Electroencephalogram Hyperscanning-based Brain-Computer Interfacing

Alicia Falcon Caro

School of Science and Technology

A thesis submitted in partial fulfilment of the requirements of Nottingham Trent University for the degree of

Doctor of Philosophy

February 2025

To the memory of my father.

This achievement is as much yours as it is mine.

Acknowledgements

I would like to express my deepest gratitude to my supervisor, Prof. Saeid Sanei, for his continuous guidance and support throughout these years. His expertise, patience, and encouragement have been invaluable in helping me navigate the challenges of this research project.

I extend my sincere thanks to my other supervisor, Dr. João Filipe Ferreira, for his insightful advice, and constructive suggestions, particularly during the final stages of this journey.

My thanks also to my colleagues, whose additional advice have enriched this thesis.

I am grateful to the Department of Computer Science at Nottingham Trent University for their financial sponsorship and support of this dissertation.

I also wish to thank all the volunteers who participated in the EEG recordings. Your selfless contributions made this research possible.

Finally, I would like to dedicate special thanks to my family. To my father, for his encouragement and inspiration, which I still feel even from afar. To my mother and sister, for their unconditional support and steadfast belief in me, especially during the most challenging times. Without your love and encouragement, I would not have been able to complete this dissertation.

Abstract

Electroencephalography (EEG) hyperscanning refers to simultaneous EEG recordings from multiple subjects. This scenario has become popular for studying group performances by social science activists to investigate human competing or collaborative actions. However, the use of EEG hyperscanning had not become practical in brain-computer interfacing (BCI) prior to this research.

In this thesis, we investigate the use of EEG hyperscanning for two BCI scenarios; one where the BCI tasks are performed in an uncontrolled environment in which the brain is naturally engaged in multi-task activities, and the objective is to enhance the BCI training accuracy. The other one is to assess how well a subject's brain can follow another brain. The latter scenario can be useful in brain rehabilitation, i.e., after stroke, or for investigating brain plasticity where new motor region substitutes the inactive zone.

For the first scenario, a new formulation for common spatial patterns for EEG hyperscanning (namely hyperCSP) has been proposed. It learns a rotating matrix to best derive the desired (common between multiple subjects) motor task and remove all the undesired (uncommon between the subjects) ones caused by the uncontrolled environment. It has been demonstrated that in an uncontrolled environment, learning from two subjects can significantly enhance the motor classification rates, achieving a best classification accuracy of 0.82 using an SVM classifier.

For the second scenario, the concept of adaptive cooperative networking has been exploited to investigate the collaboration or similarity

in function between multiple brains. A brain functional connectivity-informed single task diffusion adaptation has been used for this purpose, achieving the lowest error compared to competing methods. In this scenario, the EEG sensors for each subject are considered as the agents of a connected network, that cooperate to achieve a desired task prescribed by a second subject. The outcome of this achievement can boost brain rehabilitation process where the collaboration or difference between a healthy and a subject under rehabilitation is investigated.

Finally, given that the brain can be considered as a connected network, a prolonged physical task can be classified by assessing the variation in such network through cooperative networking using diffusion adaptation. In this study, the 3-D body movements are segmented and 1-D orthogonal vectors, in this case Bessel functions, are allocated to the segments. These orthogonal vectors are used as the targets for classification of the brain tasks, obtaining a best inter-subject validation accuracy of 0.83 under a non-ideal situation. The outcome of this research paves the path for the use of BCI with no feedback (forward BCI) for developing brain-driven cybernetic devices.

The research in this thesis is concluded by suggesting combination of hyperCSP and other proposed methods for obtaining inclusive BCI systems and a pipeline for real-time applications.

Each chapter of this thesis is followed by extensive sets of experiments and involves in-house data recordings (some made publicly available).

List of Publications

- A. Falcon-Caro, J.F. Ferreira, and S. Sanei, "Cooperative Identification of Prolonged Motor Movement from EEG for BCI without Feedback," *IEEE Access*, vol. 13, pp. 11765-11777, 2025.
- A. Falcon-Caro, E. Peytchev, and S. Sanei, "Adaptive network model for assisting people with disabilities through crowd monitoring and control," *Bioengineering*, vol. 11, no. 3, pp. 283, 2024.
- A. Falcon-Caro, S. Shirani, J.F. Ferreira, J. J. Bird, and S. Sanei, "Formulation of common spatial patterns for multi-task hyperscanning BCI," *IEEE Transactions on Biomedical Engineering*, vol. 71, no. 6, pp. 1950-1957, 2024.
- A. Falcon-Caro, M. Frincu, and S. Sanei, "A diffusion adaptation approach to model brain responses in an EEG-based hyperscanning study," in *2023 IEEE Statistical Signal Processing Workshop* (*SSP*), IEEE, 2023.
- A. Falcon-Caro and S. Sanei, "Gesture recognition via estimation of information exchange between muscles," in *2023 24th International Conference on Digital Signal Processing* (*DSP*), IEEE, 2023.
- A. Falcon-Caro and S. Sanei, "Cooperative networking approach to assisting blinds in a crowd using air trackers," in *2022 4th International Conference on Emerging Trends in Electrical, Electronic and Communications Engineering* (*ELECOM*), IEEE, 2022.
- A. Falcon-Caro and S. Sanei, "Diffusion adaptation for crowd analysis," in *2021 International Conference on e-Health and Bioengineering* (*EHB*),

IEEE, 2021.

Contents

List of Figures

List of Tables

Abbreviations and Symbols

Abbreviations

Acc Accuracy

AEP Auditory evoked potential

AIC Akaike information criterion

AR Autoregressive model

ATC Adapt-then-combine

AUC Area under the curve

BCI Brain-computer interface

BSS Blind source separation

CAR Common average referencing

CCSP Composite common spatial patterns

CLTCCSP Composite local temporal correlation common spatial patterns

CNN Convolutional neural networks

CNS Central nervous system

CSP Common spatial patterns

CTA Combine-then-adapt

DA Diffusion adaptation

DAOT Diffusion adaptation with orthogonal targets

DNNs Deep neural networks

DTF Directed transfer function

ECoG Electrocorticography

EEG Electroencephalography

EMD Empirical-mode decomposition

ERPs Event-related potentials

ErrP Error-related potential

ERS/ERD Event related synchronisation/desynchronisation

FB-CSP Filter bank common spatial patterns

FIR Finite impulse response

fMRI Functional magnetic resonance imaging

fNIRS Functional near infrared spectroscopy

GANs Generative adversarial networks

GC Granger causality

GNNs Graph neural networks

hyperCSP Hyperscanning common spatial patterns

ICA Independent component analysis

IS Imagery speech

KNN K-nearest neighbours

LDA Linear discriminant analysis

LMS Least-mean square

LSTM Long short-term memory networks

MEG Magnetoencephalography

MI Motor imagery

MMN Mismatch negativity

MSE Mean squared error

MVAR Multivariate autoregressive model

PCA Principal component analysis

PDC Partial directed coherence

PLI Phase-lag index

PLV Phase-locking value

ROC Receiver operating characteristic

RP Readiness potential

SCPs Slow cortical potentials

SNR Signal-to-noise ratio

SSVEPs Steady-state visually evoked potentials

SVM Support vector machines

VEP Visual evoked potential

List of Symbols

$\boldsymbol{\psi}_{k,i}$ Intermediate estimate of node k at time instant i in the diffusion adaptation formulation

$\boldsymbol{\Sigma'}_{x_k x_l}$ Residual covariance matrix between x_k and x_l of the reduced MVAR model

$\boldsymbol{\Sigma}_{x_k x_k}$ Residual covariance matrix between x_k and x_k of the MVAR model

$\boldsymbol{\tau}_{k,i}$ Location vector of node k relative to a global coordinate system at each time instant i in the crowd modelling scenario

$\boldsymbol{\varepsilon'}_{k,i}$ Cross-term residuals of the reduced MVAR model of node k at time instant i

$\boldsymbol{\varepsilon}_{k,i}$ Regression residuals for the MVAR model of node k at time instant i

$\boldsymbol{A'}_{kk,m}$ Reduced regression coefficients for the reduced MVAR model of node k

$\boldsymbol{A}_{kk,m}$ Regression coefficients for the MVAR model of node k

$\boldsymbol{A}_{kl,m}$ Coefficients that encapsulate the dependence of $\boldsymbol{X}_{k,i}$ on the past sample m of $\boldsymbol{X}_l$

$\boldsymbol{C_k}$ Covariance matrix for the subspace of the signals from task or subject k

$\boldsymbol{f}_a$ Normalised feature for subject or class a

$\boldsymbol{I}_{N_k}$ $N_k \times N_k$ identity matrix

$\boldsymbol{u}_{k,i}$ $1 \times M$ known regression vector, where M represents a non-negative scalar that allows us to set the size of the diffusion adaptation filter

$\boldsymbol{u}_{k,i}^T$ Transpose of $\boldsymbol{u}_{k,i}$

$\boldsymbol{v}_{k,i}$ Velocity vector of node k at time instant i in the crowd modelling scenario

$\boldsymbol{W}$ Filter coefficients that separate two classes of a brain activity, or the two populations on the CSP formulation

$\boldsymbol{w}$ Estimated target parameter in a distributed learning strategy, which contains the estimated spatial filter coefficient in a filtering approach

$\boldsymbol{w}_{k,i}$ Estimated target value and filtering coefficient for node k at time instant i

$\boldsymbol{X'}$ Filtered signal of X

$\boldsymbol{X}_i^{\eta}$ EEG sub-gesture η signal segment at time instant i

$\boldsymbol{X}_{k,i}$ MVAR model of node k at time instant i

$\boldsymbol{x}_{k,i}$ EEG signal of node k at time instant i

$\boldsymbol{x}_{k,i}^{\eta}$ EEG sub-gesture η signal segment of node k at time instant i

$\boldsymbol{Y}^{Q}$ Estimated EEG output segment representing gesture Q

$\boldsymbol{y}_{k,i}^{\eta}$ Estimated EEG sub-gesture η signal segment of node k at time instant i after applying diffusion adaptation

Δi Time step

Γ Gamma function

$\gamma_{k,i}$ Measurement noise of node k at time instant i

λ Lagrange multiplier

$\mathbb{E}$ Statistical expectation operator

$\mathbf{H}(f)$ Transfer matrix

$\mathbf{H}^{*}(f)$ Hermitian transpose of $\mathbf{H}(f)$

$\mathbf{I}_{l\rightarrow k}(f)$ Direct influence of node l on node k

$\mathbf{S}(f)$ Cross-spectral density matrix

μ Non-negative step size in the distributed learning strategies formulation, where μ_k represents the non-negative step size for node k

∇_w Gradient vector relative to the complex-conjugate transposition of $\boldsymbol{w}_{i-1}$

ν Constant parameter that determines the order of the Bessel function of the first kind

$\rho_{k,l}$ Regulation factor for the multitask diffusion adaptation formulation

$a_{l,k}$ Combination weights between node l and node k

D Dictionary of selected Bessel functions

d_i^{η} Desired target sub-gesture η at time instant i. The desired target sub-gesture takes the form of a Bessel function

$d_{k,i}$ Desired target parameter of node k at time instant i

E Euclidean distance or error between two values or signals

E_k Error of node k to achieve its target $d_{k,i}$

$E_{aggregrate}$ Aggregated error for all nodes

E_{norm} Normalised error

f Frequency

g In a cooperative hyperscanning scenario, it represents the nodes or electrodes from a second individual

G_{η} Sub-gesture η

i Time instant or point sample of a signal

$J(\boldsymbol{W})$ Traditional CSP cost function

$J_{\nu}(i)$ Bessel function of the first kind of order ν

$J_k(\boldsymbol{w})$ Cost function to estimate $\boldsymbol{w}$ for each node k

L EEG signal length or number of samples

L_s EEG segment length

L_w EEG window length

M_k Cluster of node k in the multitask diffusion adaptation formulation

N Number of nodes in a distributed network

n Non-negative integer that indexes the terms in the series representation of the Bessel function

N_k Set of neighbours of node k, including itself, in a distributed network

p MVAR model order

Q Gesture composed of η sub-gestures

S Number of EEG segments

y Dependent variable representing the solution to the Bessel's differential equation, denoting the function value at a given time i

$GC_{x_l \to x_k}$ Estimated Granger causality of node l on node k

Chapter 1

Introduction

Latest advances and improvements in neurorehabilitation techniques for patients with complete, partial, or chronic neurological injuries are not yet able to restore or improve damaged neural pathways, even more in the presence of brain plasticity. In these cases, where more traditional therapies are not enough, brain-computer interfaces (BCI), which are systems that allow people to interact with the outside world through the identification and classification of neural activities, have shown positive results as an alternative neurorehabilitative approach [1]. This approach can help restore the lost motor function, heal the damaged brain, and reduce the neurological deficits.

1.1 Brain Physiology and Motor Activity

The human brain is composed of billions of interconnected neurons that form a complex network, and is considered the most complex evolutionary biological system. This organ, together with the spinal cord, constitute the major components of the central nervous system (CNS), which is the main part of the human nervous system responsible for integrating and processing the sensory information, and coordinating the body responses based on this information.

In the CNS, the brain processes the sensory inputs as stimulus and generates neural signals as responses to these stimulus through the interaction of neurons via synapses, where neurotransmitters are released to propagate the signals. Then,

the spinal cord transmits these neural signals between the brain and the rest of the body.

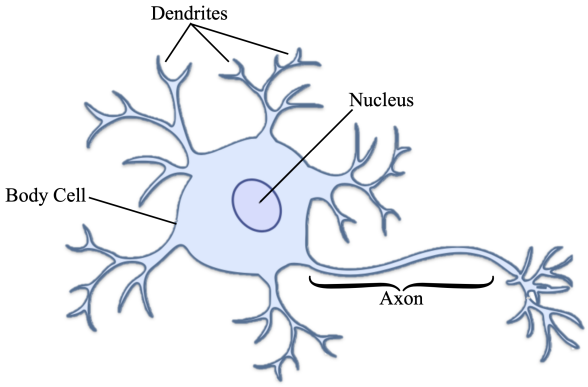

Figure 1.1: Diagram of a neuron representing its three main parts.

These neurons transmit electrical impulses, which are fundamental in all brain functions, and are formed of three main parts: the cell body, the axon, and the dendrites, as shown in Fig. 1.1. The cell body is the central part of the neuron that contains the nucleus, which stores the cell's genetic information. The axon and dendrites are thin prolongations of the neuron, where the axon transmits the electrical impulse from the neuron to other neurons, and the dendrites is used to receive the electrical impulse from other neurons. These two parts allow for the communication within the neural network. Therefore, neurons form networks that underlie all brain functions, and these networks are organised in specific ways within different brain regions, allowing the specialised functionality of each region.

Generally, the brain can be divided in three main parts: the cerebellum, located at the back of the brain and in charge of motor coordination, balance, and precision of movements; the brainstem, which connects the brain with the spinal cord and regulates many involuntary responses and vital functions, such as heartbeat and breathing; and the cerebrum, located at the uppermost region and responsible for higher cognitive functions, such as sensory perception, cognition, motor control, and decision-making. The cerebrum is divided into two hemispheres, left and right, each controlling the opposite side of the body. Each hemisphere is specialised in different functions and they can further be divided

into four main lobes: frontal, parietal, temporal, and occipital [2]. The frontal lobe is involved in decision-making, problem-solving, planning, voluntary movement and speech production, and in the control of emotions, social behaviour and personality. The parietal lobe processes sensory information from the body, such as touch, and contributes to the spatial orientation and perfection. The temporal lobe is involved in the processing of auditory information, and it is also crucial in memory formation and retrieval. Finally, the occipital lobe is mainly responsible for the processing of visual information. The cerebrum is also composed of other additional key components: the basal ganglia, a group of structures deep within the cerebrum involved in movement control and coordination; the white matter, which is the inner tissue of the cerebrum and connects different parts of the cortex and links the cortex and other brain regions and the spinal cord; the corpus callosum, a band of nerves that connect the left and right hemispheres, allowing for communication between them; and the cerebral cortex, which is the outer layer of the cerebrum and is involved in higher brain functions.

The cerebral cortex can further be divided into functional areas, referred to as brain regions, that correspond to specific sensations. During mental tasks, including motor execution, these brain regions engage in continuous communication to refine and execute the desired task. Some of the main cortex regions of interest in BCIs are [3]:

- Primary Motor Cortex: located in the frontal lobe, it is primarily responsible for executing voluntary motor commands. It contains the motor homunculus, which is a topographical map of the body where different parts of the cortex correspond to different muscle groups in the body.

- Supplementary Motor Area: located in the medial aspect of the frontal lobe, it is responsible for complex movements and the coordination of two-handed movements. This cortex area is associated with the initiation of movements that are internally guided rather than as a reaction to external stimuli.

- Primary Visual Cortex: located in the occipital lobe, it is responsible for the processing of visual stimuli.

- Primary Auditory Cortex: located in the temporal lobe, it is responsible for the processing of auditory stimuli.

A summary of different parts of the brain can be found in Fig. 1.2.

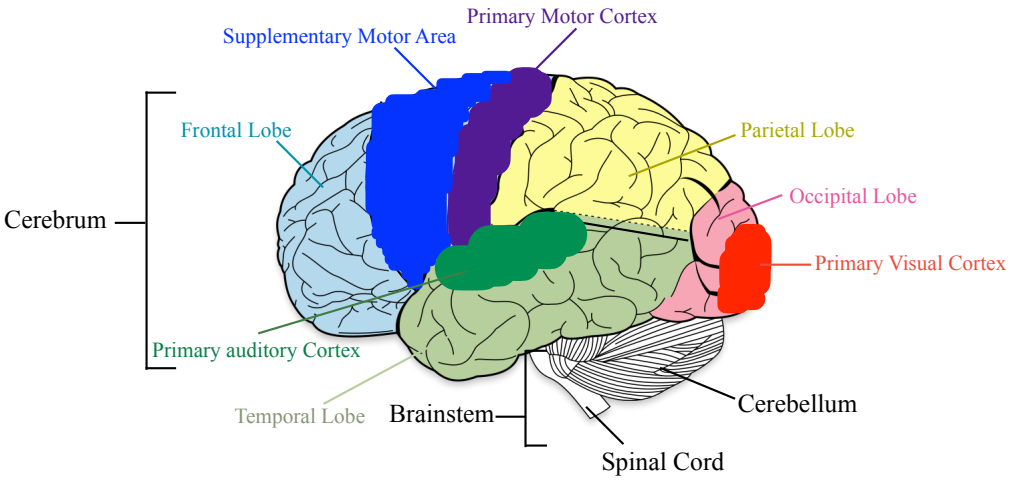

Figure 1.2: Visual representation of the different parts of the brain.

The visual and auditory cortex are highly involved in the development of BCIs since the majority of the currently developed BCIs are designed so the user receives feedback from the device or action performed. This is done so the user can understand how well they are controlling the BCI, they can adjust their mental strategy accordingly, and it is also easier for the system to recognise the neural activity that represents the user's desired task.

1.2 Brain-computer Interfaces

Traditionally, BCIs have been developed mainly for medical purposes, either as a rehabilitation tool or an assistive technology, such as through the development of neuroimaging-based prostheses [4], and have shown promising results in the

recovery of motor disorders, like stroke or spinal cord injuries, and the recognition, monitoring and improvement of cognitive diseases, such as Alzheimer's or Parkinson's [5–7]. However, thanks to the development of cheaper and more accessible high resolution non-invasive brain signal recording systems [8], and the improvement of advanced signal processing and machine learning techniques for biomedical signals in the last decade [9], BCIs have also started to be implemented in other fields for non-medical applications, such as in arts [10] and games [11].

We can divide BCIs in two types: passive and active BCIs. Passive BCIs are considered systems which deduce the user's cognitive or emotional state without the need for any intentional control. On the other hand, we can define active BCIs as systems that interpret the user's intentional brain activity associated with a specific task with the purpose of controlling an external device.

Some types of BCIs can be considered as passive or active interchangeably. Passive BCIs tend to be associated with systems focused on emotion recognition [12, 13], detection of stress or fatigue [14], and monitoring of cognitive workload [15]. Active BCIs are more commonly associated with systems related to error-related potential (ErrP), detecting brain responses to errors and unexpected events and commonly used to help fine-tune other BCIs by correcting their mistakes through a feedback mechanism; auditory evoked potential (AEP) and visual evoked potential (VEP), making use of brain's responses to auditory stimuli and visual impulses respectively; motor imagery (MI) and imagery speech (IS), which involves the mental execution of motor movements or imagined words and generate unique patterns in the brain associated with the tasks performed or imagined; and spellers, relying on the recognition of patterns in the brain responses associated with choosing specific letters or symbols, and can be used to help the user communicate [16–18].

1.2.1 Neuroimaging Techniques in BCI

The synchronised activity of neurons in the different regions of the brain can be detected by different neuroimaging techniques. Therefore, different neuroimaging systems can be used for the acquisition of brain signals for the development of BCIs, which can be divided into invasive and non-invasive methods. Under in-

vasive methods, the most common is electrocorticography (ECoG). On the other hand, under non-invasive methods, we can find functional near infrared spectroscopy (fNIRS), magnetoencephalography (MEG), functional magnetic resonance imaging (fMRI), and electroencephalography (EEG).

Although all these neuroimaging techniques can be used for the development of any of the previously mentioned passive and active BCIs, EEG is the most popular method for non-invasive BCIs, especially active BCIs, due to its low cost, high accessibility, high temporal resolution, and high portability [19].

1.2.2 Electroencephalography

The brain is composed of neurons that communicate through electrical impulses. These neurons use electrochemical processes, generating a small electrical charge, to generate and transmit signals across different brain regions. When a group of neurons generate electrical charge or signals simultaneously, they generate electromagnetic fields in the brain. These fields are mostly concentrated near the surface in the cerebral cortex, and this ultimately creates an electrical activity that can be detected externally.

EEG is able to capture this surface-level electrical activity using electrodes placed on the scalp to detect the brain's electrical fluctuations caused by the underlying neural activity, and therefore recording the brain's activity patterns using a more secure non-invasive technique. Even so, the recording of these signals from the scalp surface and not directly from the cerebral cortex, where they originate, means that the recorded signals are obtained after travelling through various brain anatomical layers. These layers are the cerebral cortex, where the brain signals are generated; the cerebrospinal fluid, which is a fluid layer surrounding the brain that slightly attenuates the signals as they pass through; the meninges, which are three protective membranes surrounding the brain; the skull, which is the bone layer that significantly attenuates the signal due to its low conductivity; and the scalp and skin, which are the outermost layers and can provide additional resistance and introduce noise from muscle and other non-brain sources to the recorded signal.

All these layers, especially the skull and scalp, distort the EEG signal, reducing

its amplitude and quality, and limiting the EEG spatial resolution. Furthermore, signals from different brain regions are distorted and might be mixed while passing through the different layers, which might result in the signals recorded at the scalp containing overlapping information from multiple sources, making it challenging to precisely localise the different sources, which is known as volume conduction.

