEEE Transaction on Information Forensics and Security.

Avola D., Di Mambro A., Doku M., Pannone D., **Advances in EEG Denoising and Artifact Removal via Transformer Architecture**. Submitted to: IEEE Access.

Collaborations in research projects

SEARCHER (Smart unmannEd AeRial vehiCles for Human likE monitoRing) project in collaboration with the Ministero della Difesa. The main purpose of this project is to equip a modern Unmanned Aerial Vehicle (UAV) with basic principles of autonomous decision, discernment and reasoning. My role was to be the task leader of Working Package 1 (WP1) and Working Package 2 (WP2). Regarding WP1, i was involved in the SOTA reporting on the logical and physical architecture of UAVs. Regarding WP2, i was involved in the SOTA reporting about deep learning algorithms related to autonomous decision, discernment and reasoning.

DRAGONS (Drone aerial images segmentation system) project in collaboration with the Ministero della Difesa. The main purpose of this project is to define an algorithm for UAV semantic segmentation over urban areas, also collecting a novel dataset for testing. My role was to be the task leader of WP1, in particular, to report on existing datasets for the urban segmentation task and related SOTA.

Bibliography

[1] B. Kerous, F. Skola, and F. Liarokapis, "Eeg-based bci and video games: a progress report," *Virtual Reality*, vol. 22, pp. 119–135, 2018.

[2] H. Altaheri, G. Muhammad, M. Alsulaiman, S. U. Amin, G. A. Altuwaijri, W. Abdul, M. A. Bencherif, and M. Faisal, "Deep learning techniques for classification of electroencephalogram (eeg) motor imagery (mi) signals: A review," *Neural Computing and Applications*, vol. 35, no. 20, pp. 14 681–14 722, 2023.

[3] S. M. Qaisar and S. F. Hussain, "Effective epileptic seizure detection by using level-crossing eeg sampling sub-bands statistical features selection and machine learning for mobile healthcare," *Computer Methods and Programs in Biomedicine*, vol. 203, p. 106034, 2021.

[4] Y. Liu, Y.-X. Huang, X. Zhang, W. Qi, J. Guo, Y. Hu, L. Zhang, and H. Su, "Deep c-lstm neural network for epileptic seizure and tumor detection using high-dimension eeg signals," *IEEE Access*, vol. 8, pp. 37 495–37 504, 2020.

[5] H. Cai, Z. Qu, Z. Li, Y. Zhang, X. Hu, and B. Hu, "Feature-level fusion approaches based on multimodal eeg data for depression recognition," *Information Fusion*, vol. 59, pp. 127–138, 2020.

[6] G. Zhang and A. Etemad, "Capsule attention for multimodal eeg-eog representation learning with application to driver vigilance estimation," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 1138–1149, 2021.

[7] A. Sors, S. Bonnet, S. Mirek, L. Vercueil, and J.-F. Payen, "A convolutional neural network for sleep stage scoring from raw single-channel eeg," *Biomedical Signal Processing and Control*, vol. 42, pp. 107–114, 2018.

[8] Y. Wang, Y. Huang, B. Gu, S. Cao, and D. Fang, "Identifying mental fatigue of construction workers using eeg and deep learning," *Automation in Construction*, vol. 151, p. 104887, 2023.

[9] X. Li, Y. Zhang, P. Tiwari, D. Song, B. Hu, M. Yang, Z. Zhao, N. Kumar, and P. Marttinen, "EEG based emotion recognition: A tutorial and review," *ACM Computing Surveys*, vol. 55, no. 4, pp. 1–57, 2022.

[10] K. Kamble and J. Sengupta, "A comprehensive survey on emotion recognition based on electroencephalograph (eeg) signals," *Multimedia Tools and Applications*, vol. 82, no. 18, pp. 27 269–27 304, 2023.

[11] A. Jain, R. Bolle, and S. Pankanti, *Biometrics: personal identification in networked society*. Springer Science & Business Media, 1999, vol. 479.

[12] K. N. Win, K. Li, J. Chen, P. F. Viger, and K. Li, "Fingerprint classification and identification algorithms for criminal investigation: A survey," *Future Generation Computer Systems*, vol. 110, pp. 758–771, 2020.

[13] C. Raghavendra, A. Kumaravel, and S. Sivasubramanian, "Iris technology: A review on iris based biometric systems for unique human identification," in *2017 International conference on algorithms, methodology, models and applications in emerging technologies (ICAMMAET)*. IEEE, 2017, pp. 1–6.

[14] W. Wu, S. J. Elliott, S. Lin, S. Sun, and Y. Tang, "Review of palm vein recognition," *IET Biometrics*, vol. 9, no. 1, pp. 1–10, 2020.

[15] M. Taskiran, N. Kahraman, and C. E. Erdem, "Face recognition: Past, present and future (a review)," *Digital Signal Processing*, vol. 106, p. 102809, 2020.

[16] H. Yadav and S. Maini, "Electroencephalogram based brain-computer interface: Applications, challenges, and opportunities," *Multimedia Tools and Applications*, vol. 82, no. 30, pp. 47 003–47 047, 2023.

[17] N. Tiwari, D. R. Edla, S. Dodia, and A. Bablani, "Brain computer interface: A comprehensive survey," *Biologically Inspired Cognitive Architectures*, vol. 26, pp. 118–129, 2018.

[18] A. Khosla, P. Khandnor, and T. Chand, "A comparative analysis of signal processing and classification methods for different applications based on eeg signals," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 2, pp. 649–690, 2020.

[19] N. S. Karuppusamy and B.-Y. Kang, "Multimodal system to detect driver fatigue using eeg, gyroscope, and image processing," *IEEE Access*, vol. 8, pp. 129 645–129 667, 2020.

[20] Y. Tan, Z. Sun, F. Duan, J. Solé-Casals, and C. F. Caiafa, "A multimodal emotion recognition method based on facial expressions and electroencephalography," *Biomedical Signal Processing and Control*, vol. 70, p. 103029, 2021.

[21] H. Cai, Z. Qu, Z. Li, Y. Zhang, X. Hu, and B. Hu, "Feature-level fusion approaches based on multimodal eeg data for depression recognition," *Information Fusion*, vol. 59, pp. 127–138, 2020.

[22] M. Wang, H. A. Abbass, and J. Hu, "Continuous authentication using EEG and face images for trusted autonomous systems," in *2016 14th Annual Conference on Privacy, Security and Trust (PST)*, 2016, pp. 368–375.

[23] S. J. Luck, *An introduction to the event-related potential technique*. MIT press, 2014.

[24] S. Sokol, "Visually evoked potentials: theory, techniques and clinical applications," *Survey of ophthalmology*, vol. 21, no. 1, pp. 18–44, 1976.