The recorded brain activity differs depending on the brain region where it originates as well as the frequency of the electrical activity, generating distinct EEG patterns, which allows us to analyse how the brain functions and to obtain information from different cognitive and physiological states. Therefore, we can say that these electrical signals fall within different frequency bands. Although these frequency ranges are not exact, each approximate range is associated with different states of consciousness, cognitive processes, and brain functions, and may be linked to specific brain regions.

The main brain frequency bands are:

- Delta (0.5-4 Hz): have a high amplitude and are most prominent during deep relaxing and restorative stages of sleep.
- Theta (4-8 Hz): associated with light sleep and creative states. It is also linked to deep state of meditation and the subconscious mind.
- Alpha (8-13 Hz): associated with a relaxed, wakeful state, and is more prominent with eyes closed, when the mind is at rest but not asleep. Normal alpha elicits around the brain occipital region.
- Mu (8-10 Hz): part of the alpha frequency band. It is generated during the performance of voluntary motor imagery or actual movement. Contrary to the alpha rhythm, which is suppressed with eyes open, this rhythm does not change independently of the state of the eyes. Mu rhythm elicits within the brain motor area.
- Beta (14-30 Hz): associated with active, alert states, cognitive processing and active concentration. Beta elicits around the brain frontal region.
- Gamma (30-100 Hz): associated with higher-order cognitive functions, such as perception, consciousness, attention, and neural synchronisation across

different brain regions.

These frequencies are lower during both, childhood and old ages.

Due to their characteristics, the mu and beta rhythms are highly relevant and frequently used in the development of motor-related BCI.

Through the study of these frequency bands and their associated brain regions, we can better understand how different neural circuits contribute to specific mental states and behaviours, and study the dynamic interactions between different brain regions during specific mental tasks.

1.2.3 EEG-based BCIs

BCIs can use different EEG manifestations or patterns to decode the user intentions. Although multiple features can be used, the most commonly used in traditional BCI systems are the following seven:

- Event related synchronisation/desynchronisation (ERS/ERD) [20]: it reflects changes in the power of certain EEG waveforms, such as alpha, beta, and mu rhythms. In these systems, ERD reflects cortical activation during tasks executions, such as in MI applications, while ERS is a post-task phenomenon that indicates the return of the sensory-motor cortex to its resting state. These EEG patterns are highly relevant especially to motor BCIs used in neurorehabilitation.

- Readiness potential (RP) [21]: it precedes voluntary movements and reflects the planning and intention for a movement. This signal activity can also be detected during imagery movement, even when no physical movement occurs, so it is commonly used in BCIs designed for users with severe motor impairments.

- Event-related potentials (ERPs) [22]: Although other components could be used, the P300 is particularly commonly used in BCIs. The P300 wave is a positive deflection which occurs approximately 300 milliseconds after an infrequent or significant stimulus, and it can be divided into two subcomponents, P3a and P3b. P3a is associated with the attention to the

novel stimuli while P3b is linked with the decision-making processes. These features are highly reliable, making the P300-based BCIs ideal for non-motor applications such as spellers.

- Mismatch negativity (MMN) [23]: also based on ERPs, this feature reflects the brain automatic detection of deviations from an expected pattern in auditory stimuli. This response occurs within 100-250 milliseconds after a deviation instant making it a rapid and unconscious mechanism. This feature is commonly used for the development of BCIs that can adapt to environmental changes and support users with impaired attention, enabling responsive and intuitive interfaces.

- Steady-state visually evoked potentials (SSVEPs) [24]: these responses are triggered by flickering visual stimuli at specific frequencies, with the brain responding at the same frequencies. They are commonly used in applications that require high-speed communication, such as VEP-based BCIs.

- Slow cortical potentials (SCPs) [25]: they represent gradual shifts in cortical activity, which is commonly linked to arousal or attention cognitive states. These manifestations are particularly useful in neurofeedback and neurorehabilitation BCIs, where the user learns to modulate their brain activity to improve their focus or regain motor control.

- Error-related potential (ErrP) [26]: this manifestation occurs from the anterior cingulate cortex when the brain detects a mismatch between an intended action and the resulting outcome. These brain responses are particularly useful in the implementation of error-correction mechanisms in BCIs that enhance system responsiveness.

The use of any of these responses depends on the paradigm used during recording of the brain signals as well as the final desired application of the system. In some BCI scenarios a combination of these responses may be utilised.

1.2.4 Pros and Cons of using EEG for BCI

EEG-based BCI was introduced in 1973, by J. Vidal [27]. Since then, other neuroimaging techniques, including fMRI, MEG, and fNIRS, have been used in various BCI paradigms.

ECoG, the predominant invasive EEG recording, has been proven to provide a better spatial resolution and a higher quality signals for BCI, allowing for a more precise identification of neural activity. However, this method requires the patient to undergo a surgery for electrode implantation, which has its own associated risks. Therefore, this technique is mainly only used for clinical or therapeutic applications where the benefits highly outweigh the risks, such as for epilepsy treatment or in cases of motor impairment [28].

On the other hand, non-invasive neuroimaging techniques are considered more practical and secure, while providing an acceptable signal quality. This makes these techniques ideal for a wide range of BCI applications. Out of the most popular non-invasive techniques, fMRI has the higher spatial resolution, which allows for the imaging and monitoring of the whole brain, including deep brain structures. However, it has a low temporal resolution, which limits its use in real-time BCI applications. MEG, on the other hand, although it offers a lower spatial resolution than fMRI, has a better temporal resolution. However, the fMRI and MEG recording systems are considerably more expensive than EEG or fNIRS. This, together with the development of wireless and more cost-effective EEG and fNRIS recording systems, makes the use of EEG and fNRIS more appropriate.

Even so, EEG and fNRIS also have their own disadvantages. Although fNIRS offers a high spatial resolution and is less susceptible to electrical noise and artifacts, it has a very low depth of penetration, which is only few millimeter. This limits its suitability for general BCI use. In contrast, EEG offers a high temporal resolution (in the order of milliseconds), making it ideal for these applications. Nevertheless, EEG has a low spatial resolution and a low signal-to-noise ratio (SNR). Still, EEG is the established neuroimaging technique for non-invasive BCI at the moment.

1.2.5 EEG-based BCI Limitations

Regardless of the increased research in BCIs in recent years, several challenges remain for the feasible implementation of these systems outside of laboratory experiments, particularly in uncontrolled environments. Some of the main challenges in BCIs are the subject dependency of the system, the SNR in EEGs, and the complexity of accurately identifying highly similar, or continuous motor tasks [29]. These challenges are further exacerbated in uncontrolled environments through the presence of additional sources of noise and environment variability, such as fluctuating lighting conditions, background stimuli, and the subject's brain natural tendency to multitask, all of which can degrade signal quality and system reliability.

Recent advancements in deep learning have considerably improved the performance of BCIs [30, 31], especially through deep learning architectures such as transformers [32, 33], and autoencoders [34, 35]. However, these methods often require large amounts of data, making them resource-intensive and difficult to scale in practical applications.

To mitigate some of these limitations, transfer learning approaches have been proposed [36], which aim to reduce training data requirements by adapting models across subjects or domains. Even so, all these approaches still have high computational demands, often requiring advanced computational resources and extensive training time, limiting their feasibility in clinical or resource-constrained environments. Furthermore, these methods are predominantly evaluated under controlled laboratory settings with common motor tasks, such as hand or leg movements, which makes comparisons across studies easier, but limits their generalisation to more complex, prolonged, or overlapping motor tasks, often present in more realistic settings [37].

As discussed previously, several methods have been developed for the identification and classification of these neural activities for BCIs, such as RP, VEPs, or ERS/ERD among others. These methods however, have difficulty in precisely and effectively distinguishing motor tasks that have highly similar neural activity, such as fine movements. Although some advances in the decoding of fine movements has been done, [38, 39], especially with the development and improvement

of deep neural networks [40], an accurate classification using low computational cost systems still remains a research challenge for BCI systems.

Existing methods also fail in identification of prolonged or continuous movements, since most of them rely on the prediction and classification of the start of a task, or the spontaneous execution of the task, but not on the prolonged and continuous execution of that. Furthermore, feedback BCI methods require the presence of some audio, visual or haptic feedback for the correct recognition of the mental task, which increases the difficulty of implementing these BCIs in real-world scenarios.

Altogether, these limitations highlight the need for the development of more robust preprocessing, feature extraction and classification strategies that can facilitate the reliable implementation of BCIs in more realistic, uncontrolled, and varying environments with minimal computational resources.

The communications between brain regions are expected to convey important information for the development of effective BCIs that can better mimic natural movements and responses even in the absence of direct feedback. Therefore, for a non-feedback BCI, these communications can be translated into actions, whether they are desired or prescribed movements. In this context, cooperative learning over networks offers a promising solution.

Previous studies [41–46] have demonstrated that cooperative learning, particularly by means of diffusion adaptation (DA) [47], is well-suited for modelling the dynamic propagation of neural information across brain regions. This method can effectively capture the distributed and cooperative nature of brain activity, allowing the system to integrate and leverage the communication flow between brain regions during the execution of cognitive processes. By incorporating these interactions, DA enhances the decoding of motor intentions and cognitive states, leading to improved accuracy and robustness in BCIs.

Although other distributed signal processing strategies, such as graph theory-based signal processing methods [48], have been successfully applied to model neural interactions, these approaches typically require extensive signal preprocessing, and often rely on a fixed graph topology or require global knowledge of the network [49, 50], limiting their adaptability to dynamic and noisy EEG environments, and their suitability for robust, real-world BCIs.

In contrast, DA operates in a fully distributed and adaptive manner, enabling real-time updates with minimal data and computational resources. Additionally, it does not require heavy preprocessing for a considerably good performance. These features make DA particularly well-suited for real-time and real-world BCI implementations.

In this thesis, for the first time, we develop novel approaches for the identification and classification of neural activities through the use of EEG hyperscanning and cooperative learning over networks to improve the BCIs performance while reducing their subject dependency.

1.3 Motivation

As discussed earlier, due to the time-variant psycho-neurophysiological and neuroanatomical factors present in individuals' brain waves, EEGs can suffer short and long-term signal variations within and across subjects. This requires each BCI to be trained for each individual for high performance. This can be tedious and time consuming, especially in situations where the same system might need to be used by several people in a short time frame, such as in game applications or for rehabilitation. These signal variations tend to be aggravated in some patients, such as stroke patients, where the brain rhythms that contain the motor states might be diminished or the brain source localisation of the motor movement might be misplaced, creating further challenges on the ability of the BCI to recognise a motor movement.

On the other hand, most recent BCI studies still tend to rely on the use of brain waves recorded in ideal scenarios where the experiment is undertaken in a controlled environment and the subjects are asked to fully concentrate on a single motor-related task. This also poses some additional challenges in the implementation of these systems in a more realistic scenario, where the brain is inherently engaged in multiple tasks.

To help overcome these challenges, we propose the use of hyperscanning [51]. EEG-based hyperscanning, also known as EEG multi-brain recordings, has great potential in BCI systems used for monitoring mental and physical rehabilitation. The hyperscanning technique involves the recording of brain activity from mul-

tiple brains simultaneously, which allows us to study not only the neural connections of a specific individual, but also how these neural connections change when interacting with other individuals [52]. An example of an EEG hyperscanning recording setting can be seen in Fig. 1.3.

Figure 1.3: A representation of an EEG hyperscanning recording with two subjects involved.

So far, most hyperscanning studies have focused on the study of social interactions between individuals [53], where they analyse how the participants' inter- and intra-brain connections change depending on the tasks they perform. These have provided some insights on how the information in the human brain is processed while in a social setting [54]. All these studies show that interpersonal coordination of action includes mutual synchronisation of neural dynamics, flow of information between brains, and causal effects of one brain upon another. Thanks to this synchronisation of neural dynamics, the information flow between brains can be measured and infused into a cooperative learning network strategy, such as DA [47], where the set of electrodes is considered as the distributed network, for the better recognition of a desired motor task from the synchronised brain waves.

Furthermore, as discussed in [55], given the results obtained after the use of hyperscanning in other social settings, the development of an effective EEG hyperscanning-based BCI for motor neurorehabilitation group therapy could improve the accuracy and speed of the rehabilitation process. This could reduce the training error and time in multi-task scenarios, opening a new and promising

direction for the development of data-efficient and adaptive BCIs in changing environmental and task conditions. However, this has not been explored in depth yet due to the technical challenges involved in the setting of an EEG hyperscanning motor task recording scenario. The two main technical challenges can be summarised as: 1) the difficulties associated with setting a motor rehabilitation scenario for hyperscanning, and 2) the complexity of analysing the EEG hyperscanning signals for motor tasks recognition.

Although the development of new wireless and easy to synchronise EEG recording devices in the last few years has helped with the first challenge, the second challenge still remains, with not much research exploring this possibility. This could be due to the lack of publicly available EEG hyperscanning data.

Therefore, here, it is intended to overcome the previously mentioned challenges in the development of an EEG hyperscanning-based BCI for motor tasks, through the following actions:

- The recording and release of a public dataset containing EEG hyperscanning motor task data to encourage further research on this relatively new research field.
- The development of a novel filtering technique that synchronises the two EEG recordings to best extract the common task and avoid the undesired ones.
- The development of novel approaches for the analysis of motor-related tasks EEG hyperscanning BCI scenarios through the use of a cooperative learning strategy.
- The development of a novel technique for the recognition of prolonged motor movements to improve the system's performance for highly similar motor tasks.

1.4 Thesis Layout and Key Contributions

In this thesis some key new developments are presented in the field of EEG hyperscanning-based BCI through a cooperative learning strategy over distributed

networks. These developments are effective and practical methodologies for several applications, with some discussed throughout this thesis. The key contributions are as follows:

Chapter 2: First, the DA strategy over distributed networks is reviewed. In addition, some application examples to model simulated data are presented. The performance of the system is evaluated and compared for non-cooperative and cooperative strategies.

Chapter 3: An overview of the recording process for an EEG motor task hyperscanning setting is presented and evaluated. The recorded EEG hyperscanning data is described, and its limitations discussed. Explanation on how to access the public dataset is also provided.

Chapter 4: A novel EEG hyperscanning analysis technique is introduced and evaluated. This technique is based on a brain connectivity-informed DA model and can be used as a new rehabilitation platform where the state of the patient under rehabilitation is analysed based on how well the patient is able to follow the tasks performed by the healthy subject.

Chapter 5: A novel approach for the recognition of prolonged motor movement and highly complex fine movements, where the system can learn a movement without requiring any audio, visual or haptic feedback, is presented. A DA strategy is used to model the interface between the brain neural activities and the corresponding gesture dynamics. The method is evaluated using real EEG multi-task data and compared against well-established deep learning models for the recognition of complex fine movements.

Chapter 6: A new filtering technique is proposed by reformulating the well-known common spatial patterns (CSP) method for hyperscanning data during the performance of collaborative tasks. This is highly effective for the development of a BCI in an uncontrolled environment. An overview of CSP is provided and the method is evaluated and the proposed hyperCSP compared against the traditional CSP.

Chapter 7: The thesis is concluded by summarising the results, and discussing possible future work.

1.5 Conclusions

The motivation for this research, together with an overview of the thesis contributions, has been presented in this chapter.

This chapter outlined the need for more robust and adaptive BCI systems, particularly for motor neurorehabilitation in real-world, uncontrolled environments. It also highlighted the potential of group therapy, introducing EEG hyperscanning as a promising approach to enhance neurorehabilitation through the analysis of shared neural dynamics.

The main neuroimaging techniques were reviewed, with EEG identified as the most suitable due to its non-invasiveness, high portability, low cost, and high temporal resolution, despite its inherent signal variability and sensitivity to noise.

The main challenges associated with the development of EEG hyperscanning-based BCIs were also discussed. To overcome these challenges, this thesis proposes the development of novel cooperative analysis methods based on hyperscanning and DA. These methods aim to model and exploit intra- and inter-subject neural communication through distributed cooperative learning strategies to improve BCI reliability and performance.

The following chapter provides a comprehensive review on the DA strategy along with other cooperative learning strategies, evaluating their suitability to the proposed EEG hyperscanning scenario. The DA formulation introduced in Chapter 2 serves as the foundation for the EEG analysis methods presented in Chapters 4 and 5 of this thesis.

Chapter 2

Cooperative Networks

Traditional EEG analysis systems often do not take into consideration the topological relationship among electrodes reflecting the brain network dynamics. As a result, most of these methods present limitations in the effective study of the neural activity patterns generated by the distributed synaptic currents. Such methods fail to capture the complex spatial and network-level interactions between different brain regions. To overcome these limitations, an analysis method based on the cooperative network theory that incorporates diffusion adaptation (DA) [47, 56] is introduced.

2.1 Cooperative Learning

Cooperative networks refer to distributed adaptive systems (or networks) where multiple agents, also referred to as nodes, work together to achieve a common goal by solving an estimation or optimisation problem.

In this system, each agent has access to local information and all the agents collaborate and exchange information with their one-hop neighbours in the network in order to solve the global common target. This allows the system to solve the problem more efficiently compared to a centralised system with a single agent [57]. Fig. 2.1 shows an example of such network, where an agent k is connected to all the agents within its neighbourhood, represented by N_k. Node k is able to communicate with all the connected agents.

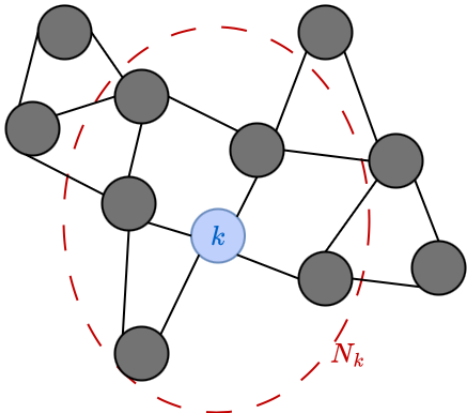

Figure 2.1: An example of a distributed adaptive network, where an agent k is connected to its neighbours within the neighbourhood N_k. Agent k is represented in blue, the other agents are represented in grey, the connected agents are depicted by black lines, and the neighbourhood N_k is depicted by a red dashed line.

Therefore, each node k of a multi-agent distributed network in N_k is interested in estimating an unknown parameter vector $\boldsymbol{w}$. Each node k has access to $d_{k,i}$, a scalar that represents the desired target, and a local measurement $1 \times M$ regression vector $\boldsymbol{u}_{k,i}$ at each time instant i. The parameters of each node are assumed to be related to $\boldsymbol{w}$ through the linear regression model:

$$d_{k,i} = \boldsymbol{u}_{k,i}\boldsymbol{w}^{\boldsymbol{o}} + \gamma_{k,i} \tag{2.1}$$

where $\gamma_{k,i}$ is the measurement noise of node k at time instant i, and $\boldsymbol{w}^{\boldsymbol{o}}$ represents the estimated optimum value of $\boldsymbol{w}$, that can be estimated for each node k by minimising the cost function:

$$J_k(\boldsymbol{w}) = \mathbb{E}|d_{k,i} - \boldsymbol{u}_{k,i}\boldsymbol{w}|^2 \tag{2.2}$$

where $\mathbb{E}$ is the statistical expectation operator. Furthermore, the global cost function of the system can be obtained through the aggregation of the cost functions of each node k, and can be represented as:

$$J(\boldsymbol{w})^{global} = \sum_{k=1}^{N_k} \mathbb{E}|d_{k,i} - \boldsymbol{u}_{k,i}\boldsymbol{w}|^2 \tag{2.3}$$

In these systems, the agents are able to adapt their behaviour based on the information they receive from their neighbours, allowing the network to dynamically adapt to the changes in the environment. The exchange or flow of information between the agents is constrained by the network topology and any restrictions or communication limitations between the agents. Therefore, the learning abilities of the agents directly influence the performance of the system.

2.2 Distributed Learning Strategies

Although different distributed learning strategies have been introduced to minimise the cost function in (2.3), three main strategies have shown a higher suitability in signal processing: incremental [58, 59], consensus [60, 61], and DA [62, 63], with DA showing the best results by better exploiting the network topology and guaranteeing convergence [57, 64]. These strategies are discussed below.

For easier and meaningful comparison between the three strategies as well as the non-cooperation strategy later in this thesis, we define the strategies using the following stochastic gradient algorithm used for centralised implementation as the starting point [59, 63]:

$$\boldsymbol{w}_i = \boldsymbol{w}_{i-1} - \mu \sum_{k=1}^{N_k} \nabla_w J_k(\boldsymbol{w}_{i-1}) \quad i \geq 0 \tag{2.4}$$

where $\nabla_w J_k$ represents the gradient vector of $J_k(\boldsymbol{w}_{i-1})$ with respect to the complex-conjugate transposition of $\boldsymbol{w}_{i-1}$, and μ is the updating time step.

2.2.1 Incremental Strategies

Incremental strategies involve the sequential processing of the information across the nodes, with each node obtaining a local estimation of the desired target and providing this information to the next node in the sequence. The local estimation is obtained based on the local information the node has access to as well as the local estimation of the target obtained and received by the preceding node. Therefore, a cyclic trajectory needs to be decided based on the network topology in order to determine the sequence in which the nodes will be processed, as shown

in Fig. 2.2.

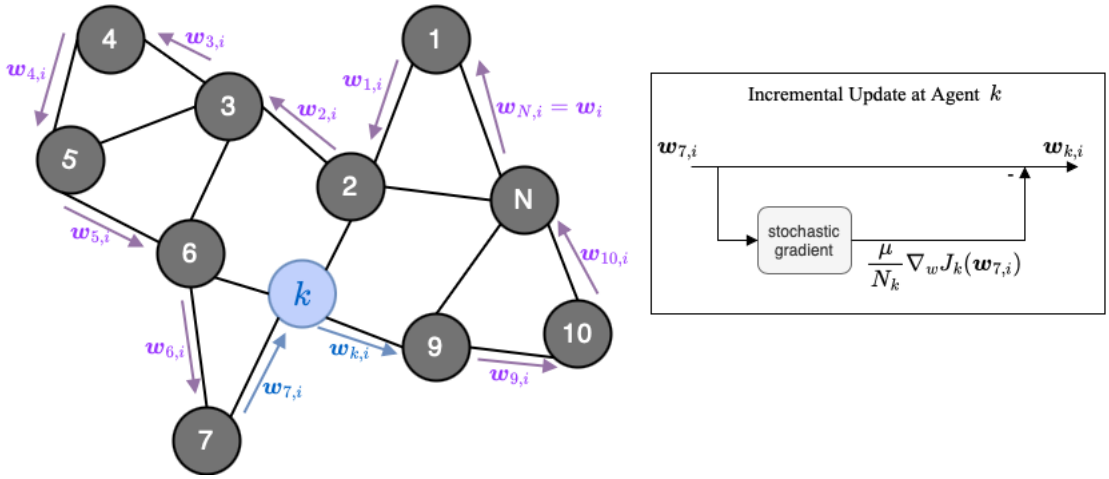

Figure 2.2: Given the network from Fig. 2.1, a cyclic path is defined so all the nodes are covered for the incremental strategy. The nodes are numbered representing the sequential order, from node 1 to node N. Information exchanged received or sent from node k is depicted in blue, while information exchanged between the other nodes is depicted in purple. The diagram at the right illustrates the incremental strategy calculations performed by agent k.