[25] G. Plourde, "Auditory evoked potentials," *Best Practice & Research Clinical Anaesthesiology*, vol. 20, no. 1, pp. 129–139, 2006.

[26] S. Morales and M. E. Bowers, "Time-frequency analysis methods and their application in developmental EEG data," *Developmental cognitive neuroscience*, vol. 54, p. 101067, 2022.

[27] E. Van Diessen, T. Numan, E. Van Dellen, A. Van Der Kooi, M. Boersma, D. Hofman, R. Van Lutterveld, B. Van Dijk, E. Van Straaten, A. Hillebrand *et al.*, "Opportunities and methodological challenges in eeg and meg resting state functional brain network research," *Clinical Neurophysiology*, vol. 126, no. 8, pp. 1468–1481, 2015.

[28] N. Palomero-Gallagher and K. Amunts, "A short review on emotion processing: A lateralized network of neuronal networks," *Brain Structure and Function*, vol. 227, no. 2, pp. 673–684, 2022.

[29] L. F. Barrett, R. Adolphs, S. Marsella, A. M. Martinez, and S. D. Pollak, "Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements," *Psychological Science in the Public Interest*, vol. 20, no. 1, pp. 1–68, 2019.

[30] S. Passardi, P. Peyk, M. Rufer, T. S. Wingenbach, and M. C. Pfaltz, "Facial mimicry, facial emotion recognition and alexithymia in post-traumatic stress disorder," *Behaviour Research and Therapy*, vol. 122, pp. 1–8, 2019.

[31] A. Semwal and N. D. Londhe, "Computer aided pain detection and intensity estimation using compact cnn based fusion network," *Applied Soft Computing*, vol. 112, pp. 1–11, 2021.

[32] R. Yuvaraj, M. Murugappan, U. R. Acharya, H. Adeli, N. M. Ibrahim, and E. Mesquita, "Brain functional connectivity patterns for emotional state classification in parkinson's disease patients without dementia," *Behavioural Brain Research*, vol. 298, no. Pt B, pp. 248–260, 2015.

[33] D. Avola, L. Cinque, M. De Marsico, A. Fagioli, and G. L. Foresti, "Lietome: Preliminary study on hand gestures for deception detection via fisher-lstm," *Pattern Recognition Letters*, vol. 138, pp. 455–461, 2020.

[34] D. Avola, M. Cascio, L. Cinque, A. Fagioli, and G. L. Foresti, "Lietome: An ensemble approach for deception detection from facial cues," *International Journal of Neural Systems*, vol. 31, no. 02, p. 2050068, 2021.

[35] M. A. Vicente-Querol, A. Fernández-Caballero, J. P. Molina, L. M. González-Gualda, P. Fernández-Sotos, and A. S. García, "Facial affect recognition in

immersive virtual reality: Where is the participant looking?" *International Journal of Neural Systems*, vol. 32, no. 10, p. 2250029, 2022.

[36] D. Avola, L. Cinque, A. Fagioli, G. L. Foresti, and C. Massaroni, "Deep temporal analysis for non-acted body affect recognition," *IEEE Transactions on Affective Computing*, vol. 13, no. 3, pp. 1366–1377, 2020.

[37] D. Avola, M. Cascio, L. Cinque, A. Fagioli, and G. L. Foresti, "Affective action and interaction recognition by multi-view representation learning from handcrafted low-level skeleton features," *International Journal of Neural Systems*, vol. 32, no. 10, p. 2250040, 2022.

[38] J. De Lope and M. Graña, "A hybrid Time-Distributed deep neural architecture for speech emotion recognition," *International Journal of Neural Systems*, vol. 32, no. 6, p. 2250024, 2022.

[39] A. Nandi, F. Xhafa, L. Subirats, and S. Fort, "Reward-Penalty weighted ensemble for emotion state classification from Multi-Modal data streams," *International Journal of Neural Systems*, vol. 32, no. 12, p. 2250049, 2022.

[40] A. Burns and H. Adeli, "Wearable technology for patients with brain and spinal cord injuries," *Reviews in the Neurosciences*, vol. 28, no. 8, pp. 913–920, 2017.

[41] A. Ortiz-Rosario and H. Adeli, "Brain-computer interface technologies: from signal to action," *Reviews in the Neurosciences*, vol. 24, no. 5, pp. 537–552, 2013.

[42] A. Olamat, P. Ozel, and S. Atasever, "Deep learning methods for multi-channel eeg-based emotion recognition," *International Journal of Neural Systems*, vol. 32, no. 05, p. 2250021, 2022.

[43] D. W. Prabowo, H. A. Nugroho, N. A. Setiawan, and J. Debayle, "A systematic literature review of emotion recognition using EEG signals," *Cognitive Systems Research*, vol. 82, p. 101152, 2023.

[44] T. Harmony, T. Fernández, J. Silva, J. Bernal, L. Díaz-Comas, A. Reyes, E. Marosi, M. Rodríguez, and M. Rodríguez, "EEG delta activity: an indicator of attention to internal processing during performance of mental tasks," *International Journal of Psychophysiology*, vol. 24, no. 1-2, pp. 161–171, 1996.

[45] W. Klimesch, "EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis," *Brain Research Reviews*, vol. 29, no. 2-3, pp. 169–195, 1999.

[46] ——, "Eeg-alpha rhythms and memory processes," *International Journal of Psychophysiology*, vol. 26, no. 1-3, pp. 319–340, 1997.

[47] A. K. Engel and P. Fries, "Beta-band oscillations—signalling the status quo?" *Current Opinion in Neurobiology*, vol. 20, no. 2, pp. 156–165, 2010.

[48] W. H. Miltner, C. Braun, M. Arnold, H. Witte, and E. Taub, "Coherence of gamma-band EEG activity as a basis for associative learning," *Nature*, vol. 397, no. 6718, pp. 434–436, 1999.

[49] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for eeg-based emotion recognition with deep neural networks," *IEEE Transactions on Autonomous Mental Development*, vol. 7, no. 3, pp. 162–175, 2015.

[50] R.-N. Duan, J.-Y. Zhu, and B.-L. Lu, "Differential entropy feature for eeg-based emotion classification," in *Proceedings of the International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE, 2013, pp. 81–84.

[51] G. Pfurtscheller and C. Neuper, "Motor imagery and direct brain–computer communication," *Proceedings of the IEEE*, vol. 89, no. 7, pp. 1123–1134, 2001.

[52] ——, "Motor imagery activates primary sensorimotor area in humans," *Neuroscience Letters*, vol. 239, no. 2-3, pp. 65–68, 1997.

[53] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain–computer interfaces for communication and control," *Clinical neurophysiology*, vol. 113, no. 6, pp. 767–791, 2002.