In this strategy, given N nodes in the network, at each time instant i, the centralised implementation (2.4) is divided into N successive incremental steps, with each step performed locally by one of the nodes [57, 59]:

$$\begin{cases} \boldsymbol{w}_{1,i} = \boldsymbol{w}_{i-1} - \dfrac{\mu}{N_k}\nabla_w J_1(\boldsymbol{w}_{i-1}) \\ \boldsymbol{w}_{2,i} = \boldsymbol{w}_{1,i} - \dfrac{\mu}{N_k}\nabla_w J_2(\boldsymbol{w}_{1,i}) \\ \boldsymbol{w}_{3,i} = \boldsymbol{w}_{2,i} - \dfrac{\mu}{N_k}\nabla_w J_3(\boldsymbol{w}_{2,i}) \\ \quad \vdots \quad = \quad \vdots \\ \boldsymbol{w}_{N,i} \equiv \boldsymbol{w}_i = \boldsymbol{w}_{N-1,i} - \dfrac{\mu}{N_k}\nabla_w J_N(\boldsymbol{w}_{N-1,i}) \end{cases} \tag{2.5}$$

Here, N_k represents the network and total number of nodes. The last iteration of the cyclic sequence at time instant i, where we obtain $\boldsymbol{w}_{N,i}$, coincides with $\boldsymbol{w}_i$ as in Eq. (2.4). Therefore, in this implementation the information travels from one agent to the next one until all the sequential nodes are covered during each

time instant i. Then, all the steps in Eq. (2.5) are repeated for each time instant, and it can be summarised as [57, 59]:

$$\boldsymbol{w}_{k,i} = \boldsymbol{w}_{k-1,i} - \frac{\mu}{N_k}\nabla_w J_k(\boldsymbol{w}_{k-1,i}) \tag{2.6}$$

where the first term, $\boldsymbol{w}_{k-1,i}$, on the right-hand side refers to the cooperation while the second term assists with the decentralisation of the system.

Although this implementation offers a distributed solution to the cost function, it presents several limitations and disadvantages compared to other strategies. Due to its sequential dependency, the information from several nodes cannot be processed in parallel, contrary to the consensus or DA strategies, which could create possible bottlenecks on the information flow. At the same time, it limits the scalability and adaptability of the system to similar scenarios with different number of nodes, such as when used for the development of an analysis method for EEGs. In this scenario, once the system is developed for a specific number of nodes, or electrodes, it will be difficult for the system to be adapted to new EEGs recorded with a different number of electrodes, which shows the limit adaptability to changes in the network topology. Furthermore, if one of the nodes of the sequence fails, it will affect all the posterior nodes.

Therefore, due to the incremental strategy characteristics, this implementation is best suited for applications where data arrives sequentially, when resources are limited, or when the data needs to be processed in a step-by-step manner. Even so, in applications related to the processing of physiological signals, this strategy has been used successfully for the sequential or real-time processing of EEG signals, such as in the development of adaptive filters [65].

2.2.2 Consensus Strategies

In a consensus strategy, multiple nodes of a distributed network work together to reach a common estimate of the unknown parameter $\boldsymbol{w}$ that minimises the cost function from (2.3) while only having access to local information, as shown in Fig. 2.3.

In this strategy, the first term on the right-hand side of Eq. (2.6) is replaced by a convex combination of the information available to each node from

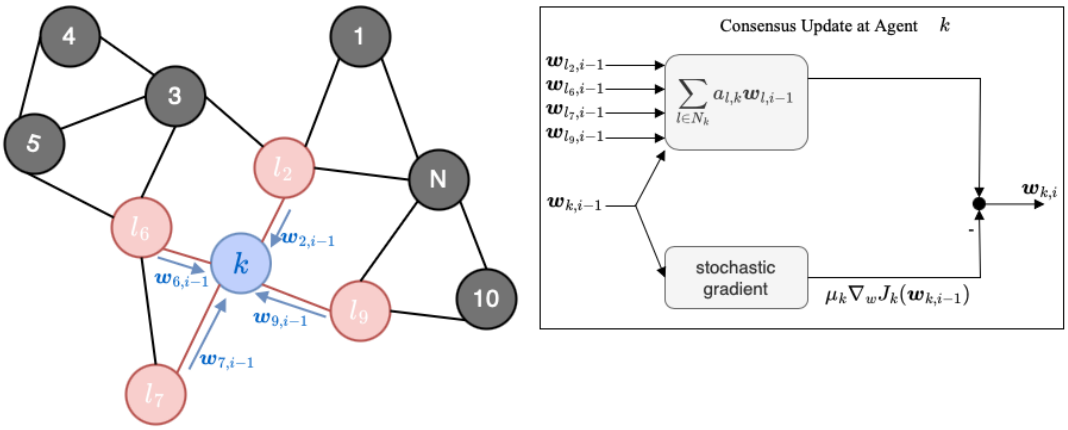

Figure 2.3: An example of a distributed adaptive network, where an agent k is connected to its neighbours. Agent k is represented in blue, the connected nodes to agent k are represented in red, and the other nodes are represented in grey. All the connected agents are depicted by a black line, except for agents connected to agent k, which are depicted by a red line. Information exchanged between node k and its connected nodes is depicted in blue. The diagram at the right illustrates the consensus strategy calculations performed by agent k, which is connected to nodes 2, 6, 7, and 9, depicted as nodes l_2, l_6, l_7, l_9.

its neighbours. On the other hand, the second $\boldsymbol{w}_{k-1,i}$ term is replaced by $\boldsymbol{w}_{k,i-1}$. Therefore, the consensus strategy for each node k is given by [57, 60]:

$$\boldsymbol{w}_{k,i} = \sum_{l\in N_k} a_{l,k}\boldsymbol{w}_{l,i-1} - \mu_k \nabla_w J_k(\boldsymbol{w}_{k,i-1}) \tag{2.7}$$

where μ_k is the agent-dependent step-size, and $a_{l,k}$ are non-negative scalar combination coefficients that satisfies the condition:

$$a_{l,k} \geq 0, \quad \sum_{l=1}^{N_k} a_{l,k} = 1, \quad \text{and} \quad a_{l,k} = 0, \quad \text{if} \quad l \notin N_k \tag{2.8}$$

and represents the weight assigned by node k to the information received from each node l in its neighbourhood.

Therefore, in this implementation each agent k performs two steps at each time instant i. In the first step, each agent aggregates all the estimates of its neighbours. In the second step, the agent updates its information by its local gradient vector.

This implementation has been used successfully in applications where data is distributed across multiple sensors, and centralised processing is either impractical or undesirable [66–68]. It has also shown good results in distributed processing scenarios where the network can afford the communication overhead necessary to reach the common solution. With respect to physiological signal processing applications, this strategy has been used in multiple scenarios, such as for eye blink artifact removal [69].

Even so, this application also presents a number of disadvantages and limitations. As explained before, this strategy presents the term $\boldsymbol{w}_{k,i-1}$ on the second step of its implementation, and a convex combination of the estimates on the first step. It has been shown that this asymmetry between the cooperation and the decentralisation terms can lead to an unstable growth over the network, especially in scenarios that require a continuous adaptation and learning of the system [47, 57, 64]. This instability due to the asymmetry issue can be solved by DA strategies, and are therefore considered the best solution for applications that involve adaptation and learning over networks.

2.2.3 Diffusion Adaptation Strategies

In the DA strategy, self-organised networks are formed of a set of nodes, where each node of the network communicates with its one-hop neighbours sharing their intermediate parameters in order to achieve a common target, as shown in Fig. 2.4.

Although several variations have been formulated to solve the DA strategy for cooperative adaptive distributed networks, two well-established main variations remain the predominant solutions: combine-then-adapt (CTA) and adapt-then-combine (ATC). These two solutions to the strategy can be applied equivalently to the same scenarios.

2.2.3.1 Combine-then-Adapt Diffusion Strategy

For the CTA variation, the same convex combination used in the consensus strategy is used for both, the cooperation and decentralisation terms of the formulation. This way, the two terms are symmetric, overcoming the asymmetry concerns

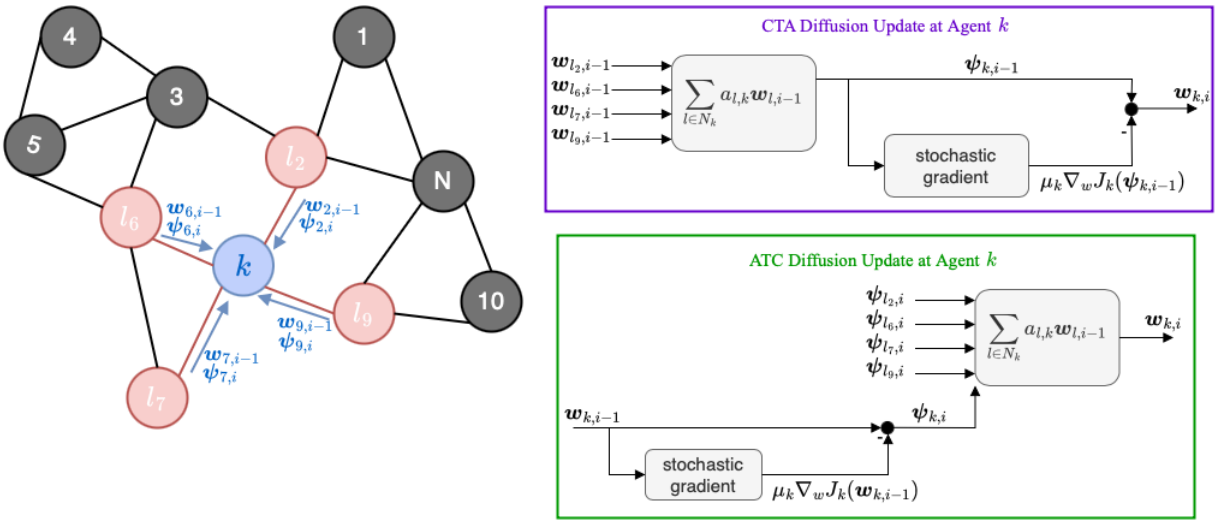

Figure 2.4: An example of a distributed adaptive network, where an agent k is connected to its neighbours. Agent k is represented in blue, the connected nodes to agent k are represented in red while the other nodes are represented in grey. All the connected agents are depicted by a black line, except for agents connected to agent k, which are depicted by a red line. Information exchanged between node k and its connected nodes is depicted in blue. The diagram at the top right illustrates the CTA strategy calculations performed by agent k, which is connected to nodes 2, 6, 7, and 9, depicted as nodes l_2, l_6, l_7, l_9. The ATC strategy is illustrated at the bottom right.

from the consensus strategy. Given all this, CTA can be formulated as [47]:

$$\boldsymbol{w}_{k,i} = \sum_{l\in N_k} a_{l,k}\boldsymbol{w}_{l,i-1} - \mu_k \nabla_w J_k(\sum_{l\in N_k} a_{l,k}\boldsymbol{w}_{l,i-1}) \tag{2.9}$$

where each node k performs two subsequent steps, combination and adaptation. This way, at each time instant i, each node k combines all the estimates $\boldsymbol{w}_{l,i-1}$ from its neighbours l to obtain its own intermediate estimate $\boldsymbol{w}_{k,i-1}$. Then, this estimate is used subsequently to approximate node k gradient vector and obtain the final estimate of node k, $\boldsymbol{w}_{k,i}$.

Therefore, and as shown in Fig. 2.4, this formulation can be divided into two equations, where the first equation, representing the combination, is used to evaluate the convex combination term into an intermediate state variable ($\boldsymbol{\psi}_{k,i-1}$), and then used to obtain the gradient update, as in:

$$\begin{cases} \boldsymbol{\psi}_{k,i-1} = \sum_{l \in N_k} a_{l,k} \boldsymbol{w}_{l,i-1} \\ \boldsymbol{w}_{k,i} = \boldsymbol{\psi}_{k,i-1} - \mu_k \nabla_w J_k(\boldsymbol{\psi}_{k,i-1}) \end{cases} \tag{2.10}$$

As for the consensus strategy, $a_{l,k}$ is the combination weight that represents the information received from each node l by node k and satisfies the same condition as for the consensus strategy, defined in Eq. (2.8).

2.2.3.2 Adapt-then-Combine Diffusion Strategy

As explained previously, the CTA and ATC diffusion strategies can be used comparably in the same scenario since they are fundamentally similar and the only difference between them is the order in which the combination and adaptation steps are performed.

While in CTA we saw that the final estimate $\boldsymbol{w}_{k,i}$ is obtained during the adaptation step, in ATC it is obtained in the combination step, as shown in Fig. 2.4. Therefore, in ATC, at each time instant i, each agent k obtains the intermediate estimate $\boldsymbol{\psi}_{k,i}$ using $\boldsymbol{w}_{k,i-1}$ and the approximate gradient vector. Then, the agent k combines all the intermediate estimates of its neighbours to obtain the updated $\boldsymbol{w}_{k,i}$.

The ATC DA strategy can be written as [47]:

$$\begin{cases} \boldsymbol{\psi}_{k,i} = \boldsymbol{w}_{k,i-1} - \mu_k \nabla_w J_k(\boldsymbol{w}_{k,i-1}) \\ \boldsymbol{w}_{k,i} = \sum_{l \in N_k} a_{l,k} \boldsymbol{\psi}_{l,i} \end{cases} \tag{2.11}$$

Although both variations can be used indistinctly in similar scenarios, ATC tends to be preferred in most cases over CTA [47, 57, 62]. This is due to the order of the adaptation and combination steps. Performing the adaptation step first allows the system to obtain a faster convergence and a better adaptation to local conditions, which can be translated into a more efficient handling of asynchronous updates of the system. This is particularly relevant in scenarios that present a highly dynamic and noisy environment, such as for the modelling of biological networks [70], or for the modelling of social networks [71].

2.3 Single Task and Multitask Distributed Networks

Cooperative learning strategies can be applied to distributed networks where there is only one common objective or target [72–74], known as single task network applications, as well as to scenarios where the network has more than one objective or target [44, 75, 76], known as multitask network applications.

2.3.1 Single Task Applications

In a single task DA strategy, as described in previous sections, each node k of a multi-agent distributed network neighbourhood N_k is interested in estimating an unknown vector of parameters $\boldsymbol{w}$ from the collected local measurements $\{d_{k,i}, \boldsymbol{u}_{k,i}\}$ at each time instant i. We consider the least-mean square error (LMS) problem, where each agent k observes realisations of zero-mean wide-sense jointly stationary data. This is achieved through estimation of the global parameter $\boldsymbol{w}$ that minimises the cost function in (2.3). Using the DA ATC strategy, Eq. (2.11) can be reformulated as [62, 63]:

$$\begin{cases} \boldsymbol{\psi}_{k,i} = \boldsymbol{w}_{k,i-1} - \mu_k \boldsymbol{u}_{k,i}^T [d_{k,i} - \boldsymbol{u}_{k,i} \boldsymbol{w}_{k,i-1}] \\ \boldsymbol{w}_{k,i} = \sum_{l \in N_k} a_{l,k} \boldsymbol{\psi}_{l,i} \end{cases} \tag{2.12}$$

where $\boldsymbol{u}_{k,i}^T$ represents the transpose of $\boldsymbol{u}_{k,i}$.

The adaptation of this formulation has been used in multiple applications [72–74, 77, 78], including for EEG analysis [46, 56]. As an example, in the following sub-section we describe how (2.12) could be reformulated to model and monitor a crowd moving through a varying width environment.

2.3.1.1 Diffusion Adaptation for Crowd Motion Modelling

This sub-section is an expansion of our published work in [79].

The proposed system models crowd movement in a dynamic environment and a distributed manner depending on the information the agents receive from each other and the changes in the environment. This allows for tracking the agents

(people), including people with disabilities, who share a connected network. The environment may include well-defined constraints, such as walls and fences, ticket control barriers, objects, or people moving in unpredictable directions. To perform this analysis, we use the concept of adaptive cooperative networks by means of the diffusion adaptation mechanism [47, 59] to model the crowd motion while passing through geometrically varying areas.

In this work, the agents (or nodes representing people) share their position coordinates and each agent communicates with other agents within its one-hop neighbourhood. As long as an agent can detect the positions of the other nodes (agents) in its neighbourhood, it adjusts its speed and distance with others and the barriers while moving towards its destination (e.g., the exit gate in a metro station). This helps in more accurately calculating and maintaining safe distances between the general public and people with disabilities as well as safe social distances in pandemic situations. To achieve this, the movement speed, the distance between the agents, and possibly their movement directions must change (within allowed limits) with respect to the variations in the pathway geometry (e.g., width) and any obstacle preventing them in reaching their target. This can be achieved simply by being aware of the nodes within a neighbourhood and the geometrical constraints. In this scenario, people can make a compromise between their distances and speeds, to keep themselves safe.

To model a crowd moving through a geometrically varying environment over time, we introduce a mobility model of people, represented as nodes or agents of a connected network, using the previously defined single task DA.

Consider the crowd as a collection of people distributed over a space $\Re^2$ with a defined geometry. The collection of people, with the ability to communicate to each other and share information, forms an adaptive network. They adapt their movement to those of agents in the neighbourhood as well as geometrical/spatial constraints while moving towards their target which, in this example, is at the end of the predefined path (e.g., an exit door). This also helps the crowd members to self-organise themselves based on the information exchanged within their one-hop neighbours. Fig. 2.5 illustrates a group of agents, their neighbourhood, and an exemplar of the surrounding environment.

The general objective of such a network is for each node k to reach the location

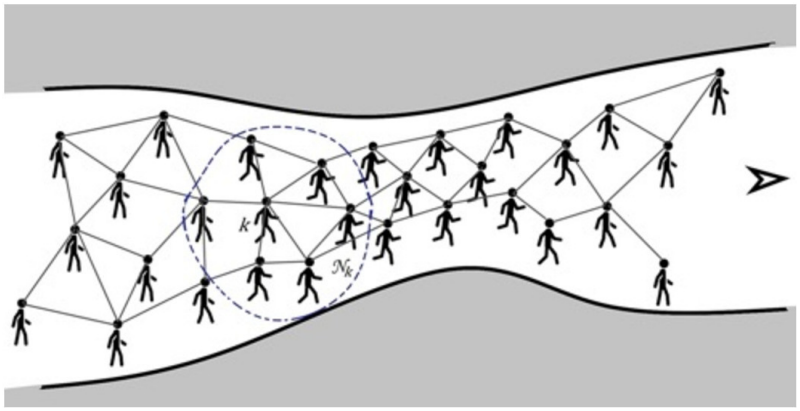

Figure 2.5: The network agents confined by two walls move in a geometrically varying environment whereby their speeds and their proximities can change accordingly. The neighbourhood of agent k, which represents the agent with disabilities in the network, is denoted by N_k represented by the dashed line.

of the target in a fully distributed manner. The model presented here is therefore a variation of the models studied in [73, 78], and [80]. This variation allows the network to move towards a target smoothly through a predefined path while avoiding possible obstacles. Consider a connected network of N nodes where each node k wants to estimate an unknown parameter $\boldsymbol{w}$ from the collected local measurements $\{d_{k,i}, \boldsymbol{u}_{k,i}, \boldsymbol{\tau}_{k,i}\}$ for each node k at time instant i. This forms a global cost function as in Eq. (2.3). In this scenario $\boldsymbol{u}_{k,i}$ represents a unit direction regression vector pointing to the direction of the target. $\boldsymbol{\tau}_{k,i}$ is the location vector of node k relative to a global coordinate system at each time instant i, and, in this scenario, $d_{k,i}$ represents the scalar distance between the location of the target and $\boldsymbol{\tau}_{k,i}$, and is given by the inner product:

$$d_{k,i} = \boldsymbol{u}_{k,i}(\boldsymbol{w} - \boldsymbol{\tau}_{k,i}) \tag{2.13}$$

To solve this optimisation problem, we apply the DA strategy given in Eq. (2.12), where the nodes in node k neighbourhood share their intermediate estimates $\{\boldsymbol{\psi}_{l,i}, d_{k,i}, \boldsymbol{u}_{k,i}\}$ after each iteration. Since the model is geometrically bearing, it can be simplified under reasonable approximations to [81]:

$$\begin{cases} \boldsymbol{\psi}_{k,i} = (1-\mu_k)\boldsymbol{w}_{k,i-1} + \mu_k \boldsymbol{w}_{k,i} \\ \boldsymbol{w}_{k,i} = \sum_{l \in N_k} a_{l,k} \boldsymbol{\psi}_{l,i} \end{cases} \tag{2.14}$$

In this application, similar to mobile networks, in the crowd movement scenario, the relationship between the movement speed and two consecutive agent locations is defined as:

$$\boldsymbol{\tau}_{k,i+1} = \boldsymbol{\tau}_{k,i} + \Delta i \cdot \boldsymbol{v}_{k,i+1} \tag{2.15}$$

where Δi is the time step (time difference between two consecutive states) and $\boldsymbol{v}_{k,i+1}$ is the velocity vector of node k in the next time instant $i+1$. From now on, we focus on estimating the velocity vector and the current position of each agent k at each time instant i, denoted by $\boldsymbol{\tau}_{k,i}$.

In this model, there are two factors that influence the velocity vector of the nodes. The first factor is the spatial constraint involved in identification of the location of node k at each time instant i. In the model, we want the crowd to navigate through a predefined path from a start point to the end point. In Fig. 2.5, we see how the distance between the two surrounding walls of the pathway can change. The moving direction for the crowd (from left to right) is denoted by an arrow. In such a scenario, while the safe social distancing is followed, in the wider areas the people can walk normally and have moderate to large distances between them. Nevertheless, in the narrower regions the subjects should move faster while allowing a minimum, smaller (yet permitted) social distancing to avoid a traffic jam in narrow areas.

The objective of the model is to estimate the position of agent $\boldsymbol{\tau}_{k,i}$, while the agent moves between the two walls and keeps its permissible distance limit from other agents.

The second factor that influences the velocity vector of the nodes is the desire of the agents to move in synchrony and avoid collisions by maintaining a safe distance r between the nodes. As described in [81], this can be achieved by updating the velocity vector as follows:

$$\boldsymbol{v}^b_{k,i+1} = \boldsymbol{v}^g_{k,i} + \gamma \boldsymbol{\delta}_{k,i} \tag{2.16}$$

where γ is a non-negative scalar and $\boldsymbol{\delta}_{k,i}$ is given by:

$$\boldsymbol{\delta}_{k,i} = \frac{1}{|N_k| - 1} \sum_{l \in N_k \setminus \{k\}} \big(\|\boldsymbol{\tau}_{l,i} - \boldsymbol{\tau}_{k,i}\| - r\big)\boldsymbol{u}_{k,i}\big(\boldsymbol{\tau}_{l,i} - \boldsymbol{\tau}_{k,i}\big) \tag{2.17}$$

$\boldsymbol{v}^g_{k,i}$ refers to a local estimate for the velocity of the centre of gravity of the network and is obtained by the ATC diffusion strategy as:

$$\begin{cases} \boldsymbol{\psi}^g_{k,i} = (1 - \mu^v_k)\boldsymbol{v}^g_{k,i-1} + \mu^v_k \boldsymbol{v}_{k,i} \\ \boldsymbol{v}^g_{k,i} = \sum_{l \in N_k} a^v_{l,k} \boldsymbol{\psi}^g_{l,i} \end{cases} \tag{2.18}$$

where μ^v_k is a positive step size, and $a^v_{l,k}$ are the combination weights from the ATC strategy.

To model a more realistic crowd motion Eq. (2.15)–(2.18) are modified so the speed of each node k and the distance between the nodes can be scaled depending on the width of the crowd pathway where node k lands at time instant i and the distances between individual nodes and the target (effectively for the narrow regions).

In order to predict (or estimate) the new speed and location for agent k, we need to re-calculate the above parameters based on the closeness of the two surrounding walls. In this simulation, we assume that the agents move inside a region restricted by two walls, where the dimensions are approximated by the chords of circles tangent to both walls at time instant i. The chord links the two tangent points. The position of node k is considered to be on the corresponding chord (i.e., the chord where k falls on). Fig. 2.6 clearly shows the concept. The agents also maintain a minimum predefined safe distance from the walls.

To enable a more realistic scenario, we assume that the people often go as slow as one step and as fast as three steps per second with approximately 0.6 to 1.2 m strides, respectively. This assumption is essential for setting the initial and the baseline crowd speed. This means that the speed of node k at time instant i ($v_{k,i}$) can vary between $v_{min} = 0.6$ m/s and $v_{max} = 3.6$ m/s. This gives an average speed of $v_{avg} = 2.1$ m/s and can be assumed fixed for all the agents representing the general public. Agents representing the individuals with disabilities are assumed to have half of the normal speed. Given their physical

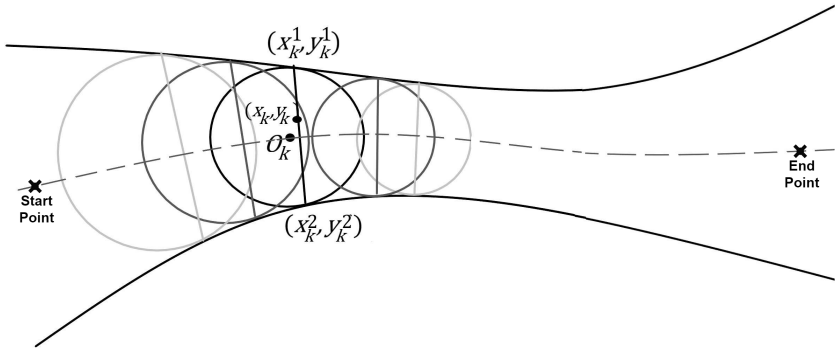

Figure 2.6: Using tangent circles to estimate the varying width of the space in $\Re^2$ at each instant for each agent k between the start and end points of the path with coordinates (x_k, y_k). The circle chord length (between the two tangent points) that contains node k best represents the width of the pathway.

condition, it is reasonable to assume that they always move at a slower speed than individuals without disabilities. The same assumption could be made for toddlers and older people if we were to include them as part of the simulated crowd.

On the other hand, the minimum social distance r can also vary inversely proportional to the speed (or according to the closeness of the walls) between a lowest (e.g., $r_{min} = 1$ m) and a highest (e.g., $r_{max} = 2$ m) value. For people with disabilities, the minimum social distance is higher than for those of the general public given the same speed.

The objective of the new model is to allow the nodes to have a higher speed and a reasonably lower distance between the nodes for the narrower regions (closer walls) and vice versa. Based on this assumption, the effective social distance in the neighbourhood of node k at time instant i can be defined as:

$$r_{k,i} = \frac{v_{max} - \left(v_{k,i} \cdot \frac{\boldsymbol{t}_{k,i} - \boldsymbol{\tau}_{k,i}}{\|\boldsymbol{t}_{k,i} - \boldsymbol{\tau}_{k,i}\|}\right)}{v_{max} - v_{min}} (r_{max} - r_{min}) + r_{min} \tag{2.19}$$

where $\boldsymbol{t}_{k,i}$ in this case represents the target (or end point) location vector at each time instant i.

This shows a linear (but negative) dependency between the social distancing of agent k and its speed at time instant i. Therefore, as long as the speed is

known by the agent, the social distance can be estimated instantly. Individuals with disabilities are expected to have double the social distance compared to individuals without disabilities.

To estimate the speed, we refer to Fig. 2.6. At each time instant i, agent k falls on the chord of a circle linking the two tangent points between the walls and the circle. This is an unique chord for each agent. The agent's speed is inversely proportional to the corresponding chord length at that time.

Therefore, the chord length of node k at time instant i, which is the main parameter for geometrical adaptation, is calculated as:

$$L_{k,i} = \left((y_k^2 - y_k^1)^2 + (x_k^2 - x_k^1)^2\right)^{1/2} \tag{2.20}$$

where, in this scenario, (x_k^1, y_k^1) and (x_k^2, y_k^2) represent the coordinates of node k at time instant i relative to the same global coordinate system. In the chord equation, they represent the tangent points of the two walls and the circle, given that all circles fall on the centre dashed line, as shown in Fig. 2.6. In this equation, we drop the time index i for simplicity.

Finally, we need to utilise the information about the maximum and minimum widths of the pathway, which represent the maximum and minimum chord lengths, respectively, to adjust the agent speed and, accordingly, social distance. Given these two values of L_{min} and L_{max}, and associating them, respectively, to v_{max} and v_{min}, the speed factor $v_{k,i}^c$ can be approximated to:

$$v_{k,i}^c = \frac{L_{max} - L_{k,i}}{L_{max} - L_{min}}(v_{max} - v_{min}) + v_{min} \tag{2.21}$$

In practice, for large crowds, we may assume that there is no chord (pathway width) of less than 1 m in width and no need for any concern about social distancing for pathways of more than 10 m in width (chord length).

The new estimated speed is applied to the overall velocity vector of node k in order to scale and adjust the speed depending on the cross-section of the area where the node is located at time instant i. To ensure that the system works if the crowd interacts with an obstacle, the new velocity vector is set as follows:

$$\boldsymbol{v}^{a}_{k,i+1} = \begin{cases} C_{k,i}\left(\frac{\boldsymbol{t}_{k,i}-\boldsymbol{\tau}_{k,i}}{\|\boldsymbol{t}_{k,i}-\boldsymbol{\tau}_{k,i}\|}\right) & \text{(a)} \\ -C_{k,i}\left(\frac{R}{\|\boldsymbol{\tau}_{k,i}-(\boldsymbol{p}_{k,i}-\alpha)\|}-1\right)\left(\frac{\boldsymbol{\tau}_{k,i}-(\boldsymbol{p}_{k,i}-\alpha)}{\|\boldsymbol{\tau}_{k,i}-(\boldsymbol{p}_{k,i}-\alpha)\|}\right) & \text{(b)} \end{cases} \tag{2.22}$$

where, in this case, R is the radius of the node k neighbourhood, $\boldsymbol{p}_{k,i}$ represents the location vector in the coordinates (x, y) of the obstacle that node k wants to avoid at time instant i, and α is an adjustable coefficient that allows the node to maintain a safe distance from such an obstacle. In this example, $C_{k,i}$ is a coefficient that regulates the speed of each node k at each time instant i depending on the path width.

When node k does not face any static obstacle such as a barrier or walls, $\boldsymbol{v}^{a}_{k,i+1}$ follows Eq. (2.22a) and moves towards the target $\boldsymbol{t}_{k,i}$. However, when node k detects an obstacle close to its location, it avoids the obstacle and moves in the direction opposite to the obstacle, following Eq. (2.22b).

According to (2.14)–(2.22), we propose the following mechanism by which $\boldsymbol{v}_{k,i+1}$ can be set for node k. This mechanism is a modification and extension of the one proposed in [73] and is defined by:

$$\boldsymbol{v}_{k,i+1} = \lambda(\beta \boldsymbol{v}^{a}_{k,i+1}) + (1-\lambda)\boldsymbol{v}^{g}_{k,i} + \gamma \boldsymbol{\delta}_{k,i} \tag{2.23}$$

where, in this scenario, $\{\lambda, \beta, \gamma\}$ are adjustable non-negative weighting coefficients. Algorithm 1 summarises the methodology.

Algorithm 1 Adaptive Cooperative Crowd Modelling using ATC

Require: $\boldsymbol{\tau}_{k,1}$, $\boldsymbol{t}_{k,i}$, $L_{k,i}$

for $i = 1$ to $NumberIterations$ **do**

for $k = 1$ to N_k **do**

Adaptation step:

if $\boldsymbol{\tau}_{k,i} \approx \boldsymbol{p}_{k,i}$ **then**

$\boldsymbol{v}^a_{k,i+1} = -C_{k,i}\big(\frac{R}{\|\boldsymbol{\tau}_{k,i}-(\boldsymbol{p}_{k,i}-\alpha)\|} - 1\big)\big(\frac{\boldsymbol{\tau}_{k,i}-(\boldsymbol{p}_{k,i}-\alpha)}{\|\boldsymbol{\tau}_{k,i}-(\boldsymbol{p}_{k,i}-\alpha)\|}\big)$

else

$\boldsymbol{v}^a_{k,i+1} = C_{k,i}\big(\frac{\boldsymbol{t}_{k,i}-\boldsymbol{\tau}_{k,i}}{\|\boldsymbol{t}_{k,i}-\boldsymbol{\tau}_{k,i}\|}\big)$

end if

$\boldsymbol{\delta}_{k,i} = \frac{1}{|N_k|-1}\sum_{l\in N_k\setminus\{k\}}\big(\|\boldsymbol{\tau}_{l,i}-\boldsymbol{\tau}_{k,i}\| - r\big)\boldsymbol{u}_{k,i}\big(\boldsymbol{\tau}_{l,i}-\boldsymbol{\tau}_{k,i}\big)$

$\boldsymbol{v}_{k,i+1} = \lambda(\beta\boldsymbol{v}^a_{k,i+1}) + (1-\lambda)\boldsymbol{v}^g_{k,i} + \gamma\boldsymbol{\delta}_{k,i}$

$\boldsymbol{\psi}^g_{k,i} = (1-\mu^v_k)\boldsymbol{v}^g_{k,i-1} + \mu^v_k\boldsymbol{v}_{k,i}$

Given $\boldsymbol{v}_{k,i+1}$, we obtain the next location vector of node k

$\boldsymbol{\tau}_{k,i+1} = \boldsymbol{\tau}_{k,i} + \Delta i \cdot \boldsymbol{v}_{k,i+1}$

Combination step:

$\boldsymbol{v}^g_{k,i} = \sum_{l\in N_k} a^v_{l,k}\boldsymbol{\psi}^g_{l,i}$

end for

end for

During each step of adaptation, people with disabilities can be advised on the direction and speed of their movement based on an estimate of the angle between the movement direction and the direction towards the destination. This can be mathematically defined as:

$$\theta_{k,i} = \cos^{-1}\left(\frac{\boldsymbol{\tau}_{k,i}-\boldsymbol{\tau}_{k,i-1}}{\|\boldsymbol{\tau}_{k,i}-\boldsymbol{\tau}_{k,i-1}\|}\cdot\frac{\boldsymbol{t}_{k,i}-\boldsymbol{\tau}_{k,i}}{\|\boldsymbol{t}_{k,i}-\boldsymbol{\tau}_{k,i}\|}\right) \tag{2.24}$$

where '$\cdot$' refers to vector internal product and $\|\cdot\|$ refers to the Euclidean distance between two location vectors. In this example, $\boldsymbol{t}_{k,i}$ is the target (or desired end point) of each node k, $\boldsymbol{\tau}_{k,i-1}$ is the previous position of node k, and $\boldsymbol{\tau}_{k,i}$, represents its current position. $\theta_{k,i}$ can be delivered to the subjects with disabilities (or to their wheelchairs) to correct their directions where necessary.

To show the correct performance of the system, the crowd motion through a predefined path from a start to an end point is simulated. For a better evaluation of the proposed method, we simulate the crowd motion under a highly constrained situation. Therefore, the pathway chosen for the simulation presents a bottleneck situation, such as what the crowd encounters when passing through a narrow corridor, as shown in Fig. 2.7a, an under bridge passage, as in Fig. 2.7b, or a metro station corridor, as in Fig. 2.7c.

Figure 2.7: Illustration of pathways with possible bottleneck represented by the simulated pathway. (a) Narrowing of a passageway, (b) an under bridge passage, and (c) the corridor of a metro station.

In the simulation, we consider the same target for all the nodes over time as the end point such as:

$$\boldsymbol{t}_{k,i} \approx \hat{\boldsymbol{t}} \tag{2.25}$$

where $\hat{\boldsymbol{t}}$, in this scenario, represents the approximate target location, given that the agents keep their social distances evenly close to the target and, therefore, they do not converge exactly to a single target point.

The simulation parameters are set as follows. Consider a crowd of 40 people each representing a node of the network, where one node is a person with disabilities and the others are the general public. The step size, μ_k, and μ_k^v are set to 0.05. For a safe distance from the walls, the coefficient α is set to 0.5. For

velocity control, the coefficients $\{\lambda, \beta, \gamma\}$ are, respectively, equal to $\{0.5, 1, 1\}$. The other parameters are set as defined previously. For a more realistic representation, some random noise was added to the speed of the nodes as well as the distances between each other.

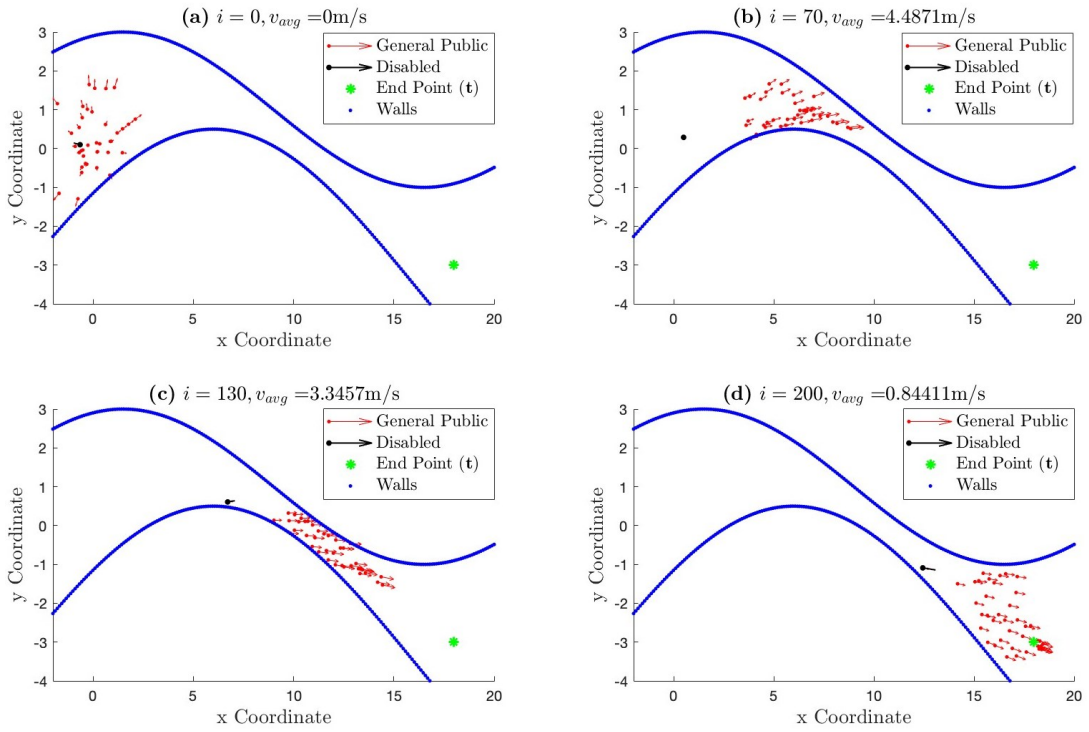

Figure 2.8: Simulation of the movement of a crowd over time. The average speed ($v_{k,i}$) of all the nodes is given for $i = 0$, $i = 70$, $i = 130$, and $i = 200$. The path is represented by blue dots, the general public by red dots, the people with disabilities by a black dot, and the end point of the path by a green "*".

Fig. 2.8 illustrates the movement of the crowd (mobile network) described above in $\Re^2$. The green symbol '*' on the right represents the end point (the target destination), the red dots represent the positions of the nodes considered as the general public, and the black dot represents the position of the people with disabilities over time. Finally, the blue lines define the walls of the path.

At the start of the simulation, in Fig. 2.8a, the nodes are located at some random positions on the left side of the path. This represents the initial locations (start point) of each node showing the stage we start running the algorithm.

These initial locations are generated randomly for simulation purposes. Later, all the nodes move towards their desired destination within the defined path. In Fig. 2.8b, c, the nodes adapt their speeds and distances depending on the width of the path. Finally, in Fig. 2.8d, which shows the end of the simulation time, the nodes gradually approach the desired destination.

In addition, Fig. 2.9 provides the evidence that the nodes have effectively reached or will reach the target over time. This figure represents the Euclidean distances calculated using the equation:

$$E_{k,i} = \left((\boldsymbol{t}_x - \boldsymbol{\tau}_{k,i_x})^2 - (\boldsymbol{t}_y - \boldsymbol{\tau}_{k,i_y})^2 \right)^{1/2} \tag{2.26}$$

The Euclidean distance is calculated for each node k at each time instant i. Fig. 2.9 shows that the average distance between all the nodes and the target decreases over time until it reaches a distance close to 0, which means that the nodes have reached the target. At the end of the simulation, some nodes have already reached their desired destination, while others, including those of the people with disabilities, are still approaching it. As explained previously, the people with disabilities maintain a lower speed, so it will always take them longer to reach their destination than other nodes. This is also in line with the simulation presented in Fig. 2.8, where it can be appreciated how people with disabilities fall behind the general public.

Table 2.1 shows how the chord length changes with respect to the path, and how the speed of the nodes is changed accordingly. The larger or smaller the chord, the lower or higher the speed of the node. The changes to the speed of the people with disabilities can also be appreciated in this table. For a similar chord length, the speed for node $k = 7$, the general public, is considerably higher than the speed for node $k = 5$, the people with disabilities.

These results show that the proposed method obtains a considerable improvement in accurate modelling of the movement of the crowd through a constrained path, compared to the results presented in [78,80], while proving that cooperative communication networks through DA can be used successfully for crowd motion modelling.

This application demonstrated how cooperative networks, and specifically the

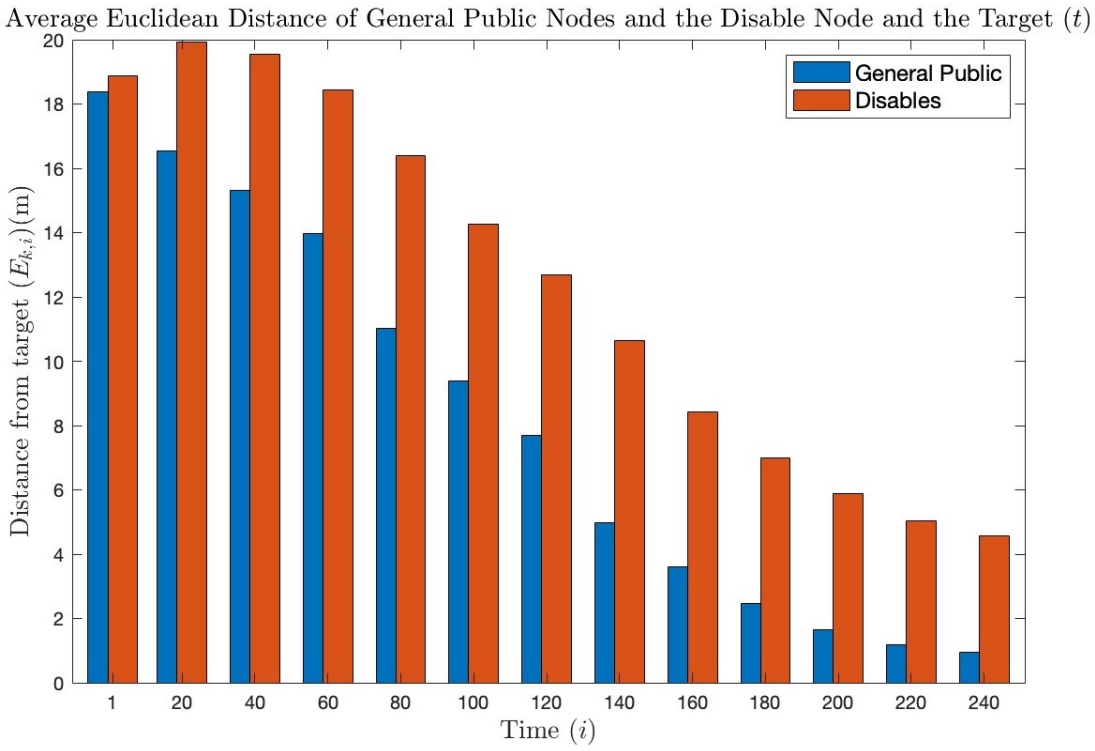

Figure 2.9: Representation of the average Euclidean distance between node k ($\boldsymbol{\tau}_{k,i}$) and the target ($\boldsymbol{t}$) at each time instant i. The blue bars represent the average distance of all the general public nodes, while the red bars represent the people with disabilities.

DA strategy, can effectively model complex real-world interactions, such as a crowd moving towards a common target or exit, or brain regions interacting to perform a cognitive task. It also illustrates the convergence of the DA model, appreciated through the concurrence of all the nodes of the network in the final target location.

As reviewed and proven in [47, 57], the stability and convergence of the DA strategy is strongly influenced by the learning step size, a key algorithmic parameter. Therefore, the selection of this parameter requires careful consideration. Although a large value could accelerate the convergence of the system, it may also lead to the instability of the system and the generation of oscillations in its estimates. On the other hand, a small value may lead to the slow convergence of the network and may fail to track dynamic changes. This value is then chosen

Table 2.1: Length of the chord ($L_{k,i}$) that goes through nodes $k = 7$ and $k = 5$ at several time instants i and speed ($v^c_{k,i}$) of nodes $k = 7$ and $k = 5$, respectively. Node $k = 5$ represents the individual with disabilities.

	$i = 2$	$i = 40$	$i = 80$	$i = 100$	$i = 150$	$i = 200$
$L_{7,i}$ (m)	5.47	3.31	2.36	2.03	3.32	5.75
$v^c_{7,i}$ (m/s)	2.11	2.83	3.15	3.26	2.83	2.02
$L_{5,i}$ (m)	4.57	4.63	3.99	3.68	2.13	2.53
$v^c_{5,i}$ (m/s)	1.2	1.2	1.30	1.35	1.61	1.54

empirically to maintain a balance between the network convergence speed and its stability.