[54] R. Ortner, B. Z. Allison, G. Korisek, H. Gaggl, and G. Pfurtscheller, "An ssvep bci to control a hand orthosis for persons with tetraplegia," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 19, no. 1, pp. 1–5, 2011.

[55] T. M. Vaughan, W. J. Heetderks, L. J. Trejo, W. Z. Rymer, M. Weinrich, M. M. Moore, A. Kübler, B. H. Dobkin, N. Birbaumer, E. Donchin *et al.*, "Brain–computer interface technology: a review of the second international meeting," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 11, no. 2, pp. 94–109, 2003.

[56] A. Kübler, B. Kotchoubey, J. Kaiser, J. R. Wolpaw, and N. Birbaumer, "Brain–computer communication: unlocking the locked in," *Psychological Bulletin*, vol. 127, no. 3, pp. 358–375, 2001.

[57] M. Jeannerod, "Mental imagery in the motor context," *Neuropsychologia*, vol. 33, no. 11, pp. 1419–1432, 1995.

[58] J. Decety, "The neurophysiological basis of motor imagery," *Behavioural brain research*, vol. 77, no. 1-2, pp. 45–52, 1996.

[59] M. Lotze, P. Montoya, M. Erb, V. Hömberg, H. Flor, and N. Birbaumer, "Activation of cortical and cerebellar motor areas during executed and imagined hand movements: an fmri study," *Journal of Cognitive Neuroscience*, vol. 11, no. 5, pp. 491–501, 1999.

[60] B. H. Dobkin, "Brain–computer interface technology as a tool to augment neurorehabilitation," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 78, no. 6, pp. 698–699, 2007.

[61] S. Silvoni, A. Ramos-Murguialday, M. Cavinato, C. Volpato, G. Cisotto, A. Turolla, F. Piccione, and N. Birbaumer, "Brain-computer interface in stroke: a review of progress," *Clinical EEG and neuroscience*, vol. 42, no. 4, pp. 245–252, 2011.

[62] A. H. Do, P. T. Wang, C. E. King, S. N. Chun, and Z. Nenadic, "Brain-computer interface controlled robotic gait orthosis," *Journal of neuroengineering and rehabilitation*, vol. 10, pp. 1–9, 2013.

[63] X. Ma, S. Qiu, and H. He, "Multi-channel eeg recording during motor imagery of different joints from the same limb," *Scientific data*, vol. 7, no. 1, p. 191, 2020.

[64] J. Kalcher, D. Flotzinger, C. Neuper, S. Gölly, and G. Pfurtscheller, "Graz brain-computer interface ii: towards communication between humans and computers based on online classification of three different eeg patterns," *Medical and biological engineering and computing*, vol. 34, pp. 382–388, 1996.

[65] G. Pfurtscheller, C. Guger, G. Müller, G. Krausz, and C. Neuper, "Brain oscillations control hand orthosis in a tetraplegic," *Neuroscience letters*, vol. 292, no. 3, pp. 211–214, 2000.

[66] B. Fanini and L. Cinque, "Encoding, exchange and manipulation of captured immersive VR sessions for learning environments: The prismin framework," *Applied Sciences*, vol. 10, no. 6, pp. 1–16, 2020.

[67] C. A. Fidas and D. Lyras, "A review of EEG-based user authentication: Trends and future research directions," *IEEE Access*, 2023.

[68] M. V. Ruiz-Blondet, Z. Jin, and S. Laszlo, "Cerebre: A novel method for very high accuracy event-related potential biometric identification," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 7, pp. 1618–1629, 2016.

[69] H. Zhao, Y. Chen, W. Pei, H. Chen, and Y. Wang, "Towards online applications of EEG biometrics using visual evoked potentials," *Expert Systems with Applications*, vol. 177, p. 114961, 2021.

[70] R. Das, E. Maiorana, and P. Campisi, "EEG biometrics using visual stimuli: A longitudinal study," *IEEE Signal Processing Letters*, vol. 23, no. 3, pp. 341–345, 2016.

[71] Q. Gui, Z. Jin, and W. Xu, "Exploring EEG-based biometrics for user identification and authentication," in *2014 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*. IEEE, 2014, pp. 1–6.

[72] M. Phothisonothai, "An investigation of using SSVEP for EEG-based user authentication system," in *2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*. IEEE, 2015, pp. 923–926.

[73] E. Piciucco, E. Maiorana, O. Falzon, K. P. Camilleri, and P. Campisi, "Steady-state visual evoked potentials for EEG-based biometric identification," in *2017 International Conference of the Biometrics Special Interest Group (BIOSIG)*. IEEE, 2017, pp. 1–5.

[74] Z. Gao, W. Dang, M. Liu, W. Guo, K. Ma, and G. Chen, "Classification of EEG signals on VEP-based BCI systems with broad learning," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 11, pp. 7143–7151, 2020.

[75] B.-K. Min, H.-I. Suk, M.-H. Ahn, M.-H. Lee, and S.-W. Lee, "Individual identification using cognitive electroencephalographic neurodynamics," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 9, pp. 2159–2167, 2017.

[76] Y. Peng, C. M. Wong, Z. Wang, A. C. Rosa, H. T. Wang, and F. Wan, "Fatigue detection in SSVEP-BCIs based on wavelet entropy of EEG," *IEEE Access*, vol. 9, pp. 114 905–114 913, 2021.

[77] S. N. A. Seha and D. Hatzinakos, "EEG-based human recognition using steady-state AEPs and subject-unique spatial filters," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 3901–3910, 2020.

[78] ——, "A new approach for EEG-based biometric authentication using auditory stimulation," in *2019 International Conference on Biometrics (ICB)*. IEEE, 2019, pp. 1–6.

[79] R. Zhang, B. Yan, L. Tong, J. Shu, X. Song, and Y. Zeng, "Identity authentication using portable electroencephalography signals in resting states," *IEEE Access*, vol. 7, pp. 160 671–160 682, 2019.

[80] L. A. Moctezuma and M. Molinas, "Towards a minimal EEG channel array for a biometric system using resting-state and a genetic algorithm for channel selection," *Scientific RepoRtS*, vol. 10, no. 1, p. 14917, 2020.

[81] D. Kim and K. Kim, "Resting state EEG-based biometric system using concatenation of quadrantal functional networks," *Ieee Access*, vol. 7, pp. 65 745–65 756, 2019.

[82] D. La Rocca, P. Campisi, B. Vegso, P. Cserti, G. Kozmann, F. Babiloni, and F. D. V. Fallani, "Human brain distinctiveness based on EEG spectral coherence connectivity," *IEEE transactions on Biomedical Engineering*, vol. 61, no. 9, pp. 2406–2412, 2014.

[83] B. B. Das, P. Kumar, D. Kar, S. K. Ram, K. S. Babu, and R. K. Mohapatra, "A spatio-temporal model for EEG-based person identification," *Multimedia Tools and Applications*, vol. 78, no. 19, pp. 28 157–28 177, 2019.