In this example, a fixed step size is assigned to all agents and remains constant throughout the adaptation process. This uniformity not only simplifies the convergence analysis but also ensures consistent learning dynamics across the network.

Similar principles apply to the hyperscanning scenario presented in this thesis, where instead of crowd members, it is necessary to model the dynamic interactions between brain regions engaged in motor or cognitive tasks. This application provides the foundation for the DA-based algorithms developed in Chapters 4 and 5, where the underlying principles of stability, cooperation and parameter sensitivity remain.

2.3.2 Multitask Network Applications

In some scenarios, each group of the network agents might be interested in a separate objective instead of all having one in common. In this situation, the network consists of different clusters of agents with different objectives [75, 76, 82, 83], as shown in Fig. 2.10.

In a multitask DA strategy, we consider a connected network N_k of N nodes, where each node is interested in estimating a set of unknown vectors of parameters $\boldsymbol{w}_{M_q}$ from the collected local measurements $\{d_{k,i}, \boldsymbol{u}_{k,i}\}$ at each time instant i. This can be obtained through the minimisation of the cost function in (2.3). The number of unknown vectors of parameters represent the number of tasks of the

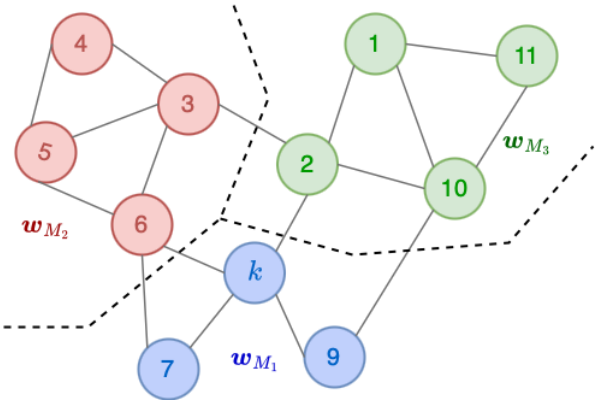

Figure 2.10: An example of a multitask distributed adaptive network, where an agent k is connected to its neighbours. The network consists of $N = 11$ nodes and 3 clusters, also referred to as tasks, where nodes from each cluster are represented in a different colour. The clusters are represented by $\boldsymbol{w}_{M_1}$, $\boldsymbol{w}_{M_2}$, and $\boldsymbol{w}_{M_3}$ respectively.

system, and therefore the number of clusters. In this case, each node k has to estimate its own optimum vector of parameters $\boldsymbol{w}_k^o$, where the optimum vectors of parameters are only constrained to be equal within each cluster. It is assumed that there are similarities and relationships between the neighbouring clusters, which can be represented as:

$$\boldsymbol{w}_k = \boldsymbol{w}_{M_q}, \quad \text{for} \quad k \in M_q \tag{2.27}$$

$$\boldsymbol{w}_{M_p} \sim \boldsymbol{w}_{M_q}, \quad \text{if} \quad M_p, M_q \quad \text{are connected} \tag{2.28}$$

where M_p and M_q denote two cluster indexes, and $\sim$ represents the similarity between the neighbouring clusters. It is assumed that two clusters are connected if there is at least one edge linking a node from one cluster to a node in another cluster.

Therefore, considering a cluster M_k to which node k belongs, Eq. (2.3) can be rewritten as [82]:

$$J_k(\boldsymbol{w}_{M_k}) = \mathbb{E}|d_{k,i} - \boldsymbol{u}_{k,i}\boldsymbol{w}_{M_k}|^2 \tag{2.29}$$

In this formulation, some prior knowledge on the relationship between the

tasks is necessary. These similarities can be defined by the user or established in several ways, such as through mean regularisation [84], low rank regularisation [85], or clustered regularisation [86] among others. In this case, we will focus on the formulation for the multitask network where the similarity between nodes is determined based on the distance between them, obtained through the squared Euclidean distance as:

$$\Delta(\boldsymbol{w}_{M_k}, \boldsymbol{w}_{M_l}) = ||\boldsymbol{w}_{M_k} - \boldsymbol{w}_{M_l}||^2 \tag{2.30}$$

Combining both equations, we obtain a regularised global cost function:

$$J^{global}(\boldsymbol{w}_{M_1}, \cdots, \boldsymbol{w}_{M_q}) = \sum_{k=1}^{N_k} \mathbb{E}|d_{k,i} - \boldsymbol{u}_{k,i}^T \boldsymbol{w}_{M_k}|^2 + \mu_M \sum_{k=1}^{N_k} \sum_{l \in N_k} \rho_{l,k} ||\boldsymbol{w}_{M_k} - \boldsymbol{w}_{M_l}||^2 \tag{2.31}$$

where $\mu_M \geq 0$ is the regularisation strength, and $\rho_{l,k}$ are the regulation factors that satisfy:

$$\sum_{l=1}^{N_k} \rho_{l,k} = 1, \quad \text{and} \quad \rho_{l,k} = 0 \quad \text{if} \quad l \notin N_k \backslash M_k \tag{2.32}$$

and penalise the distance between the vectors of parameters $\boldsymbol{w}_{k,i}$ and $\boldsymbol{w}_{l,i}$.

Similar to the single task global optimisation problem over distributed networks, the cost function in (2.31) can be solved through several strategies. In this case, using the ATC DA strategy defined in Eq. (2.11), Eq. (2.31) can be represented as [82]:

$$\begin{cases} \boldsymbol{\psi}_{k,i+1} = \boldsymbol{w}_{k,i} + \mu_k \Big(\sum_{l \in N_k \cap M_k} \boldsymbol{u}_{l,i} [d_{l,i} - \boldsymbol{u}_{l,i}^T \boldsymbol{w}_{k,i}] + \mu_M \sum_{l \in N_k \backslash M_k} \rho_{l,k} (\boldsymbol{w}_{l,i} - \boldsymbol{w}_{k,i}) \Big) \\ \boldsymbol{w}_{k,i+1} = \sum_{l \in N_k \cap M_k} a_{l,k} \boldsymbol{\psi}_{k,i+1} \end{cases} \tag{2.33}$$

Although this formulation presents certain limitations due to its necessity to some prior knowledge about the relationship between the agents and the tasks, the authors in [83] presented an adaptation of the multitask DA strategy where the nodes do not know which other nodes share similar objectives. This adaptation overcomes the limitations of the example presented here.

The multitask scenario is common in many applications, and the above formulation has been adapted to several applications, such as in web page categorisation [87], web-search ranking [88], disease progression modelling [89], and many others, including EEG analysis applications [44, 45]. As an example, it has been shown that the DA multitask network formulation can be highly useful in BCI applications, where multiple motor tasks need to be classified.

2.4 Conclusions

Despite its major applications in communications and social networks, cooperative networks, as a new and powerful approach, has also been used successfully in the analysis of physiological signals. Specifically, in EEG signal processing, the set of electrodes forms the distributed network (as shown in Fig. 2.11), and each electrode, representing each agent or node, captures a part of the brain's electrical activity. In this scenario, a neuron receives and sends information to other neurons depending on the brain topological connectivity. Therefore, the signals from all the electrodes are processed in a distributed manner through one of the established distributed learning strategies. This method allows for a more accurate and comprehensive analysis of brain dynamics, especially when considering the topological relationship between the electrodes [56].

When choosing the most appropriate distributed learning strategy for this scenario, it is extremely important to take the EEG characteristics into consideration. One aspect to consider is that the EEG could be obtained through a variable number of electrodes. Therefore, the development of an analysis method that can adapt to a varying network size and topology is essential. Due to this and considering the non-stationary nature of these signals, the incremental strategy, which suffers from several limitations including its difficulty to adapt to systems with variable number of nodes, is not appropriate.

As explained previously, these signals present high variability in time. Therefore, it is necessary that the chosen learning strategy can handle the real-time data variability while providing the best results for a system that requires continuous learning and adaptation. This makes the consensus strategy not appropriate due to its asymmetry issues, which can cause instability of the optimisation al-

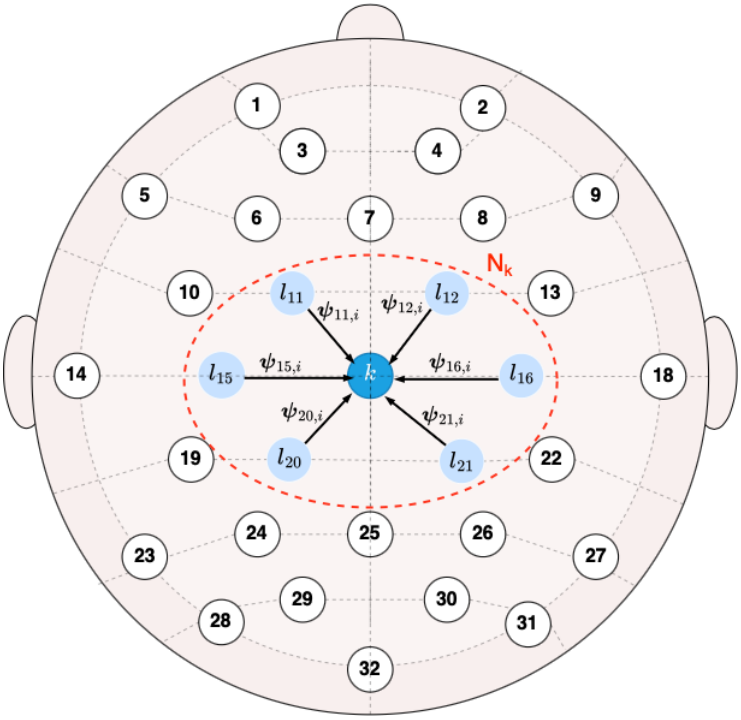

Figure 2.11: An illustration of a set of electrodes which forms the distributed network, where each electrode represents a node or agent. Agent k, which is connected to nodes 11, 12, 15, 16, 20 and 21, depicted as nodes $l_{11}, l_{12}, l_{15}, l_{16}, l_{20}, l_{21}$, is represented in blue. The connected nodes to agent k are represented in light blue while the other nodes are represented in white. Information exchanged through the diffusion adaptation strategy between node k and its connected nodes is depicted in black.

gorithm [64]. Therefore, DA is the most appropriate strategy for the application presented in this thesis.

On the other hand, the suitability of DA for EEG applications has already been proven in several studies, where it also showed the high efficiency of this technique in the study of brain dynamics [41, 46] as well as for the classification of complex EEG signals [43, 44] compared to other learning strategies and non-cooperative models. When applying the DA strategy in this scenario, the nodes share the estimated vector of parameters ($\boldsymbol{w}$) after each iteration of the adaptation process.

Even so, so far DA has not been used in the analysis and study of the brain

dynamics in an hyperscanning scenario. The suitability of DA for the analysis and classification of complex EEG signals shown in previous studies, informed our decision to choose this strategy as the based of the analysis and classification techniques proposed in this thesis for EEG hyperscanning scenarios.

The following chapter provides a detailed description and discussion of the EEG hyperscanning scenarios, with the EEG hyperscanning dataset used to evaluate the proposed DA-based methods described, and its limitations discussed.

Chapter 3

EEG Hyperscanning for BCI; System Setup and Multi-subject Data Recording

Nowadays, a primary factor limiting BCI research is the scarcity and difficulty of accessing the necessary data. This is due to the diverse range of body movements and postures creating considerable ambiguity. There are some publicly available datasets containing physiological data that could be used for BCI development, such as the BCI Competitions datasets [90]. However, most of them contain data from similar BCI paradigms, which limits the scope of the research and advances in the field. Most of these datasets have also been recorded in controlled environments, which hinder the development and testing of techniques that can be applied in daily life.

As previously discussed in Chapter 1, the transition from controlled laboratory settings to uncontrolled environments, such as in real-world neurorehabilitation clinics, introduces additional challenges in the development of accurate and reliable BCIs. In these scenarios, the quality and consistency of EEG recordings can further be degraded due to external factors like ambient noise, variable lighting, background movement, and spontaneous subject behaviour. These factors can introduce additional unrelated neural activity by the subject, either consciously or unconsciously, increasing the presence of artifacts, reducing the SNR, and

increasing the difficulty to extract meaningful neural information, posing additional challenges for signal processing and classification. Moreover, in these settings, subject dependency becomes more pronounced, as inter-subject differences in brain dynamics are further influenced by environmental unpredictability. The high similarity between fine motor movements further adds to these challenges, making it difficult to distinguish between these movements without advanced preprocessing and modelling techniques [29].

The lack of standardisation in recording and analysis of these data prevents researchers from sharing and combining different publicly available datasets into a single BCI paradigm, to better validate their methods under various conditions and develop more versatile applications. Even so, with the development of new, more precise and cheaper wireless EEG recording systems, such as the Emotiv Epoc [91], in the last few years, it has become easier for researchers to record their data using designed paradigms, allowing them to work on a wider range of BCI applications.

The challenges associated with the analysis of single-subject EEG data are intensified in the recording of hyperscanning data. At the time of writing, only two hyperscanning datasets have been made publicly available [92, 93], where none of them contains mental or motor-related tasks that could be of help in the development of new hyperscanning motor rehabilitation BCIs. For this reason, it was decided to record new suitable EEG hyperscanning data.

3.1 EEG Hyperscanning Data Acquisition

In multi-subject settings, the human brain is inherently engaged in multi-tasks at the same time. The tasks often involve both mental and movement activities. Therefore, extraction of EEG activity related to a particular movement in a natural environment involves a great deal of uncertainty and error. The main aim of EEG hyperscanning is to reduce this uncertainty by mathematically restoring the common desired task from the others. Using simultaneous recording of multiple interacting brains, recent studies have been able to examine the brain function underlying social cognition beyond passive observation [94].

Multiple hyperscanning studies have been carried out, most of them focusing

on studying the neural synchronisation and interaction between the subjects in different social settings [52–54], with few studies analysing movement paradigms, such as [95–97]. Some multi-brain studies have also been implemented and analysed in gaming setups, as reviewed in [98]. However, in these implementations, although the brain waves from multiple subjects are recorded simultaneously, the signals from each subject are analysed independently. Therefore, they should not be mistaken as pure EEG hyperscanning-based BCIs, where the signals from multiple subjects are recorded, preprocessed, and jointly analysed.

The findings of these movement related hyperscanning studies, together with the previously mentioned development of more affordable and portable recording systems, have paved the way for a higher interest in using this BCI modality mainly for rehabilitation purposes [55]. These studies have discovered that there is an increased synchrony between subjects during imitation tasks where the gestures are not completely mirrored. This indicates that inter-brain synchronisation, known as the synchronisation between multiple subjects' brains, may not exclusively depend on the precise execution of a particular movement. This is particularly relevant to motor rehabilitation scenarios or paradigms that involve a healthy subject and a subject under rehabilitation. For this reason, in this research we focus on the recording of EEG motor-related hyperscanning data for the first time. This follows the idea that, during interpersonal interaction between two or more individuals, time-varying relationships in brain activation may arise and reveal important findings about inter-brain dynamics.

In an EEG hyperscanning recording session, the data acquisition follows the same procedure used for single-subject EEG recording sessions, but with consideration of the time synchronisation between the subjects.

3.2 EEG Signal Recording

EEG was first introduced by Hans Berger, who recorded human EEG for the first time in 1924, making him the first person to detect and document the brain's electrical activity using a non-invasive neuroimaging technique [99].

Although some other non-invasive neuroimaging technologies, such as fMRI, provide a higher spatial resolution than EEG, they suffer from low time resolution

and other limitations, as discussed in Section 1.2.4. Given all that, over the years EEG has proven to be the most popular method for the development of BCI due to its low cost, good brain coverage, sufficiently high time resolution, and portability.

As discussed in Section 1.2.2, EEG measures the electrical brain activity caused by neural active potentials during the synaptic excitation of the neural dendrites within the cortex. These electrical signals are aggregated and recorded by the scalp EEG electrodes. The position of these electrodes can vary depending on the desired application. However, the majority of EEG recording systems tend to follow a specific standardised positioning, such as the popular 10-20 international electrode placement system [100, 101]. Different types of electrodes might be available, each with their own properties, advantages, and disadvantages. The EEG electrodes may be wet or dry.

Wet electrodes require conductive gel to establish the electrical contact between the scalp and the electrode. This conductive gel reduces the impedance of the electrodes, allowing them to record higher quality signals. Due to the lower impedance, the recorded signals tend to be less noisy, which is highly beneficial in applications requiring high-resolution data. However, the application of the gel to each electrode can be tedious, time consuming and difficult, and often requires trained personnel for the optimal setup of the system, and their posterior clean-up. Furthermore, the gel that is applied to the electrodes can dry out during the recording of long sessions, requiring its re-application.

Dry electrodes, on the other hand, do not need any conductive gel, and rely on conductive materials that are directly in contact with the skin. While this reduces problems related to the use of conductive gel, and leads to a faster system setup and clean-up, it may suffer from lower conductivity between the scalp and the electrodes. This results in lower quality signals with more noise.

Therefore, wet electrodes are preferred in clinical and research applications, where high quality signals are required for accurate diagnosis and detailed brain activity analyses. Dry electrodes, in contrast, are preferred in consumer EEG devices and in-field applications, where an easier and more convenient setup is important and the signal quality is less critical. Even so, new advances and improvements in dry electrodes suggests that in near future, these system are

likely to provide similar signal quality to that of wet electrodes, allowing their implementation in clinics [102].

The most common sampling frequencies for EEG recording systems are 250, 500 and 1000 Hz [103]. The higher the sampling frequency, the higher the signal resolution with more detailed brain activity recorded. On the other hand, following the 10-20 electrode placement system introduced above, we might find systems with a low number of electrodes (typically 21 or 19-32). These systems are commonly used for clinical settings for routine EEG recording. They are easy to use and have low, but sufficient, spatial resolution suitable for monitoring the main brainwaves. We can also find systems with a high number of electrodes (between 128-256 electrodes), also known as high-density systems. Although these systems offer a high spatial resolution and detailed spatial information about brain activity, they are more complex and expensive. For this reason, these systems are usually used in research that requires high-resolution brain mapping, such as cognitive neuroscience studies. Finally, we can also find multiple systems that use a medium number of electrodes (between 32-64), having higher portability than high-density systems. These systems are popular for wireless EEG recording devices and are considered the most suitable for real-world applications like BCIs or field studies.

Depending on the desired application, the most appropriate system setup is chosen. For example, for motor-related tasks, the EEG signals of interest are often recorded from over the sensorimotor cortex, where motor-related potentials and rhythms (mu and beta) are modulated during motor planning and execution. These minimum requirements for the recording system allow us to obtain brain signals that capture detailed enough motor information.

A next step in conditioning the EEG data is preprocessing, which is performed for mitigating noise and artifacts. Often, electrode re-referencing can reduce the existing noise due to the nature of differential amplifiers in EEG systems. Notch or finite impulse response (FIR) stop-band filters are used to remove the national grid frequency. Common average referencing (CAR) is used to reduce noise [103, 104]. Other more advanced techniques can also be used for the online removal of muscle, ocular and heart artifacts, such as regression, blind source separation (BSS), empirical-mode decomposition (EMD) or wavelet transform

techniques, as reviewed in [105, 106].

3.3 Multi-system Setup

The setting for an hyperscanning recording session is considered highly complex since it requires careful experimental design, synchronisation techniques, and robust data aligning methods. As mentioned in [55], although a number of commercial EEG recording systems can be used for the collection of EEG hyperscanning data, well-designed EEG hyperscanning paradigms require precise time synchronisation between all the systems, which is not always easily achievable. Small discrepancies in timing can introduce artifacts or distortions, making it difficult to accurately interpret the neural interactions between the participants. Furthermore, when multiple EEG recording devices are used in the same environment, there is possibility of electromagnetic interference between the devices, which can further worsen the signal quality.

Figure 3.1: A g.Nautilus system with 32 wet electrodes from g.tec medical engineering GmbH.

Taking this into account, the data were recorded using two g.Nautilus systems

(see Fig. 3.1) following the recording standards and recommendations given in the system user manual. These systems were chosen for their high signal quality recording and wireless characteristics. Each recording system has 32 wet electrodes and the data were recorded at 250 Hz sampling frequency. Therefore, in total, 64 channels, 32 channels per subject, were recorded. The electrode locations for each system followed the standard 10-20 international EEG electrode placement, as shown in Fig. 3.2. Each system has a GND electrode located on the forehead of the subject and a reference on the right earlobe.

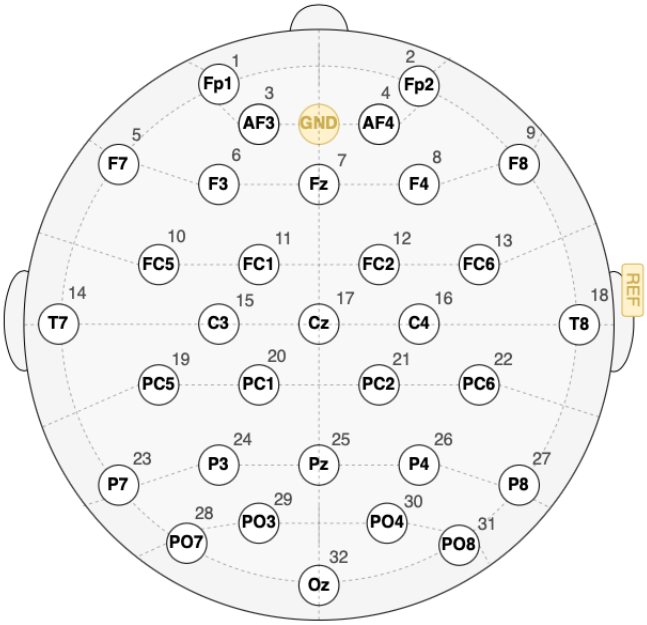

Figure 3.2: The electrode setup used for the experiment. The locations of the electrodes follow the standard 10-20 international EEG electrode placement system for 32 electrodes.

Due to the limited available recording systems and resources, each hyperscanning experiment only consisted of two subjects. To allow for the recording of more realistic data, the recordings were performed in a standard office space, more similar to a standard neurorehabilitation clinic space, where some electrical devices might be present. Following the standard procedure, the office space used for the recordings had a number of electrical devices, which may influence the recording of the signals. Possible distractions such as external noises, stimuli, and other

distractors were reduced from the recording space to increase the concentration of the subjects on the tasks of the experiments. However, the environment is still affected by outside noise, sound clutter, varying window lighting, and likely engagement of the brains in irrelevant mental or physical tasks, which provides an uncontrolled recording situation.

3.3.1 Challenges in Multi-subject Recordings and Data Handling

Several challenges can arise during an hyperscanning recording session, including the experiments' paradigms, the setup and synchronisation of the systems, the availability of subjects, and those associated with the handling of the recorded data.