[84] R. Das, E. Maiorana, and P. Campisi, "Motor imagery for EEG biometrics using convolutional neural network," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 2062–2066.

[85] J. Chuang, H. Nguyen, C. Wang, and B. Johnson, "I think, therefore i am: Usability and security of authentication using brainwaves," in *Financial Cryptography and Data Security: FC 2013 Workshops, USEC and WAHC 2013, Okinawa, Japan, April 1, 2013, Revised Selected Papers 17*. Springer, 2013, pp. 1–16.

[86] M. Wang, J. Hu, and H. A. Abbass, "Brainprint: EEG biometric identification based on analyzing brain connectivity graphs," *Pattern Recognition*, vol. 105, p. 107381, 2020.

[87] Y. Du, Y. Xu, X. Wang, L. Liu, and P. Ma, "EEG temporal–spatial transformer for person identification," *Scientific Reports*, vol. 12, no. 1, p. 14378, 2022.

[88] E. Debie, N. Moustafa, and A. Vasilakos, "Session invariant EEG signatures using elicitation protocol fusion and convolutional neural network," *IEEE Transactions on Dependable and Secure Computing*, vol. 19, no. 4, pp. 2488–2500, 2021.

[89] P. Kumari and A. Vaish, "Brainwave based user identification system: A pilot study in robotics environment," *Robotics and Autonomous Systems*, vol. 65, pp. 15–23, 2015.

[90] E. Maiorana, "Deep learning for EEG-based biometric recognition," *Neurocomputing*, vol. 410, pp. 374–386, 2020.

[91] M. Wang, H. El-Fiqi, J. Hu, and H. A. Abbass, "Convolutional neural networks using dynamic functional connectivity for EEG-based person identification in diverse human states," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 12, pp. 3259–3272, 2019.

[92] B. Platt and G. Riedel, "The cholinergic system, EEG and sleep," *Behavioural Brain Research*, vol. 221, no. 2, pp. 499–504, 2011, the cholinergic system and brain function.

[93] D. Aeschbach, J. R. Matthews, T. T. Postolache, M. A. Jackson, H. A. Giesen, and T. A. Wehr, "Two circadian rhythms in the human electroencephalogram during wakefulness," *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, vol. 277, no. 6, pp. R1771–R1779, 1999.

[94] T. M. Sokhadze, R. L. Cannon, and D. L. Trudeau, "EEG biofeedback as a treatment for substance use disorders: review, rating of efficacy, and recommendations for further research," *Applied psychophysiology and biofeedback*, vol. 33, pp. 1–28, 2008.

[95] T. C. Ferree, P. Luu, G. S. Russell, and D. M. Tucker, "Scalp electrode impedance, infection risk, and EEG data quality," *Clinical Neurophysiology*, vol. 112, no. 3, pp. 536–544, 2001.

[96] A. A. Fingelkurts, A. A. Fingelkurts, and S. Kähkönen, "Functional connectivity in the brain—is it an elusive concept?" *Neuroscience & Biobehavioral Reviews*, vol. 28, no. 8, pp. 827–836, 2005.

[97] K. J. Friston, "Functional and effective connectivity: a review," *Brain connectivity*, vol. 1, no. 1, pp. 13–36, 2011.

[98] T. Behrouzi and D. Hatzinakos, "Graph variational auto-encoder for deriving EEG-based graph embedding," *Pattern Recognition*, vol. 121, p. 108202, 2022.

[99] A. Riera, A. Soria-Frisch, M. Caparrini, C. Grau, and G. Ruffini, "Unobtrusive biometric system based on electroencephalogram analysis," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, pp. 1–8, 2007.

[100] T. Schons, G. J. Moreira, P. H. Silva, V. N. Coelho, and E. J. Luz, "Convolutional network for EEG-based biometric," in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications: 22nd Iberoamerican Congress, CIARP 2017, Valparaíso, Chile, November 7–10, 2017, Proceedings 22*. Springer, 2018, pp. 601–608.

[101] Y. Sun, F. P.-W. Lo, and B. Lo, "EEG-based user identification system using 1d-convolutional long short-term memory neural networks," *Expert Systems with Applications*, vol. 125, pp. 259–267, 2019.

[102] Z. Wu, Z. Liu, J. Lin, Y. Lin, and S. Han, "Lite transformer with long-short range attention," 2020.

[103] Y. Wang, Y. Yang, J. Bai, M. Zhang, J. Bai, J. Yu, C. Zhang, G. Huang, and Y. Tong, "Evolving attention with residual convolutions," 2021.

[104] S. Lemm, B. Blankertz, G. Curio, and K.-R. Muller, "Spatio-spectral filters for improving the classification of single trial EEG," *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 9, pp. 1541–1548, 2005.

[105] R. Jenke, A. Peer, and M. Buss, "Feature extraction and selection for emotion recognition from EEG," *IEEE Transactions on Affective Computing*, vol. 5, no. 3, pp. 327–339, 2014.

[106] Y.-P. Lin, C.-H. Wang, T.-P. Jung, T.-L. Wu, S.-K. Jeng, J.-R. Duann, and J.-H. Chen, "Eeg-based emotion recognition in music listening," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 7, pp. 1798–1806, 2010.

[107] M. Akin, "Comparison of wavelet transform and fft methods in the analysis of EEG signals," *Journal of Medical Systems*, vol. 26, pp. 241–247, 2002.

[108] E. H. Houssein, A. Hammad, and A. A. Ali, "Human emotion recognition from eeg-based brain–computer interface using machine learning: a comprehensive review," *Neural Computing and Applications*, vol. 34, no. 15, pp. 12 527–12 557, 2022.

[109] M. Miao, L. Zheng, B. Xu, Z. Yang, and W. Hu, "A multiple frequency bands parallel spatial–temporal 3d deep residual learning framework for eeg-based emotion recognition," *Biomedical Signal Processing and Control*, vol. 79, p. 104141, 2023.

[110] J. Liu, G. Wu, Y. Luo, S. Qiu, S. Yang, W. Li, and Y. Bi, "Eeg-based emotion classification using a deep neural network and sparse autoencoder," *Frontiers in Systems Neuroscience*, vol. 14, pp. 1–43, 2020.

[111] Y. Li, B. Fu, F. Li, G. Shi, and W. Zheng, "A novel transferability attention neural network model for EEG emotion recognition," *Neurocomputing*, vol. 447, pp. 92–101, 2021.

[112] C. Wei, L.-l. Chen, Z.-z. Song, X.-g. Lou, and D.-d. Li, "Eeg-based emotion recognition using simple recurrent units network and ensemble learning," *Biomedical Signal Processing and Control*, vol. 58, p. 101756, 2020.