The current publicly available hyperscanning experiments' paradigms are mostly customised for the study of social interactions between subjects, such as the experiments reviewed in [52], where the researchers are not concerned with synchronised or mirrored motor movements between the subjects. In contrast, for single EEG recording experiments, as discussed in [107], despite several studies, the majority still follow the experiment paradigms presented in [90]. This shows how well-established and evaluated paradigms are usually preferred. Such setups provide quality data for good motor movement recognition results and their outputs are easily comparable to previously developed techniques.

Even so, the paradigm used in our research is unique, has not been used before, and is tailored to our objectives. This incremented the complexity already associated with the design of EEG experiments. To overcome this challenge, the experiments' paradigms were designed following popular BCI motor-related recordings best practices, as in [90], and EEG hyperscanning paradigms proposed in [54]. This paradigm is suitable for the recording and analysis of the EEG for competing or cooperative brain activities. Nevertheless, along the way we made the necessary changes to the paradigm to best suit our objectives.

In EEG hyperscanning, the higher the number of subjects, the more complex it gets to setup and synchronise all the systems. In motor movement paradigms with subjects freely moving around, each subject should have their own indepen-

dent wireless system. This increases the cost of the recording session, and the possibility of introducing delays in the transmission of simultaneous data from such a high number of systems. Furthermore, the synchronisation and connection of all these systems together can be challenging since a high number of commercially available EEG recording systems do not currently support hyperscanning settings or the simultaneous connection of multiple systems to a single PC. A possible solution in that case would be to have each recording system connected to an independent PC and all the PCs synchronised. However, this still further increases the cost of the recording session. The PCs may also introduce artifacts on the recorded signals, degrading the quality of the signals. However, in a perfect EEG-based multi-subject BCI setup and synchronisation, using higher number of participants can boost the training accuracy of the system. This is because the common motor task can be isolated with higher accuracy.

More challenges are related to system preparation, and availability and diversity of the participants. Furthermore, there are complexities associated with the simultaneous recording of multiple subjects whose brain signals are different in nature, such as a group of paralysed and non-paralysed subjects, or a group of young and old individuals. These differences between groups of people could also affect the speed of movement between the subjects, which would need to be mitigated or effectively exploited either during the recording or the preprocessing phase.

Finally, another challenge associated with the recording of EEG hyperscanning data is how to save and handle the recorded data. For a real hyperscanning setting, the data should not only be recorded simultaneously, but it should also be analysed simultaneously, sample by sample for all subjects. The higher the number of subjects involved in the experiment, the more storage space and computational resources that would be necessary, and the more computationally costly the preprocessing and analysis of the data. Furthermore, it is highly important that the synchronisation between the EEG data from all the subjects is maintained. For example, a commonly used technique for the removal of bad data segments or artifacts such as eye blinks, is the removal of the bad segment or bad channel. However, this technique must be applied with caution when working with hyperscanning data. To maintain consistency across all data, if a

bad channel is removed from one subject, often the same channel should also be removed from all the other subjects of the group. Furthermore, to maintain the synchronisation between the subjects, if a bad segment containing an artifact is removed from one subject, it needs to be removed from those of all the other subjects. The higher the number of subjects, the higher the possibility of some subjects presenting multiple bad segments and channels, which could decrease the quality of the data and the results.

3.3.2 Multi-subject Multi-task Dataset Description

Most EEG-based BCI studies tend to rely on the use of large training datasets. This poses some challenges in the later implementation of these systems in an uncontrolled environment and when working with non-ideal and smaller datasets given that the brain is inherently engaged in multiple tasks. Furthermore, as mentioned previously, the data is not readily available and the recording of such data is difficult. Therefore, it is essential to develop a system that can perform well even for small datasets and in uncontrolled environments for an easier implementation of the system in real-world. To support this, we present and release a small but highly valuable dataset containing non-ideal data from several motor-related hyperscanning scenarios.

For the recording of these BCI paradigms, we consider non-ideal data where in addition to the common desired physical tasks between the subjects, there are other possibly strong undesired mental or motor tasks, inherently or intentionally performed by the subjects.

In this experiment, the data were recorded under the ethical approval from the Nottingham Trent University, School of Science and Technology non-invasive Ethical Committee, under the application number 20/21/103. All the volunteers gave their written consents.

The dataset contains the EEG hyperscanning data recorded from a total of five subjects, distributed in three pairs (Subjects 1 and 2, Subjects 3 and 4, and Subjects 1 and 5). For each pair, all the experiments were recorded on the same day. All the subjects were healthy and between 20-30 years old. Out of the five subjects, there were four males and one female. Experiments 1 to 5 contain

between 19 to 29 trials each. Additional experiments that also included other subjects were conducted and analysed, but not used or released with this thesis.

A total of five multi-task experiments were recorded and made publicly available under CC BY-NC 4.0 in [108]. During each one, both subjects performed a common (similar to each other) task and an uncommon (different from each other) task simultaneously. The subjects performed all the experiments with their eyes open. Here, we describe each of the experiments.

During experiments 1 and 2, we recorded two sessions of the experiment for each pair of subjects, and the subjects were asked to sit comfortably on their chairs one next to each other. However, for experiment 3, 4, and 5, only one session of each experiment for each pair of subjects was recorded, and the subjects were standing in front of each other.

Experiment 1: During this experiment, the two subjects were asked to open and close their right or left hands in a random order determined by a sequence of orders given to Subject 1. The left or right movement was also alternated with some randomly determined short free movement time. During the free movement time, the subjects were relaxing and free to move. During the movement, the whole arm was extended and flexed following the opening and closing of the hands. The arm was extending when opening the hand and flexing when closing it. Both subjects were asked to perform the movements at a slow pace and try to synchronise their movements. Subject 1 was leading the movement following the instructions received through a visual stimulus presented on a monitor, while Subject 2 was following Subject 1. Fig. 3.3a shows how the whole arm is moved when the hand is opened or closed.

Experiment 2: Like in the previous experiment, the two subjects were asked to open and close their right and left hands while extending and flexing the corresponding arm in a random order determined by a sequence of orders given to Subject 1. During this experiment, both subjects were also asked to simultaneously move another part of their body in a random order decided by themselves. Here, for the secondary movement, Subject 1 moves their right leg back and forth, while Subject 2 taps their feet. Fig. 3.3b shows the secondary movements performed by both subjects.

Experiment 3: During this experiment, both subjects were asked to open

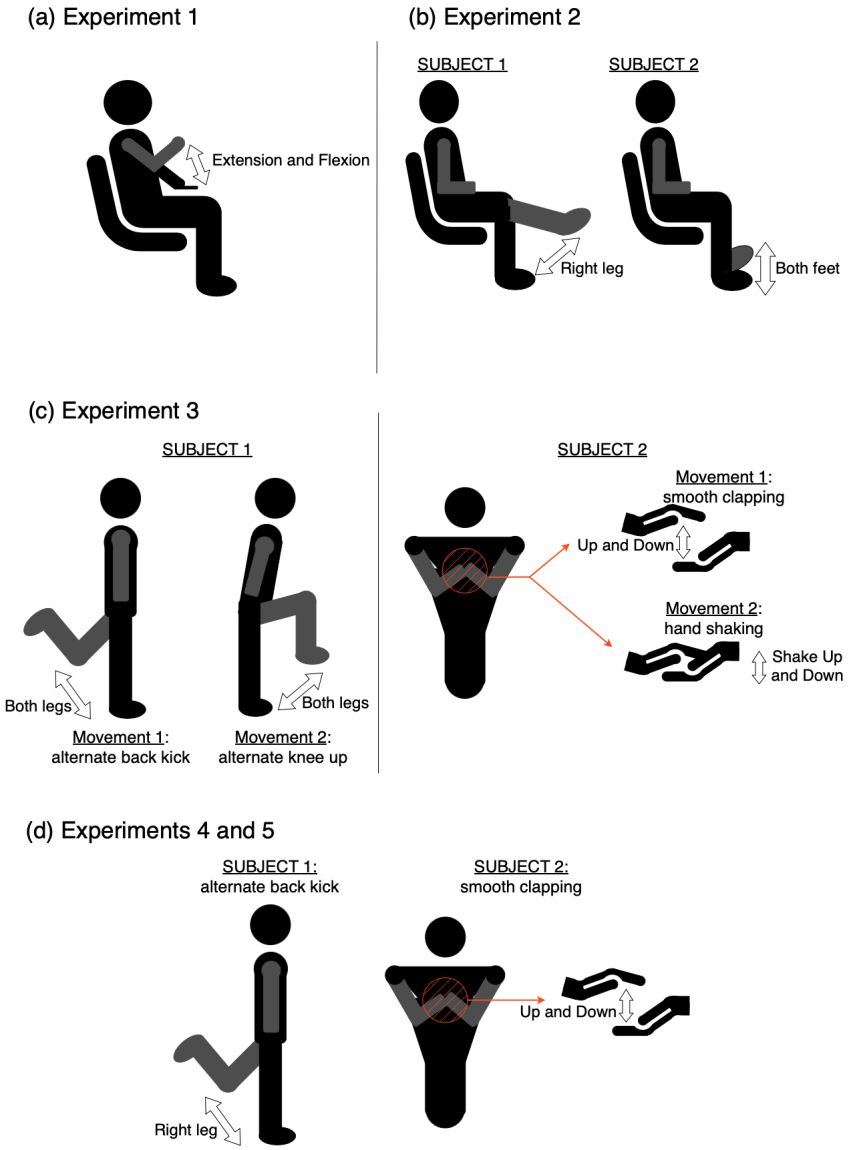

Figure 3.3: The schematic of movement performed by the subjects during all the experiments, (a) open and close hand movement performed while extending and flexing the corresponding arm for experiment 1 and 2, (b) secondary movements performed by subjects 1 and 2 during experiment 2, (c) two secondary movements performed by each subject during experiment 3, and (d) secondary movement performed by each subject during experiments 4 and 5.

and close their both hands simultaneously following a slow motion. During this experiment, as opposed to experiments 1 and 2, only the hands were moving.

Both subjects were asked to try to synchronise their opening and closing hand movements. It was observed that the subjects were able to synchronise their movements most of the time. In a random order, Subject 1 was given three sets of instructions: to freely move their hands, to open and close their hands and to perform a secondary movement, while Subject 2 follows Subject 1 for the prescribed movement. During the freely-moving time, both subjects were allowed to relax their hands and move them freely, as in previous experiments. During the secondary movement time, both subjects performed two movements each in a randomly self-decided order. Here, Subject 1 alternatively performed back kicks with each of their legs or alternatively moved one of their knees up. Subject 2 either performed a smooth hand shake between both hands or smoothly clapped. Subject 2 performed continuous smooth movement of their arms and hands. Fig. 3.3c shows the secondary movements performed by both subjects. In these experiments, the main objective is to isolate the common task performed by both participants.

Experiment 4: This experiment was a repetition of experiment 3, with a single change. Here, when the subjects were asked to perform the secondary movement, both performed a separate single movement instead of choosing between two movements. Subject 1 performed back kicks with their right leg, while Subject 2 simulated a smooth continuous hand clapping. Fig. 3.3d shows the movement performed by both subjects during the secondary movement time.

Experiment 5: This experiment was a repetition of experiment 4, with a single change. During this experiment, instead of only Subject 1, both subjects were given the instructions as to when to move freely, perform the secondary movement, or open and close their hands. They were still asked to try to synchronise their movements.

For the evaluation of some of the developed techniques, another experiment was carried out. This experiment was performed by the previous pair and a new pair of subjects, consisted of one subject in common with previous experiments and another new unrelated subject. The data from this experiment was not made publicly available. Two sessions per pair of subjects were recorded for this experiment, with several trials recorded as part of this experiment per session.

Experiment 6: In this experiment, two right-handed subjects were asked to

Table 3.1: Summary of EEG hyperscanning dataset. Left and right are abbreviated as L and R, and S1 and S2 represents Subjects 1 and 2 respectively.

Experiment	Subjects	Trials	Primary Task	Secondary Task(s)	Posture	Instructions
1	4 (2 pairs)	29	Open/close L/R hand with full arm movement	None	Sitting	Visual stimulus to Subject 1
2	4 (2 pairs)	29	Same as Experiment 1	S1: right leg movement S2: feet tapping	Sitting	Visual stimulus to Subject 1
3	4 (2 pairs)	23	Open/close both hands (no arm movement)	S1: back kicks or knee raises S2: hand shake or smooth clapping	Standing	Visual stimulus to Subject 1
4	4 (2 pairs)	19	Same as Experiment 3	S1: back kicks S2: smooth clapping	Standing	Visual stimulus to Subject 1
5	4 (2 pairs)	22	Same as Experiment 4	Same as Experiment 4	Standing	Verbal instructions given to both subjects
6	6 (3 pair)	12	Draw clockwise and anti-clockwise circles	None	Sitting	Verbal instructions given to Subject 2 only

perform a cooperative task, with one subject acting as the leading subject (in this case Subject 2) and the other subject acting as the follower subject (in this case Subject 1). Both subjects had their eyes open during the experiment. Both subjects were asked to draw clockwise and anti-clockwise circles following a slow and continuous movement. The leading subject was provided instructions on how many circles to draw in each direction, following a set sequence, before the start of the recording. The follower subject was asked to draw a circle following the sequence of directions from the leading subject. The experiment ran for approximately 10 minutes.

The sequence of directions given to the leading subject was:

1. Draw a circle clockwise for three times.
2. Rest for 30s.
3. Draw a circle anti-clockwise for three times.
4. Rest for 30s.
5. Repeat steps (1)-(4).

Table 3.1 provides a summary of the key characteristics of each experiment presented here, including the number of trials per session, and a brief task de-

scription. Although other recording sessions and experiments were obtained, they were discarded due to their low signal quality or lack of synchronisation between the data received by both systems.

3.3.3 EEG Hyperscanning Dataset Limitations

As most datasets, the dataset described in this chapter presents certain limitations. It consists on recordings from only five subjects. Although this is consistent with other EEG hyperscanning studies, such as [109, 110], the relatively small number of participants could impact the dataset's ability to robustly assess inter-subject variability and generalisability in BCI performance.

To mitigate this limitation, careful consideration was given to the selection of the subjects, emphasising their diverse demographic, gender, and behavioural characteristics.

On the other hand, although the number of trials per experiment may be considered modest, especially in comparison to large-scale EEG datasets, the number of trials is evenly distributed across the different experimental conditions. This balance helps mitigate possible biases during model training and evaluation, while supporting a fair comparison across different task conditions. Moreover, the sample size is also similar to other small size EEG motor-related datasets, such as the datasets reviewed in [111].

Nevertheless, taking the sample size into consideration, cross-validation has been adopted as part of the evaluation strategy in later chapters. This approach facilitates the optimal use of the available data while providing a robust and reliable estimate of the model performance.

Even so, despite its limitations, the recorded data contains high-quality motor-related EEG hyperscanning recordings, making it highly valuable for the exploration and development of EEG hyperscanning-based BCI approaches.

3.4 Conclusions

EEG data recording for multi-subject BCI (a.k.a. hyperscanning) has been attempted in this chapter. Although the number of subjects, and data size are

limited, the presented dataset remains suitable due to the diversity of subjects, and the high quality signals.

Increasing the number of participants under ideal and non-ideal conditions, as well as the data size and the experiments can enhance the accuracy of the BCI training system and the corresponding results. However, it increases the complexity in the setup and requires the availability of diverse participants for longer periods of time, and well-tuned and synchronised recording systems, which can be challenging and not always possible.

Due to the necessary parallel preprocessing and analysis of the hyperscanning data, many common and well-established EEG techniques cannot be used. Therefore, in addition to the basic signal preprocessing techniques, the development of new techniques especially tailored for hyperscanning settings, such as new pattern recognition methods, is essential. However, other related popular preprocessing methods, such as CAR or independent component analysis (ICA) [112, 113], can be used after minor modifications or when applied to each subject's data independently as long as the synchronisation of the data is maintained.

Despite various problems, limitations, and challenges, a useful dataset from two pairs of participants have been recorded, conditioned, and preprocessed to be used by the algorithms proposed in the later chapters of this thesis. In the next chapter, a novel hyperscanning analysis technique is presented and evaluated with the help of the dataset described in this chapter. The proposed method is based on the DA strategy previously introduced in Chapter 2, and can be used as a new rehabilitation platform where the state of a patient under rehabilitation is assessed based on their ability to follow the tasks performed by a healthy subject.

Chapter 4

Diffusion Adaptation for Hyperscanning BCI

This chapter is an expansion of our published work in [114].

As discussed in Chapters 1 and 3, the hyperscanning technique allows researchers to study not only the neural connections of a single person, but also how the transfer of neural information change when individuals interact with each other [52, 115]. Although multiple hyperscanning studies have been carried out so far, they were mostly related to studying social interactions between individuals [53]. To study these interactions, the researchers focused on studying the connectivity between different regions of the brain, within a subject's brain regions (referred to as intra-brain connectivity), or between different subject's brain regions (referred to as inter-brain connectivity). Statistical approaches are used to study the causal relationship for the above two cases. However, the use of such information within the hyperscanning data has not been attempted in BCI context so far.

Depending on the type of activities and interactions the subjects are involved in an hyperscanning setting, we can separate these mostly social studies in two main groups: collaborative and competitive [54]. In cooperative settings, such as the scenarios discussed and presented in Chapter 3, the subjects perform at least one task in common. On the other hand, in a competition hyperscanning setting, the subjects perform tasks against each other, such as playing a card

game [110]. Due to such brain interactions, there are new added information exchange among the network nodes and relationships between the subjects' brain regions as compared with the traditional, non-hyperscanning setting.

As discussed in Chapter 1 and [55], more advanced preprocessing, analysis and classification techniques specifically tailored for hyperscanning data become necessary. These techniques are mainly to exploit the intra- and inter-brain communications in BCI systems used for neurorehabilitation.

More specifically, we propose a brain connectivity-informed diffusion adaptation (DA) method. The proposed method is applied to the EEGs from two brains recorded while performing collaborative tasks. In such scenario, at least two subjects perform a collaborative task together, where one of the subjects leads the task while the other subject follows. The state of the patient can then be identified depending on his ability to follow the task performed by the leading subject (usually a healthy subject).

4.1 Brain as a Connected Network

As discussed in Chapter 2, the human brain can be considered as a connected network, composed of billions of neurons organised into regions that are structurally and functionally interconnected. These regions communicate through complex pathways that enable the brain to process information, make decisions, and control actions. Using scalp EEG, the nodes can represent the brain zones under the electrodes.

Several studies have demonstrated the necessity to study not only the functionality of each part of the brain, but also the connections and relationships between the brain regions, how they change over time, and how they relate to different mental or motor tasks. Furthermore, some studies have demonstrated that the changes in brain connectivity help identify different neural disorders [116–120]. Several methods have been proposed to estimate the causal relationships of brain zones, namely casual and functional brain connectivities.

As discussed in Section 2.4, cooperative learning strategies have been used previously for analysis of physiological signals, where the brain is represented as a connected network, as shown in Fig. 2.11. In such a network, each electrode

is considered as a node or agent of the network. In these studies, cooperative learning strategies, especially DA, have proven to be an excellent resource to model the pathway between the brain network dynamics and the desired motor task [42, 44, 45]. This method can be even more useful when a subject sets the target and other subjects try to mirror, follow, or achieve that. This is the scenario proposed in this chapter.

Therefore, we establish a cooperative adaptive model informed by the brain connectivity information from one brain to follow the other.

4.1.1 Brain Connectivity Estimation

Several methods can be used for the study of brain connectivity and, depending on the purpose and scope of the analysis, they can be divided into three main groups: structural, functional and effective connectivities [121]. Structural connectivity is anatomical and time independent. On the other hand, functional connectivity reflects the statistical relationships, such as temporal correlation or synchrony, between neural activities in different regions over time. Finally, effective connectivity focuses on the causal influence that one brain region has over another. Therefore, when estimating the connectivity strength and direction between different regions, we focus on functional or effective connectivity study methods.

Although effective connectivity methods provide the direction of the information flow between brain regions, as discussed in [121] and [122], functional connectivity methods are more widely used in these studies.

There are different methods for estimating the functional connectivity of EEG signals, such as the methods reviewed in [122, 123]. These methods may further be divided into two groups: non-parametric and parametric. Methods based on the multivariate autoregressive (MVAR) model are non-parametric as the data distribution is not involved in the estimation. The most commonly used methods to measure brain connectivity in EEGs are cross-correlation, when measuring connectivity in time domain [121]; phase-locking value (PLV) [124] or phase-lag index (PLI) [125] for phase-based measures; and partial directed coherence (PDC) [126], directed transfer function (DTF) [127] or the multivariate autoregressive

extension of Granger causality (GC) [128] for multivariate time series measures. GC was initially introduced based on linear autoregressive (AR) models of time series [129], and was later extended to the multivariate formulation for its more effective use for the analysis of multichannel data such as EEG.

Although most of these methods can be used for brain connectivity analysis in hyperscanning settings, some of them are more appropriate than others for the scenario presented in this chapter. In hyperscanning settings, it is common to have a time lag between the two brains. These time lags as the result of desynchronisation between the brains' neural activities can be measured and analysed to provide some important insights into the brain dynamics of a multi-subject cooperative task. However, in a scenario such as the one presented in this chapter, where a subject is mirroring or following the task of another subject, the time lag between the leading brain and the following brain can be considerable. Therefore, in this scenario, other measures for intra- or inter-brain connectivity, such as GC, DTF or PDC, would be sufficiently suitable.

Given all this, in this chapter we employ the multivariate extension of GC, which is one of the commonly used hyperscanning connectivity estimates. This method also shows if there is some level of cooperation, presented as similarity between the two EEG patterns.

The traditional GC studies the existence of a causal influence to a driven process from a driving process such as from $x_{1,i}$ to $x_{2,i}$ in the time domain. This is known as a pairwise causal analysis. Therefore, the traditional form of GC may better estimate the relationship or statistical dependency between any two time series in more detail compared to other methods by exploiting their true probabilistic models. Furthermore, the multivariate linear GC analysis allows us to assess the direct influence between pairs of signals.

4.2 Modelling Brain Motor Activity

In this section, we propose a method to model the brain responses of an individual based on the brain responses of a second individual under a cooperative EEG-based hyperscanning scenario. We establish a cooperative adaptive model between two brains informed by the connectivity information from the EEG sig-

nals.

Cooperative distributed learning has been used previously for EEG processing and task classification [45, 130], as a filtering approach for modelling brain responses to motor tasks [42], and in BCI multi-tasks scenarios [44]. However, it has never been used to model the connection between two brains.

4.2.1 Different Possible Models

Considering the brain as a multi-agent distributed network of connected nodes, the brain of each subject can be represented as a connected network, where each electrode represents a node or agent of that network, as shown in Fig. 2.11. Therefore, the distributed networks consist of N_k nodes each. In these networks, the nodes try to estimate the unknown vector ($\boldsymbol{w}$) given the shared information $\{d_{k_i}, \boldsymbol{u}_{k_i}\}$ that minimises the cost function in (2.2). As discussed in Section 2.4, although there are multiple cooperative learning strategies for solving this optimisation problem, the most appropriate and chosen learning strategy for this scenario are the DA strategies.