[113] S. Liu, Z. Wang, Y. An, J. Zhao, Y. Zhao, and Y.-D. Zhang, "EEG emotion recognition based on the attention mechanism and pre-trained convolution capsule network," *Knowledge-Based Systems*, vol. 265, p. 110372, 2023.

[114] D. Li, L. Xie, Z. Wang, and H. Yang, "Brain emotion perception inspired EEG emotion recognition with deep reinforcement learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. Early Access, pp. 1–14, 2023.

[115] J. Hu, C. Wang, Q. Jia, Q. Bu, R. Sutcliffe, and J. Feng, "Scalingnet: extracting features from raw EEG data for emotion recognition," *Neurocomputing*, vol. 463, pp. 177–184, 2021.

[116] Y. Li, L. Wang, W. Zheng, Y. Zong, L. Qi, Z. Cui, T. Zhang, and T. Song, "A novel bi-hemispheric discrepancy model for EEG emotion recognition," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 13, no. 2, pp. 354–367, 2020.

[117] T. Zhang, W. Zheng, Z. Cui, Y. Zong, and Y. Li, "Spatial–temporal recurrent neural network for emotion recognition," *IEEE Transactions on Cybernetics*, vol. 49, no. 3, pp. 839–847, 2018.

[118] W. Guo, G. Xu, and Y. Wang, "Multi-source domain adaptation with spatio-temporal feature extractor for EEG emotion recognition," *Biomedical Signal Processing and Control*, vol. 84, p. 104998, 2023.

[119] Y. Yang, Q. Wu, M. Qiu, Y. Wang, and X. Chen, "Emotion recognition from multi-channel EEG through parallel convolutional recurrent neural network," in *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2018, pp. 1–7.

[120] X. Du, C. Ma, G. Zhang, J. Li, Y.-K. Lai, G. Zhao, X. Deng, Y.-J. Liu, and H. Wang, "An efficient lstm network for emotion recognition from multichannel EEG signals," *IEEE Transactions on Affective Computing*, vol. 13, no. 3, pp. 1528–1540, 2020.

[121] S. Liu, Y. Zhao, Y. An, J. Zhao, S.-H. Wang, and J. Yan, "Glfanet: A global to local feature aggregation network for EEG emotion recognition," *Biomedical Signal Processing and Control*, vol. 85, p. 104799, 2023.

[122] P. Zhong, D. Wang, and C. Miao, "Eeg-based emotion recognition using regularized graph neural networks," *IEEE Transactions on Affective Computing*, vol. 13, no. 3, pp. 1290–1301, 2020.

[123] T. Song, W. Zheng, P. Song, and Z. Cui, "EEG emotion recognition using dynamical graph convolutional neural networks," *IEEE Transactions on Affective Computing*, vol. 11, no. 3, pp. 532–541, 2018.

[124] Y. Zhou, F. Li, Y. Li, Y. Ji, G. Shi, W. Zheng, L. Zhang, Y. Chen, and R. Cheng, "Progressive graph convolution network for EEG emotion recognition," *Neurocomputing*, vol. 544, p. 126262, 2023.

[125] Y. Yin, X. Zheng, B. Hu, Y. Zhang, and X. Cui, "EEG emotion recognition using fusion model of graph convolutional neural networks and lstm," *Applied Soft Computing*, vol. 100, p. 106954, 2021.

[126] P. Comon, "Independent component analysis, a new concept?" *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994, higher Order Statistics.

[127] J. Ruan, X. Wu, B. Zhou, X. Guo, and Z. Lv, "An automatic channel selection approach for ica-based motor imagery brain computer interface," *Journal of Medical Systems*, vol. 42, pp. 1–13, 2018.

[128] V. S. Handiru and V. A. Prasad, "Optimized bi-objective eeg channel selection and cross-subject generalization with brain–computer interfaces," *IEEE Transactions on Human-Machine Systems*, vol. 46, no. 6, pp. 777–786, 2016.

[129] H. Varsehi and S. M. P. Firoozabadi, "An eeg channel selection method for motor imagery based brain–computer interface and neurofeedback using granger causality," *Neural Networks*, vol. 133, pp. 193–206, 2021.

[130] A. K. Seth, A. B. Barrett, and L. Barnett, "Granger causality analysis in neuroscience and neuroimaging," *Journal of Neuroscience*, vol. 35, no. 8, pp. 3293–3297, 2015.

[131] P. Gaur, H. Gupta, A. Chowdhury, K. McCreadie, R. B. Pachori, and H. Wang, "A sliding window common spatial pattern for enhancing motor imagery classification in eeg-bci," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–9, 2021.

[132] C. Park, C. C. Took, and D. P. Mandic, "Augmented complex common spatial patterns for classification of noncircular eeg from motor imagery tasks," *IEEE Transactions on neural systems and rehabilitation engineering*, vol. 22, no. 1, pp. 1–10, 2013.

[133] B. Xu, L. Zhang, A. Song, C. Wu, W. Li, D. Zhang, G. Xu, H. Li, and H. Zeng, "Wavelet transform time-frequency image and convolutional network-based motor imagery eeg classification," *Ieee Access*, vol. 7, pp. 6084–6093, 2018.

[134] M. Tolić and F. Jović, "Classification of wavelet transformed eeg signals with neural network for imagined mental and motor tasks," *Kinesiology*, vol. 45, no. 1., pp. 130–138, 2013.

[135] M. H. Alomari, A. AbuBaker, A. Turani, A. M. Baniyounes, and A. Manasreh, "Eeg mouse: A machine learning-based brain computer interface," *International Journal of Advanced Computer Science and Applications*, vol. 5, no. 4, 2014.

[136] Y. Kim, J. Ryu, K. K. Kim, C. C. Took, D. P. Mandic, and C. Park, "Motor imagery classification using mu and beta rhythms of eeg with strong uncorrelating transform based complex common spatial patterns," *Computational intelligence and neuroscience*, vol. 2016, no. 1, p. 1489692, 2016.

[137] G. Zhang, V. Davoodnia, A. Sepas-Moghaddam, Y. Zhang, and A. Etemad, "Classification of hand movements from eeg using a deep attention-based lstm network," *IEEE Sensors Journal*, vol. 20, no. 6, pp. 3113–3122, 2019.

[138] A. Loboda, A. Margineanu, G. Rotariu, and A. M. Lazar, "Discrimination of eeg-based motor imagery tasks by means of a simple phase information method," *International Journal of Advanced Research in Artificial Intelligence*, vol. 3, no. 10, 2014.

[139] D. Zhang, K. Chen, D. Jian, and L. Yao, "Motor imagery classification via temporal attention cues of graph embedded eeg signals," *IEEE journal of biomedical and health informatics*, vol. 24, no. 9, pp. 2570–2579, 2020.