In the application presented in this chapter, only the set of electrodes from one of the two individuals, the subject whose brain responses will be modelled, are considered as a distributed network. The corresponding electrode signal from the second individual is instead considered as the global target of the first individual's distributed network. This scenario could also be considered as a multitask distributed network application, where each electrode from the first individual has a corresponding electrode from the second individual as target of the network [44].

In a DA network the agents, or nodes, cooperate with each other to solve a global optimisation problem over the network, as discussed in detail in Chapter 2. In classic DA, the objective for all the nodes is the same. However, in the scenario proposed in this chapter, each node, representing each EEG channel, tries to estimate its own $M \times 1$ unknown vector ($\boldsymbol{w}_k$) given the shared information that minimises the cost function in (2.2). This is known as a multitask DA strategy. However, this scenario could also be considered as a single task application instead, where the set of electrodes from the second individual is considered as a single multidimensional target, as proposed in [41]. Most importantly, despite

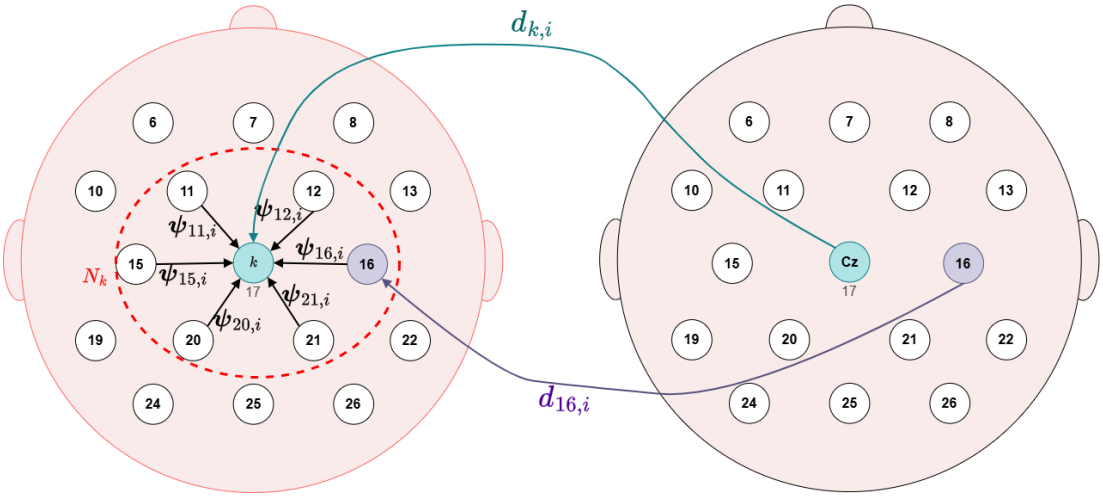

Figure 4.1: An illustration of the cooperation in the proposed hyperscanning setup. The nodes from the left head cooperate between each other while trying to follow the signal from their equivalent electrode from the right head. The right head represents node g from the individual leading the activity while the head at the left represents node k from the individual following the activity. N_k represents the neighbourhood of node k when $k = 17$.

various brain activations, here there is only one physical task to be considered and studied. In this case, the desired signal target ($d_{k,i}$) may be represented by the average of the signals that correspond to the desired target class (the leading subject). Another option is to have separate distributed networks each with an independent objective, such that the objective or target is the signal from the corresponding electrode of the leading subject. Therefore, we model the relationship between the signal of each node k of one individual based on the signal of the corresponding node g of another individual, as in Fig. 4.1.

For simplicity and to reduce the computational cost of the proposed system, we consider the proposed scenario as a single task DA problem. Therefore, to solve the optimisation problem in Eq. (2.2), we use the ATC DA strategy defined in Eq. (2.12).

In the ATC formulation the combination weights $a_{l,k}$ represent the weight of information received by node k from its neighbours l. It is represented as a $N_k \times N_k$ combination matrix of non-negative values at time instant i. N_k is the neighbourhood size. In single task DA applications where the weights are not

measurable, it is presumed that these combination weights are the same and equal to $a_{l,k} = 1/N_k$ for all the nodes l in the neighbourhood of each node k. However, in the scenario presented here, the $a_{l,k}$ coefficients are estimated through brain connectivity measurement to provide a more realistic and accurate information exchange between the nodes, as proposed in [41] and [42] for traditional, non-hyperscanning BCI settings.

4.2.2 Incorporating Connectivity Measures into the Diffusion Adaptation Formulation

Here, the intra-brain connectivity measures from the first individual are used as the combination weights for the DA formulation.

As discussed in Section 4.1.1, although different methods have been established to study the information flow between a group of EEG signals, some methods are more appropriate than others to be used in the scenario proposed here.

For the model proposed in this chapter, we calculate the combination weights $a_{l,k}$ based on the multivariate GC analysis in the frequency domain.

Consider $\boldsymbol{X}_{k,i} = [x_{k_1}(i), x_{k_2}(i), \quad \dots \quad x_{k_{N_k}}(i)]$ a signal that satisfies a MVAR of the form [131]:

$$\boldsymbol{X}_{k,i} = \sum_{m=1}^{p} \boldsymbol{A}_{kk,m} \boldsymbol{X}_{k,i-m} + \sum_{\substack{l=1 \\ l \neq k}}^{N_k} \sum_{m=1}^{p} \boldsymbol{A}_{kl,m} \boldsymbol{X}_{l,i-m} + \boldsymbol{\varepsilon}_{k,i} \tag{4.1}$$

where N_k is the total number of EEG channels that forms the regression model, p denotes the MVAR model order, $\boldsymbol{\varepsilon}_{k,i}$ is the regression residuals, $\boldsymbol{A}_{kk,m}$ are the regression coefficients, and $\boldsymbol{A}_{kl,m}$ are the coefficients that encapsulate the dependence of $\boldsymbol{X}_k$ on the past sample m of $\boldsymbol{X}_l$, given its own past. Eq. (4.1) can be summarised to:

$$\boldsymbol{X}_{k,i} = \sum_{m=1}^{p} \boldsymbol{A}'_{kk,m} \boldsymbol{X}_{k,i-m} + \boldsymbol{\varepsilon}'_{k,i} \tag{4.2}$$

where $\boldsymbol{\varepsilon}'_{k,i}$ are the cross-term residuals, and $\boldsymbol{A}'_{kk,m}$ are the reduced regression coefficients.

We transform Eq. (4.2) to the frequency domain and obtain the transfer matrix $\boldsymbol{H}(f)$, which contains all the information about the relations between the nodes (k, l):

$$\mathbf{H}(f) = \left[\boldsymbol{I}_{N_k} - \sum_{m=1}^{p} \boldsymbol{A}'_{kk,m} e^{-j2\pi f m}\right]^{-1} \tag{4.3}$$

where $\boldsymbol{I}_{N_k}$ is the $N_k \times N_k$ identity matrix, and $e^{-j2\pi f m}$ is the Fourier transform kernel, where $j = \sqrt{-1}$.

Given the transfer matrix and the reduced MVAR model, the cross-spectral density matrix is found as [132, 133]:

$$\mathbf{S}(f) = \mathbf{H}(f)\boldsymbol{\Sigma}'_{x_k x_l}\mathbf{H}^*(f) \tag{4.4}$$

where each element $\mathbf{S}_{kl}(f)$ of $\mathbf{S}(f)$ represents the spectral relationship between signals $\boldsymbol{x}_k$ and $\boldsymbol{x}_l$, $\mathbf{H}^*(f)$ represents the Hermitian transpose of $\mathbf{H}(f)$, and $\boldsymbol{\Sigma}'_{x_k x_l}$ is the residual covariance matrix of the reduced MVAR model such as $\boldsymbol{\Sigma}'_{x_k x_l} = \text{cov}(\boldsymbol{\varepsilon}'_{k,i})$.

Given this, we can assess the direct influence of node x_l on node x_k, also referred as the GC index, as:

$$\mathbf{I}_{l\to k}(f) = \frac{|\mathbf{S}_{kk}(f)|}{|\mathbf{S}_{kk}(f) - \mathbf{H}_{kl}(f)\boldsymbol{\Sigma}'_{x_l x_l}\mathbf{H}_{kl}(f)^*|} \tag{4.5}$$

given $\boldsymbol{\Sigma}'_{x_k x_l} \equiv 0$ for $l \neq k$.

Furthermore, given Eq. (4.3) and Eq. (4.4), DTF, which estimates casual influences based on the transfer function $\mathbf{H}(f)$ and provides directional connectivity across full frequency range, is estimated as:

$$\mathbf{DTF}_{l\to k}(f) = \frac{|H_{kl}(f)|^2}{\sum_{m=1}^{N_k} |H_{km}(f)|^2} \tag{4.6}$$

where $H_{kl}(f)$ is the element in the k-th row and l-th column of $\mathbf{H}(f)$. Furthermore, PDC, which identifies direct causal influences between signals in the frequency domain, is obtained as:

$$\mathbf{PDC}_{l\to k}(f) = \frac{|A_{kl}(f)|^2}{\sum_{m=1}^{N_k} |A_{km}(f)|^2} \tag{4.7}$$

where $A_{kl}(f)$ is the element in the k-th row and l-th column of $\mathbf{A}(f)$, which is the Fourier-transformed, or frequency-domain equivalent, of the MVAR model coefficient matrix, defined as:

$$\mathbf{A}(f) = \boldsymbol{I}_{N_k} - \sum_{m=1}^{p} \boldsymbol{A}'_{kk,m} e^{-j2\pi fm} \tag{4.8}$$

The above three connectivity measurements provide similar information, and the choice between GC, DTF, and PDC depends on the scenario where they are applied to, and the exact information to be extracted. Often, the three measures can be equivalently applied to the same scenario. For the application presented here, given the coefficients obtained for the multivariate GC ($\mathbf{I}_{l\to k}(f)$), where the MVAR model is estimated from the EEG signals, we obtain the combination weights $a_{l,k}$ by normalising (between [0 1]) and averaging them over the frequency band of interest, such as 8-20 Hz for mu rhythm. This $a_{l,k}$ is then used in Eq. (2.12).

The length of the EEG signal segment, together with the selection of the MVAR model order p, significantly influence the quality and reliability of the estimated connectivity measures. Estimating an MVAR model assumes that the underlying signal is stationary over the analysis window. A longer segment provides a greater number of data points, which improves the statistical robustness and stability of parameter estimation. However, this may come at the cost of violating the stationarity assumption, thereby degrading the accuracy of the MVAR and derived connectivity estimates. Conversely, a shorter window is more likely to preserve local stationarity but may result in unreliable MVAR estimation due to insufficient data, potentially failing to capture important temporal dependencies.

Furthermore, the model order p determines how many past samples are considered in predicting the current signal value for each channel. A higher order allows the model to capture more complex temporal and cross-channel dependencies, which can enhance the detection of functional interactions, while it also increases the risk of overfitting and the inclusion of spurious dependencies. In

contrast, a model order that is too low may fail to capture essential interactions and underlying dependencies.

Given this, the window lengths and the MVAR model order should be selected taking the nature of the neural activity into consideration, providing a reasonable compromise between estimation reliability and responsiveness to dynamic changes in brain connectivity.

4.2.3 Diffusion Adaptation for Translating Mind to Action for BCI

Fig. 4.2 summarises the proposed connectivity-informed DA model. Suppose we acquire the EEG signals from N_k electrodes from the individual whose brain signal we want to model ($\boldsymbol{u}_{k_i}$) and acquire the same number of EEG signals from the individual used as the system target (d_{k_i}). Each electrode k forms a neighbourhood N_k with all the electrodes that are at a maximum of one-hop distance from electrode k. Each electrode k has access to time realisations $\{d_{k_i}, \boldsymbol{u}_{k_i}\}$ of the electrodes from its neighbourhood and the equivalent electrode, node g, from the other individual (see Fig. 4.1).

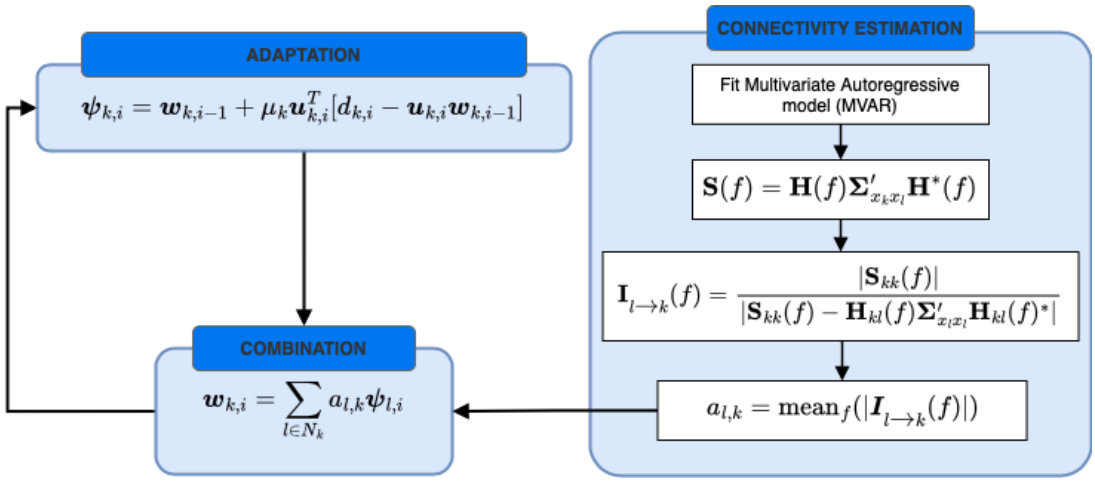

Figure 4.2: An illustration of the proposed method. $a_{l,k}$ represents the average of the connectivity matrix obtained from the GC analysis for the frequency band of interest after being normalised.

For each electrode k from the first individual, we use the EEG signals to form

a $1 \times M$ regression vector in the form:

$$\boldsymbol{u}_{k,i} = [u_k(i) \quad u_k(i-1) \quad \cdots \quad u_k(i+M-1)] \tag{4.9}$$

where M is a non-negative scalar that sets the cooperative filter length and the corresponding signal segment duration.

As discussed previously, d_{k_i} is taken from a global or local desired signal. In the scenario presented here, d_{k_i} is taken from the EEG signals acquired from the second individual, which is the person leading the task during the cooperative EEG-based hyperscanning study. The main goal of the system is to create a model that corresponds to the brain activity from the target individual by defining the network topology with parameters from the brain connectivity information of the first individual, the follower. Using this model, we can analyse if the first individual follows the activities from the second individual and to what degree.

We use ATC defined in Eq. (2.12) at each time instant i to update the estimates $\boldsymbol{w}_{k_i}$. Before applying the combination weights $a_{l,k}$ obtained from the proposed connectivity analysis to Eq. (2.12), they are normalised to satisfy the condition in Eq. (2.8). The DA vector of parameters is then iteratively estimated.

In real-world, such an adaptive system can be applied to the brain under study (rehabilitation) and evaluate the degree of brain recovery. In a BCI scenario, the model is used to assess how well the first individual follows the second one.

The selection of algorithm parameters highly influences the performance and convergence ability of the proposed method. The filter length M defines the number of past samples that will be included in the regression vector $\boldsymbol{u}_{k,i}$, which determines the temporal memory of the adaptive system, influencing the information included in the selected feature space, and the stability and convergence of the system. As discussed in Chapter 2, a balance between the convergence speed and the stability of the system is necessary through the selection of the algorithm parameters.

In the scenario presented in this chapter, a longer filter length allows the model to capture more temporal dependencies from the EEG signals, highly valuable in longer intervals of neural dynamics exchanges, such as during continuous motor movements. However, this also increases the dimensionality of the filter vec-

tor, increasing the computational complexity of the system in each adaptation step. This will also require more data to be available, which could destabilise the adaptation process. In contrast, a shorter filter length reduces the complexity of the system and improves its convergence stability, but increases the possibility of missing important temporal dependencies. Furthermore, in the proposed method, M also determines the signal segment duration, used to estimate the MVAR model.

Therefore, careful consideration is taken during the selection of this parameter, chosen through manual observation based on the sampling rate and neural activities under study to maintaining a balance between the temporal resolution and the learning dynamics stability of the system.

4.3 Data and Experiment

We employ the cooperative hyperscanning data from Experiment 1 (referred to as Dataset 1) and Experiment 6 (referred to as Dataset 2), both described in Section 3.3.2. For Dataset 1, both subjects were asked to open and close the right and left hand in a random order. The leading subject (represented by Subject 1) was asked to lead the movements following a given sequence of instructions. The follower subject (represented by Subject 2) was asked to follow the sequence of movements from the leading subject. For Dataset 2, the leading subject was asked to draw a perfect circle instead following specific given instructions while the follower subject was asked to draw the same circle following the movements from the leading subject.

For both datasets, a total of 64 EEG signals were recorded from two systems, each associated to one subject, during the cooperative hyperscanning study. Each EEG system has 32 channels following the standard 10-20 EEG electrode placement system, as described in Section 3.3.2. Since the experiment consists of the two subjects performing some motor tasks, as shown in Fig. 4.3, in this scenario, from the 32 channels of each system, we only retain and use 10 channels per system associated with the brain motor area. Although this is a moderate number of electrodes, the spatial resolution of the experiment is still acceptable since the number of brain sources used in this experiment is expected to be less than

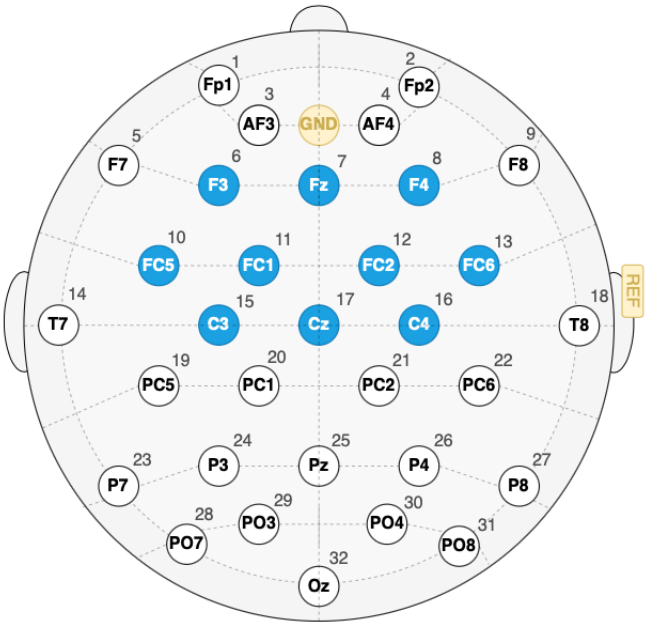

Figure 4.3: An illustration of the electrode setup used for the experiment. The location of the electrodes follows the standard 10-20 EEG electrode placement system. The highlighted electrodes are the electrodes corresponding to the motor area, which are the electrodes used during the performance of the method proposed here.

the number of sources. Even so, if we were to use a higher spatial resolution, it is likely to obtain a higher resolution and, therefore, the system might achieve better results.

Any bad channel or channel with sever artifact is removed from the EEGs of both systems. Then, Infomax ICA from EEGLAB [134] is used to mitigate other artifacts (such as eye-blinks). Then, we perform bandpass filtering to the signals to retain only the frequency bands of interest (alpha and beta, i.e. 8-20 Hz). Finally, the model order of the MVAR model is decided using the Akaike Information Criterion (AIC) [135]. The multivariate GC in the frequency domain ($\mathbf{I}_{l\to k}(f)$) is estimated with the help of FieldTrip Toolbox [136].

4.4 Results

In this section, we present the results obtained by the proposed method when used to model the EEG signals acquired during the above experiments.

4.4.1 EEG Hyperscanning Analysis with Single task Diffusion Adaption

We can confirm that the signals acquired during the experiment present some cooperation between Subjects 1 and 2 based on the similarity between the two patterns and the estimated inter-subject functional connectivity, as shown in Fig. 4.4. However, from this connectivity study, only the normalised intra-brain connections ($a_{l,k}$) from the second subject (follower individual) are used. In this work, the inputs ($\boldsymbol{u}_{k_i}$) to the system are the EEG channels (over motor area) of the second subject and the global target (d_{k_i}) are the temporally and spatially filtered EEGs of the first subject. The filtering process provides a smooth signal less affected by noise and the neighbouring channels.

Fig. 4.5 shows the GC-based connectivity patterns, obtained following Eq. (4.6), between the channels of the second subject for the hyperscanning experiment over the alpha and beta frequency bands after the EEG signals are preprocessed.

To assess the efficiency of the proposed method, we calculate the error between the desired EEG signals (signals from leading subject) and the modelled EEG signals, which are the signals from the follower subject generated after applying the proposed connectivity-informed DA model.

The error for each node k to achieve its target d_{k_i} may be estimated as:

$$E_k = \frac{1}{L}\sum_{i=1}^{L} |d_{k_i} - \boldsymbol{w}_k \boldsymbol{x}_{k_i}|^2 \tag{4.10}$$

where, in this scenario, L represents the number of EEG channel samples, d_{k_i} represents the EEG signals from Subject 1, as defined in Section 4.2, and $\boldsymbol{w}_k \boldsymbol{x}_{k_i}$ represents the modelled EEG signals, constructed with the set of coefficients obtained from the proposed method for Subject 2. $\boldsymbol{x}_{k_i}$ represents the EEG signal

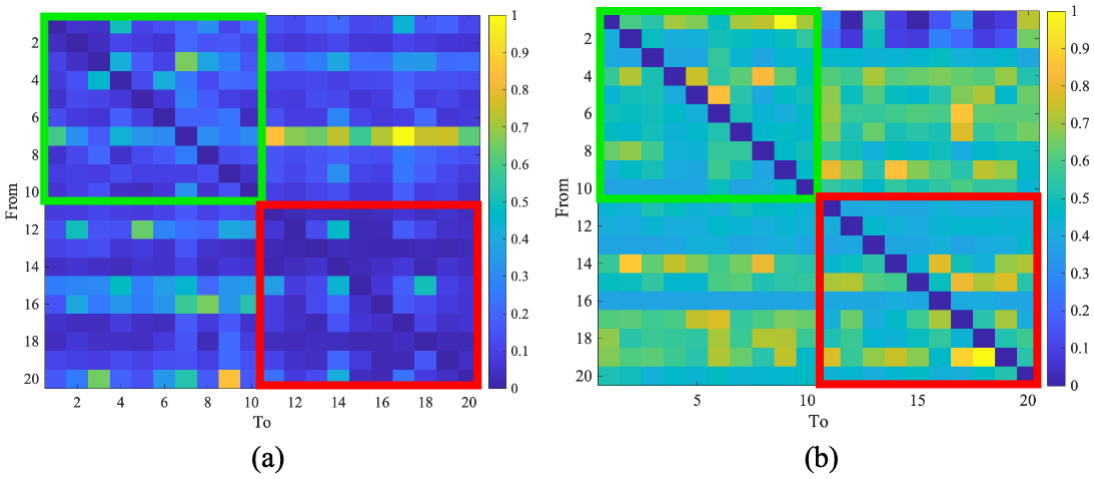

Figure 4.4: Inter-subject GC-based connectivity analysis estimated through $\mathbf{I}_{l\to k}(f)$ between the leading and following subjects over the alpha and beta frequency bands after the EEG signals have been preprocessed. (a) the connectivity matrix for Dataset 1. (b) the connectivity matrix for Dataset 2. The highlighted green and red squares show the intra-subject connectivity for the follower and leading subjects respectively, while the remaining quadrants, left-bottom and right-top, represent the inter-subject connectivity. The numbers in the axis represents the channels: 1 to 10 from follower and 11 to 20 from leading subject.

of node k at time instant i.