[140] H. Dose, J. S. Møller, H. K. Iversen, and S. Puthusserypady, "An end-to-end deep learning approach to mi-eeg signal classification for bcis," *Expert Systems with Applications*, vol. 114, pp. 532–542, 2018.

[141] W. Huang, W. Chang, G. Yan, Z. Yang, H. Luo, and H. Pei, "Eeg-based motor imagery classification using convolutional neural networks with local reparameterization trick," *Expert Systems with Applications*, vol. 187, p. 115968, 2022.

[142] T. Karácsony, J. P. Hansen, H. K. Iversen, and S. Puthusserypady, "Brain computer interface for neuro-rehabilitation with deep learning classification and virtual reality feedback," in *Proceedings of the 10th Augmented Human International Conference 2019*, 2019, pp. 1–8.

[143] C. Stangor and J. Walinga, "Introduction to psychology," 2014.

[144] T. C. Gent, C. L. Bassetti, and A. R. Adamantidis, "Sleep-wake control and the thalamus," *Current opinion in neurobiology*, vol. 52, pp. 188–197, 2018.

[145] L. M. Ward, "The thalamus: gateway to the mind," *Wiley Interdisciplinary Reviews: Cognitive Science*, vol. 4, no. 6, pp. 609–622, 2013.

[146] C. M. Sweeney-Reed, L. Buentjen, J. Voges, F. C. Schmitt, T. Zaehle, J. W. Kam, J. Kaufmann, H.-J. Heinze, H. Hinrichs, R. T. Knight *et al.*, "The role of the anterior nuclei of the thalamus in human memory processing," *Neuroscience & Biobehavioral Reviews*, vol. 126, pp. 146–158, 2021.

[147] D. Perry, T. Hendler, and S. G. Shamay-Tsoory, "Projecting memories: the role of the hippocampus in emotional mentalizing," *Neuroimage*, vol. 54, no. 2, pp. 1669–1676, 2011.

[148] R. P. Kesner, M. R. Hunsaker, and W. Ziegler, "The role of the dorsal and ventral hippocampus in olfactory working memory," *Neurobiology of learning and memory*, vol. 96, no. 2, pp. 361–366, 2011.

[149] M. Çırak, K. Yağmurlu, K. N. Kearns, E. C. Ribas, K. Urgun, M. E. Shaffrey, and M. Y. S. Kalani, "The caudate nucleus: its connections, surgical implications, and related complications," *World neurosurgery*, vol. 139, pp. e428–e438, 2020.

[150] P. Risold, R. Thompson, and L. Swanson, "The structural organization of connections between hypothalamus and cerebral cortex," *Brain research reviews*, vol. 24, no. 2-3, pp. 197–254, 1997.

[151] M. Glickstein and K. Doron, "Cerebellum: connections and functions," *The Cerebellum*, vol. 7, pp. 589–594, 2008.

[152] K. S. Traylor and B. F. Branstetter, "Cranial nerve anatomy," *Neuroimaging Clinics*, vol. 32, no. 3, pp. 565–576, 2022.

[153] E. F. Wijdicks, "The ascending reticular activating system," *Neurocritical care*, vol. 31, pp. 419–422, 2019.

[154] W. contributors, "Cranial nerves — Wikipedia, the free encyclopedia," https://en.wikipedia.org/w/index.php?title=Cranial_nerves&oldid=1252834288, 2024.

[155] S. Herculano-Houzel, "The human brain in numbers: a linearly scaled-up primate brain," *Frontiers in human neuroscience*, vol. 3, p. 857, 2009.

[156] N. Palomero-Gallagher and K. Zilles, "Cortical layers: Cyto-, myelo-, receptor- and synaptic architecture in human cortical areas," *Neuroimage*, vol. 197, pp. 716–741, 2019.

[157] K. Zilles, N. Palomero-Gallagher, and K. Amunts, *Myeloarchitecture and Maps of the Cerebral Cortex*. Elsevier, 12 2015, vol. 2, pp. 137–156.

[158] T. Kirschstein and R. Köhling, "What is the source of the eeg?" *Clinical EEG and neuroscience*, vol. 40, no. 3, pp. 146–149, 2009.

[159] F. Walter, F. Röhrbein, and A. Knoll, "Computation by time," *Neural Processing Letters*, vol. 44, pp. 103–124, 2016.

[160] Y. Kubota, "Untangling gabaergic wiring in the cortical microcircuit," *Current opinion in neurobiology*, vol. 26, pp. 7–14, 2014.

[161] C. Chayer and M. Freedman, "Frontal lobe functions," *Current neurology and neuroscience reports*, vol. 1, no. 6, pp. 547–552, 2001.

[162] L. Fogassi and G. Luppino, "Motor functions of the parietal lobe," *Current Opinion in Neurobiology*, vol. 15, no. 6, pp. 626–631, 2005, motor sytems / Neurobiology of behaviour.

[163] L. Cohen, S. Dehaene, F. Chochon, S. Lehericy, and L. Naccache, "Language and calculation within the parietal lobe: a combined cognitive, anatomical and fmri study," *Neuropsychologia*, vol. 38, no. 10, pp. 1426–1440, 2000.

[164] M. A. L. Ralph, E. Jefferies, K. Patterson, and T. T. Rogers, "The neural and computational bases of semantic cognition," *Nature reviews neuroscience*, vol. 18, no. 1, pp. 42–55, 2017.

[165] M. T. de Schotten, M. Urbanski, R. Valabregue, D. J. Bayle, and E. Volle, "Subdivision of the occipital lobes: an anatomical and functional mri connectivity study," *Cortex*, vol. 56, pp. 121–137, 2014.

[166] L. Donnelly, "The brain: functional divisions," *Anaesthesia & Intensive Care Medicine*, vol. 24, no. 6, pp. 358–363, 2023.

[167] J. Friedrich, M. Mückschel, and C. Beste, "Somatosensory lateral inhibition processes modulate motor response inhibition-an eeg source localization study," *Scientific Reports*, vol. 7, no. 1, p. 4454, 2017.

[168] J. T. Nicholas, A. M. Johannessen, and T. Van Nunen, *Tactile working memory scale: A professional manual*. Nordens välfärdscenter/Nordic Welfare Centre, 2019.

[169] J. Holsheimer and B. Feenstra, "Volume conduction and eeg measurements within the brain: a quantitative approach to the influence of electrical spread on the linear relationship of activity measured at different locations," *Electroencephalography and clinical neurophysiology*, vol. 43, no. 1, pp. 52–58, 1977.

[170] S. B. Rutkove, "Introduction to volume conduction," *The clinical neurophysiology primer*, pp. 43–53, 2007.

[171] R. Armitage, G. J. Emslie, R. F. Hoffmann, J. Rintelmann, and A. J. Rush, "Delta sleep eeg in depressed adolescent females and healthy controls," *Journal of affective disorders*, vol. 63, no. 1-3, pp. 139–148, 2001.

[172] S. Karakaş, "A review of theta oscillation and its functional correlates," *International Journal of Psychophysiology*, vol. 157, pp. 82–99, 2020.

[173] N. Fallon, Y. Chiu, T. Nurmikko, and A. Stancak, "Altered theta oscillations in resting eeg of fibromyalgia syndrome patients," *European journal of pain*, vol. 22, no. 1, pp. 49–57, 2018.

[174] S. M. Shadli, L. C. Ando, J. McIntosh, V. Lodhia, B. R. Russell, I. J. Kirk, P. Glue, and N. McNaughton, "Right frontal anxiolytic-sensitive eeg 'theta'rhythm in the stop-signal task is a theory-based anxiety disorder biomarker," *Scientific reports*, vol. 11, no. 1, p. 19746, 2021.

[175] M. H. MacLean, K. M. Arnell, and K. A. Cote, "Resting eeg in alpha and beta bands predicts individual differences in attentional blink magnitude," *Brain and cognition*, vol. 78, no. 3, pp. 218–229, 2012.

[176] R. J. Somsen, B. J. van't Klooster, M. W. van der Molen, H. M. van Leeuwen, and R. Licht, "Growth spurts in brain maturation during middle childhood as indexed by eeg power spectra," *Biological psychology*, vol. 44, no. 3, pp. 187–209, 1997.

[177] M. Gola, M. Magnuski, I. Szumska, and A. Wróbel, "Eeg beta band activity is related to attention and attentional deficits in the visual performance of elderly subjects," *International Journal of Psychophysiology*, vol. 89, no. 3, pp. 334–341, 2013.

[178] C. Spironelli, M. Manfredi, and A. Angrilli, "Beta eeg band: A measure of functional brain damage and language reorganization in aphasic patients after recovery," *Cortex*, vol. 49, no. 10, pp. 2650–2660, 2013.

[179] P. J. Uhlhaas, C. Haenschel, D. Nikolić, and W. Singer, "The role of oscillations and synchrony in cortical networks and their putative relevance for the pathophysiology of schizophrenia," *Schizophrenia bulletin*, vol. 34, no. 5, pp. 927–943, 2008.

[180] A. Lutz, L. L. Greischar, N. B. Rawlings, M. Ricard, and R. J. Davidson, "Long-term meditators self-induce high-amplitude gamma synchrony during mental practice," *Proceedings of the national Academy of Sciences*, vol. 101, no. 46, pp. 16 369–16 373, 2004.

[181] X. Liu, S. Liu, M. Li, F. Su, S. Chen, Y. Ke, and D. Ming, "Altered gamma oscillations and beta–gamma coupling in drug-naive first-episode major depressive disorder: association with sleep and cognitive disturbance," *Journal of Affective Disorders*, vol. 316, pp. 99–108, 2022.

[182] J. S. Kwon, B. F. O'Donnell, G. V. Wallenstein, R. W. Greene, Y. Hirayasu, P. G. Nestor, M. E. Hasselmo, G. F. Potts, M. E. Shenton, and R. W. McCarley, "Gamma frequency–range abnormalities to auditory stimulation in schizophrenia," *Archives of general psychiatry*, vol. 56, no. 11, pp. 1001–1005, 1999.

[183] S. Rajaraman, J. A. Bragg, J. D. Ross, and M. G. Allen, "Micromachined three-dimensional electrode arrays for transcutaneous nerve tracking," *Journal of Micromechanics and Microengineering*, vol. 21, no. 8, p. 085014, 2011.

[184] M. Seeck, L. Koessler, T. Bast, F. Leijten, C. Michel, C. Baumgartner, B. He, and S. Beniczky, "The standardized eeg electrode array of the ifcn," *Clinical neurophysiology*, vol. 128, no. 10, pp. 2070–2077, 2017.

[185] R. W. Homan, J. Herman, and P. Purdy, "Cerebral location of international 10–20 system electrode placement," *Electroencephalography and clinical neurophysiology*, vol. 66, no. 4, pp. 376–382, 1987.

[186] I. Rosenzweig, A. Fogarasi, B. Johnsen, J. Alving, M. E. Fabricius, M. Scherg, M. Y. Neufeld, R. Pressler, T. W. Kjaer, W. van Emde Boas *et al.*, "Beyond the double banana: improved recognition of temporal lobe seizures in long-term eeg," *Journal of Clinical Neurophysiology*, vol. 31, no. 1, pp. 1–9, 2014.

[187] L. Koessler, T. Cecchin, S. Colnat-Coulbois, J.-P. Vignal, J. Jonas, H. Vespignani, G. Ramantani, and L. G. Maillard, "Catching the invisible: mesial temporal source contribution to simultaneous eeg and seeg recordings," *Brain topography*, vol. 28, pp. 5–20, 2015.

[188] W. Commons, "electrodes of international 10-20 system for eeg — wikimedia commons, the free media repository," 2023.

[189] ——, "File:eeg 10-10 system.svg — wikimedia commons, the free media repository," 2023.

[190] T. W. Picton, "The p300 wave of the human event-related potential," *Journal of clinical neurophysiology*, vol. 9, no. 4, pp. 456–479, 1992.

[191] A. Bashashati, M. Fatourechi, R. K. Ward, and G. E. Birch, "A survey of signal processing algorithms in brain–computer interfaces based on electrical brain signals," *Journal of Neural engineering*, vol. 4, no. 2, p. R32, 2007.

[192] F.-B. Vialatte, M. Maurice, J. Dauwels, and A. Cichocki, "Steady-state visually evoked potentials: focus on essential paradigms and future perspectives," *Progress in neurobiology*, vol. 90, no. 4, pp. 418–438, 2010.

[193] Y. Sun, X. Chen, B. Liu, L. Liang, Y. Wang, S. Gao, and X. Gao, "Signal acquisition of brain-computer interfaces: a medical-engineering crossover perspective review," *Fundamental Research*, 2024.

[194] P. Comon, "Independent component analysis, a new concept?" *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994, higher Order Statistics.

[195] J. Onton, M. Westerfield, J. Townsend, and S. Makeig, "Imaging human eeg dynamics using independent component analysis," *Neuroscience & Biobehavioral Reviews*, vol. 30, no. 6, pp. 808–822, 2006, methodological and Conceptual Advances in the Study of Brain-Behavior Dynamics: A Multivariate Lifespan Perspective.

[196] T.-W. Lee, "Independent component analysis," in *Independent component analysis*. Springer, 1998, pp. 27–66.

[197] J. Dien, "Issues in the application of the average reference: Review, critiques, and recommendations," *Behavior Research Methods, Instruments, & Computers*, vol. 30, no. 1, pp. 34–43, 1998.

[198] R. B. Blackman and J. W. Tukey, "The measurement of power spectra from the point of view of communications engineering—part i," *Bell System Technical Journal*, vol. 37, no. 1, pp. 185–282, 1958.

[199] J.-P. Lachaux, E. Rodriguez, J. Martinerie, and F. J. Varela, "Measuring phase synchrony in brain signals," *Human Brain Mapping*, vol. 8, no. 4, pp. 194–208, 1999.

[200] M. Fey, J. E. Lenssen, F. Weichert, and H. Müller, "Splinecnn: Fast geometric deep learning with continuous b-spline kernels," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 869–877.

[201] L. Piegl and W. Tiller, *The NURBS book*. Springer Science & Business Media, 2012.

[202] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[203] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International conference on machine learning*. pmlr, 2015, pp. 448–456.

[204] M. Limper, Y. Jung, J. Behr, T. Sturm, T. Franke, K. Schwenk, and A. Kuijper, "Fast, progressive loading of binary-encoded declarative-3d web content," *IEEE computer graphics and applications*, vol. 33, no. 5, pp. 26–36, 2013.

[205] B. Fanini and L. Cinque, "Encoding immersive sessions for online, interactive VR analytics," *Virtual Reality*, vol. 24, no. 3, pp. 423–438, 2020.

[206] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional lstm and other neural network architectures," *Neural Networks*, vol. 18, no. 5-6, pp. 602–610, 2005.

[207] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[208] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[209] T. Takase, "Dynamic batch size tuning based on stopping criterion for neural network training," *Neurocomputing*, vol. 429, pp. 1–11, 2021.

[210] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals," *circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[211] G. Huang, Z. Hu, W. Chen, S. Zhang, Z. Liang, L. Li, L. Zhang, and Z. Zhang, "M3cv: A multi-subject, multi-session, and multi-task database for EEG-based biometrics challenge," *NeuroImage*, vol. 264, p. 119666, 2022.

[212] M. R. Nuwer, G. Comi, R. Emerson, A. Fuglsang-Frederiksen, J.-M. Guérit, H. Hinrichs, A. Ikeda, F. Jose C. Luccas, and P. Rappelsburger, "IFCN standards for digital recording of clinical EEG," *Electroencephalography and Clinical Neurophysiology*, vol. 106, no. 3, pp. 259–261, 1998.

[213] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, "PyTorch: An imperative style, high-performance deep learning library," *Advances in neural information processing systems*, vol. 32, 2019.

[214] M. Fey and J. E. Lenssen, "Fast graph representation learning with PyTorch geometric," *arXiv preprint arXiv:1903.02428*, 2019.

[215] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[216] D. Avola, M. Cascio, L. Cinque, A. Fagioli, G. L. Foresti, M. R. Marini, and D. Pannone, "Analyzing EEG data with machine and deep learning: A benchmark," in *Proceedings of the International Conference on Neural Image Analysis and Processing (ICIAP)*. Springer, Cham, 2022, pp. 335–345.

[217] S. Khan, M. Naseer, M. Hayat, S. W. Zamir, F. S. Khan, and M. Shah, "Transformers in vision: A survey," *ACM computing surveys (CSUR)*, vol. 54, no. 10s, pp. 1–41, 2022.

[218] A. Gramfort, M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier, C. Brodbeck, R. Goj, M. Jas, T. Brooks, L. Parkkonen *et al.*, "Meg and eeg data analysis with mne-python," *Frontiers in Neuroinformatics*, vol. 7, p. 267, 2013.

[219] H. Dalianis and H. Dalianis, "Evaluation metrics and evaluation," *Clinical Text Mining: secondary use of electronic patient records*, pp. 45–53, 2018.

[220] M. L. McHugh, "Interrater reliability: the kappa statistic," *Biochemia medica*, vol. 22, no. 3, pp. 276–282, 2012.

[221] M. Sun, Z. Song, X. Jiang, J. Pan, and Y. Pang, "Learning pooling for convolutional neural network," *Neurocomputing*, vol. 224, pp. 96–104, 2017.

[222] A. R. Javed, S. Ur Rehman, M. U. Khan, M. Alazab, and T. Reddy, "Can-intelliids: Detecting in-vehicle intrusion attacks on a controller area network using CNN and attention-based GRU," *IEEE transactions on network science and engineering*, vol. 8, no. 2, pp. 1456–1466, 2021.

[223] X. Yin, Z. Liu, D. Liu, and X. Ren, "A novel cnn-based bi-lstm parallel model with attention mechanism for human activity recognition with noisy data," *Scientific Reports*, vol. 12, no. 1, p. 7878, 2022.

[224] M. H. Rafiei and H. Adeli, "A new neural dynamic classification algorithm," *IEEE Trans Neural Netw Learn Syst*, vol. 28, no. 12, pp. 3074–3083, 2017.

[225] K. M. R. Alam, N. Siddique, and H. Adeli, "A dynamic ensemble learning algorithm for neural networks," *Neural Computing and Applications*, vol. 32, no. 12, pp. 8675–8690, 2020.

[226] D. R. Pereira, M. A. Piteri, A. N. Souza, J. P. Papa, and H. Adeli, "FEMa: a finite element machine for fast learning," *Neural Computing and Applications*, vol. 32, no. 10, pp. 6393–6404, 2020.

[227] Z. Cai, L. Wang, M. Guo, G. Xu, L. Guo, and Y. Li, "From intricacy to conciseness: A progressive transfer strategy for EEG-Based Cross-Subject emotion recognition," *International Journal of Neural Systems*, vol. 32, no. 3, p. 2250005, 2022.

[228] B. García-Martínez, A. Fernández-Caballero, A. Martínez-Rodrigo, R. Alcaraz, and P. Novais, "Evaluation of brain functional connectivity from electroencephalographic signals under different emotional states," *International Journal of Neural Systems*, vol. 32, no. 10, p. 2250026, 2022.

[229] M. H. Rafiei, L. V. Gauthier, H. Adeli, and D. Takabi, "Self-Supervised learning for electroencephalography," *IEEE Transactions on Neural Networks and Learning Systems*, vol. PP, 2022.

[230] R. J. Vlek, D. Steines, D. Szibbo, A. Kübler, M.-J. Schneider, P. Haselager, and F. Nijboer, "Ethical issues in brain–computer interface research, development, and dissemination," *Journal of neurologic physical therapy*, vol. 36, no. 2, pp. 94–99, 2012.

[231] M. Farisco, S. Laureys, and K. Evers, "Externalization of consciousness. scientific possibilities and clinical implications," *Ethical issues in behavioral neuroscience*, pp. 205–222, 2015.