4.4.2 EEG Hyperscanning Analysis with Multitask Diffusion Adaption

As discussed in Section 4.2.2, the desired target of the distributed network is the neural activity from the brain of the leading subject, which is recorded through multiple channels. In the scenario presented here, the desired target of each node k is a corresponding node such as g. Therefore, it can be considered as a DA multitask scenario, where each node g represents its own cluster. This related optimisation problem is then solved through multitask DA.

For the multitask approach, the same considerations and steps defined in Section 4.2.3 for the single task approach are followed. The single task formulation, defined as Eq. (2.12), is replaced by the multitask formulation, defined by Eq. (2.33). In the multitask formulation, the combination weights $a_{l,k}$ are obtained

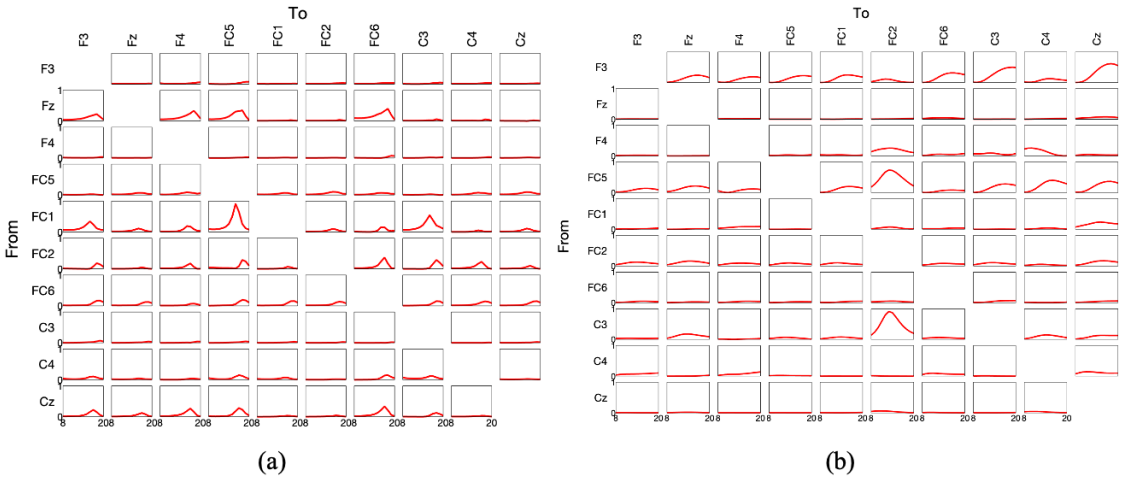

Figure 4.5: Combination weights after normalisation in the frequency domain between a set of nodes represented by the EEG channels. (a) the combination weights from Dataset 1. (b) the combination weights from Dataset 2.

from the connectivity estimation ($\mathbf{I}_{l\to k}(f)$) from the following subject. The regulation factors $\rho_{l,k}$ from Eq. (2.33), that satisfy the conditions in (2.32), are obtained through the connectivity estimation ($\mathbf{I}_{l\to k}(f)$) from the leading subject.

To evaluate the multitask approach against the single task approach, E_k and the model's average running time for all trials are obtained and compared for both approaches. For easier comparison, the proposed and comparison models were evaluated on an Intel Core i5 CPU.

Fig. 4.6 shows the comparison between connectivity-informed single task, multitask, and classic single task DA (where the combination weights are similar) methods in terms of error in achieving the targets. The error is calculated as the average over the number of trials.

From this figure, the proposed connectivity-informed single task and multitask DA methods show considerably better performance than the classic approach for all the channels when applied to the same scenario. This confirms that using the combination weights obtained from a connectivity analysis showing the realistic information flow for each channel leads to significantly more accurate results, especially in complex scenarios such as the scenario from Dataset 1.

When comparing the performance of multitask and single task DA approaches,

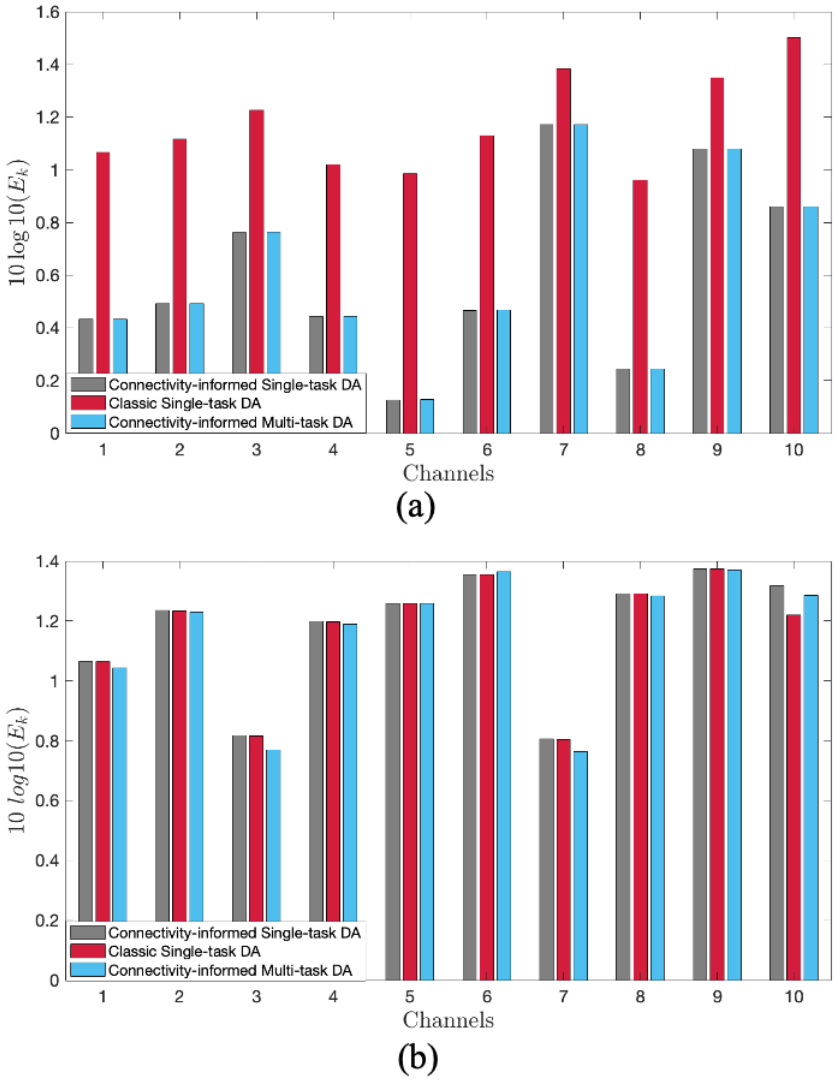

Figure 4.6: Comparison between the average estimation errors of the proposed connectivity-informed single task, multitask, and classic single task DA algorithms over the trials. (a) The performances of the three models for Dataset 1, and (b) the performances for Dataset 2.

it can be appreciated from Fig. 4.6 that the performance of both approaches is highly similar. Fig. 4.7 shows the running time for the two proposed models and the classic approach, showcasing the considerably higher computational cost required by the multitask approach over the single task one. Given the higher computational cost of the multitask approach, which provides an approximately

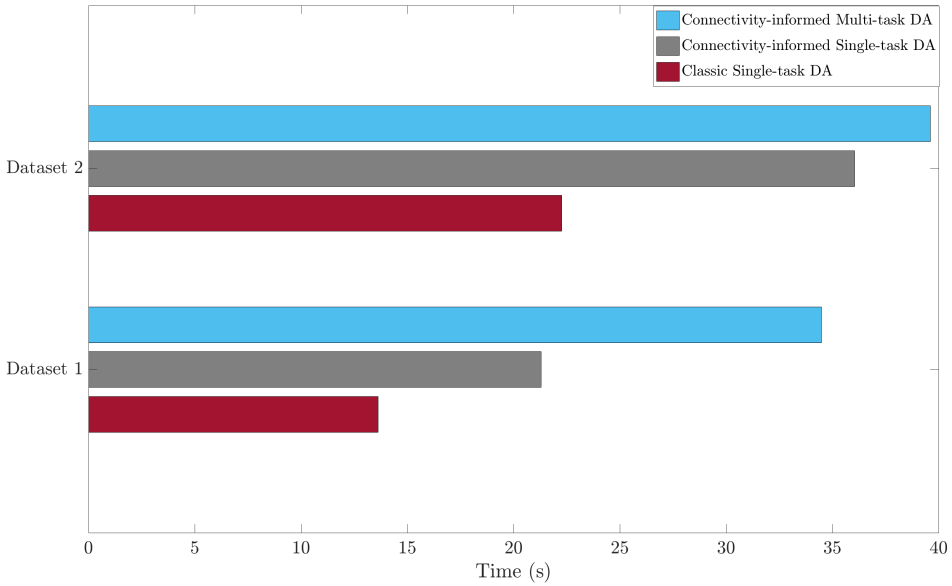

Figure 4.7: Comparison between the running time of the proposed connectivity-informed single task, multitask and classic single task DA algorithms over a single trial.

similar performance to the single task approach, it can be concluded that, although both approaches are suitable, the single task approach is preferred in general due to its lower computational cost. However, the multitask approach might be preferred when dealing with more complex data or neural motor activity, when the higher computational cost might be acceptable in exchange for the better performance of the system.

Given the results presented in this section, it can be appreciated that the proposed method can lead to better analysis and classification of brain responses in hyperscanning scenarios compared to the classic DA model. These results encourage researchers to use BCI hyperscanning studies in the rehabilitation field.

4.5 Conclusions

In this chapter, we proposed a connectivity-informed adaptive cooperative network based on a single task DA that can model the brain responses of a person

based on the brain activities from a second person.

The multitask DA formulation is also implemented and compared against the single task DA formulation for the proposed application. Both formulations are compared in terms of error between the reconstructed and the desired signal as well as in terms of computational cost. After evaluating both approaches, it is established that the suitability of the single or multitask approach depends on the complexity of the system and the involved computational cost.

Two distinct motor task scenarios were employed to evaluate the proposed method: one involving hand opening and closing (Dataset 1), and the other involving continuous circular drawing (Dataset 2). While both tasks require bilateral coordination, they differ significantly in motor complexity, temporal structure, and the nature of the underlying neural coordination.

Task synchronisation in Dataset 1 presented greater challenges due to the involvement of a higher number of fine motor actions, such as individual finger movements, which are inherently more variable and demanding in terms of neural control. In contrast, the circular drawing task in Dataset 2, although still requiring coordinated motor execution, involved more continuous and gross motor movements. This reduced the likelihood of temporal mismatches between the two subjects, leading to a higher degree of movement synchrony and motor pattern similarity.

Such potential inter-subject desynchronisation, whether due to task complexity, timing delays, or inconsistencies in gesture execution, can negatively affect the performance of the proposed method. Nevertheless, the results presented in this chapter demonstrate that the proposed approach effectively models inter-brain relationships across both task types. This suggests that task complexity, in terms of motor control precision, duration, and cognitive engagement, can influence the method's performance. However, the results also confirm that the proposed method maintains a reasonable degree of robustness, even in the presence of moderate task-related variability and desynchronisation between subjects' motor behaviours.

In addition, both datasets were recorded in simulated uncontrolled environments that included natural sources of noise, such as changing lighting conditions and spontaneous ambient disturbances. These factors introduced real-world vari-

ability into the EEG recordings, manifesting as fluctuations in signal quality, muscle artifacts, and inconsistencies in subject attention and engagement.

Despite only minimal preprocessing, the proposed method maintained acceptable performance, demonstrating its resilience to moderate levels of environmental and physiological noise. Even so, it is important to acknowledge that, like most EEG-based modelling techniques, the system remains sensitive to certain noise sources. High levels of muscle contamination, suboptimal electrode impedance, or severe inter-subject asynchrony could impact connectivity estimation and affect the signal reconstruction accuracy.

The proposed method suggests new possibilities for the analysis of motor task related hyperscanning studies and therefore, a further and improved use of hyperscanning in the rehabilitation scenario. Furthermore, although in this chapter the proposed method has been implemented in a hyperscanning scenario with only two subjects, this method can be extended and applied to a higher number of subjects, where one subject is the leading subject and the remaining subjects are treated as multiple followers. In that case, the same proposed method would be applied to each pair of leading-following subjects.

In the next chapter, a new approach for the recognition of prolonged motor movements, such as the continuous drawing of a circle discussed in this chapter, as well as fine movements, such as hand opening and closing, is introduced. The proposed method makes use of the introduced DA strategy to model the relationship between the brain neural activities and the corresponding gesture dynamics. This method can complement the method proposed in this chapter, contributing to the development of a more accurate motor-related EEG hyperscanning-based BCI.

Chapter 5

Cooperative Classification of Prolonged Body Movement from EEG for BCI without Feedback

This chapter is an expansion of our published work in [137].

As discussed in Section 1.2.3, EEG-based BCI systems rely on the identification and classification of neural activities identified and extracted from the EEG signals, such as RP, ERPs, or ERS/ERD. However, these methods have difficulty in precisely and effectively distinguishing mental motor tasks that have highly similar neural activity, such as different fine or prolonged movements.

Although research in BCI that focuses on decoding fine motor activity from brain signals has attracted considerable attention for many years [39, 138, 139], its accuracy highly deteriorates in uncontrolled environments. Furthermore, to the best of our knowledge, the decoding or detection of prolonged uninterrupted motor tasks, also known as prolonged movements, in the absence of any audio, visual or haptic feedback, has not been explored yet.

The ability to accurately detect and evaluate the continuation or sudden interruption of a motor movement is important. Moreover, identification of the desired movement trajectory, and therefore, a specific prolonged movement, could be highly beneficial in applications such as rehabilitation of stroke patients or the development of more precise assistive technology for fully paralysed individuals.

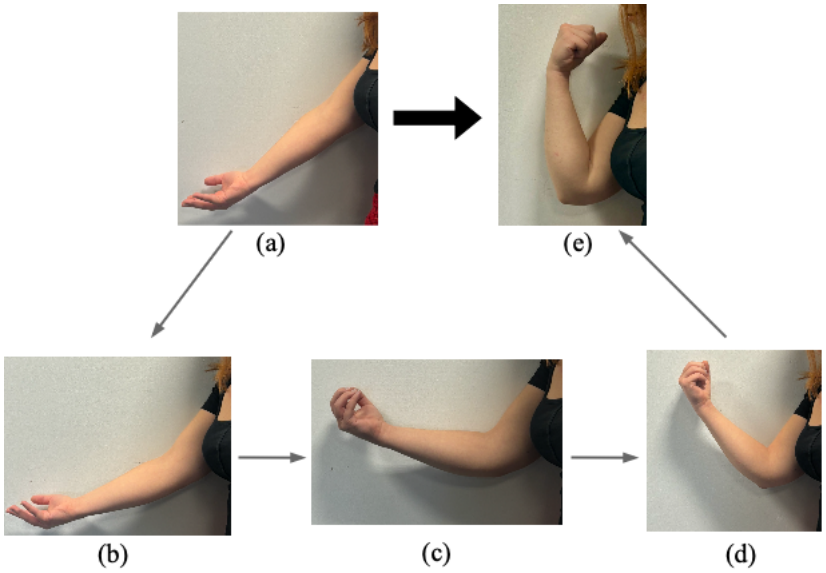

Figure 5.1: An illustration of a prolonged physical motor movement; in this case the flexion of an arm, that also contains progressive closing of the hand. The start and end of the movement are shown respectively in (a) and (e). (b)-(d) show the sequence of sub-gestures that compose the prolonged movement, where each image represents a sub-gesture of the sequence.

In this chapter, we present a novel approach for the recognition of prolonged motor movements from a subject's EEG using orthogonal functions to model a sequence of sub-gestures. In this approach, an individual's EEG signals corresponding to physical (or imagery) continuous movement for different gestures are divided into segments associated with their related sub-gestures, as shown in Fig. 5.1. Then, a diffusion adaptation (DA) approach is introduced to model the interface between the brain neural activity and the corresponding gesture dynamics. In such a formulation, orthogonal Bessel functions are utilised to represent different gestures and used as the target for the adaptation algorithm. This method aims at detecting and evaluating the prolonged motor movements as well as identifying highly complex sub-gestures.

Some of the most common challenges in the implementation of EEG-based BCI in neurorehabilitation include high subject dependency and long training time required by these systems to adjust to each subject. This means that such systems need to be re-trained every time they are used for a new subject [140].

Therefore, the proposed method in this chapter aims at achieving a good inter-subject performance.

5.1 EEG-based Prolonged Movement Decoding

As previously discussed, most BCIs rely on the prediction and classification of the start of a task, associated with a specific brain response, but not on its prolonged and continuous execution. Due to the complex and noisy nature of EEG signals, it can be difficult to differentiate between closely related motor movements. Examples include sequential finger movements, small movement segments involved in flexion and extension of an arm, or closing and opening of a hand. Decoding these details enhances the accuracy in modelling the entire movement. This increases the complexity of decoding a prolonged movement, even when treating it as a set of consecutive small similar gestures.

There have been incredible advances in the BCI field with excellent results on the recognition of motor movements. However, most of these studies focus only on the recognition of the decision to move a specific part of the body, such as to move a left or right arm, a left or right leg, or tongue, but they do not consider any specific trajectory of such movements without feedback. Therefore, they cannot be used to decode and differentiate multiple prolonged movements. Some successful works have explored the continuous 2D reconstruction of fine movement kinematics and trajectories from brain activity based on parameters such as movement speed or direction, as in [141] and [142].

5.2 EEG-based BCI for Fine Movement Classification

The proposed method in this chapter for the decoding of prolonged movements relies on treating a prolonged movement as a set of sequential sub-gestures.

Accurate decoding of fine movements, such as finger movements of the same hand, from EEG is challenging and has been under research for years. The successful outcome of this research can lead to modelling long and continuous

movements without any audio or visual feedback.

Traditional classifiers such as support vector machines (SVM), linear discriminant analysis (LDA), or k-nearest neighbours (KNN), have been extensively used for the classification of EEG signals [107]. Even so, only a small amount of studies have attempted the decoding of complex fine movements, such as in [38, 39, 143, 144]. These studies relied on the combination of traditional classifiers with advanced preprocessing and feature extraction techniques, such as wavelet-CSP [145] or principal component analysis (PCA) [146]. However, these studies obtained only an average of 40-70% accuracy during the classification of multiple fingers from the same hand, showcasing the difficulty of the task.

In recent years, thanks to the development and improvement of deep neural networks (DNNs), more studies have focused on these complex tasks, obtaining better results compared to using traditional classifiers [39, 40, 147].

5.3 Deep Learning Algorithms for Complex EEG Classification Tasks

In the last few years, deep learning methods have shown outstanding results in applications such as computer vision [148], speech processing [149], or image classification [150], motivating their applications for the classification of more complex MI EEG signals. These methods are able to learn high-level and latent complex features directly from the EEG data [151]. However, the main disadvantage of these models, and the reason why DNNs algorithms are still not being used extensively in EEG classification tasks, is due to lack of sufficient data.

Nevertheless, the use of DNNs for recognition of EEG motor-related tasks has become prevalent in the last decade. Given their ability to recognise motor movement patterns directly from the raw EEG, these methods require minimal preprocessing steps, which allows for a more realistic online implementation of the classification systems. Furthermore, the combination of DNNs and more advanced preprocessing techniques for feature extraction preceding the deep learning model, has proven to improve the classification results while introducing minimal acceptable delays to the online BCI implementation and reducing the amount of

data necessary to train the DNNs [152–154].

Deep learning techniques have shown promising results in the classification of more complex EEG motor movements, such as fine hand [147] and finger movements [40]. Even so, the performance of these models for fine movements still require improvement. So far, DNN methods implemented for multi-task fine movements classification still require access to a high volume of data and significant computational resources, hindering their implementation in neurorehabilitation clinics, where access to these resources and large amounts of data may not always be possible.

Although many types of DNNs have been used for EEG classification, for EEG motor movements, and more specifically, for fine movements, two main types of DNNs have proved to provide the best results: convolutional neural networks (CNNs) and long short-term memory (LSTM) networks [30, 31, 151].

LSTMs have been proven to be highly successful in processing and finding patterns in time series data due to their ability to capture temporal dependencies and long-range patterns in sequential data. It has also been demonstrated that they can provide a considerably better performance than traditional machine learning algorithms and CNNs [155–157]. Given their ability to learn long-term sequential patterns, they are considered to be highly suitable for the processing and classification of continuous time series data, such as the classification of continuous prolonged movements, as the ones presented in this chapter. However, they can only learn patterns sequentially, which limits their ability to process data in parallel in comparison with other DNNs. They also require larger datasets compared to other methods to avoid overfitting. This makes these systems computational intensive compared to other DNNs, limiting its feasibility to be implemented in real-world scenarios. Furthermore, due to the high inter-subject variability of EEGs and the model's ability to learn subject-specific temporal features, these models usually have a low robustness to subject-variability compared to traditional methods and CNNs, meaning that they require model adaptation and fine-tuning when used for a different subject.

In contrast, CNNs can be used to learn spatial features directly from raw EEG signals [158, 159]. They can also perform feature extraction and classification simultaneously, allowing the system to provide acceptable results without the need

for extensive preprocessing steps, and sometimes, even outperform LSTM models [151, 160, 161]. Furthermore, their ability to learn spatial features allow them to process the data in parallel across time and space, reducing their dependency on large datasets. This means that they can make better use of the available resources, leading to lower computational requirements and faster training times, making them more practical for real-time and real-world applications.

Therefore, although LSTMs might provide better final results for EEG classification tasks, especially in the classification of prolonged movements and fine movements, CNNs are still preferred. This is due to their lower computational cost, and smaller dependency on large datasets, resulting in higher feasibility to be implemented in real-world scenarios. Even so, although smaller than LSTMs, CNNs still require access to a considerably large amount of data and resources for its correct performance.

Given that, these methods alone are still unable to obtain high performance while maintaining a low computational cost. Hence, the accurate classification of fine movements using low computational cost systems still remains a research challenge for BCI motor rehabilitation systems.

5.4 Problem Formulation for Diffusion Adaption with Orthogonal Target

Through the use of cooperative learning and DNNs, the proposed method models the interface between each sub-gesture and the related neural activity. The aggregate of these sub-gestures provides sufficient and detailed information for the classification of the entire movement.

In this section, we propose a novel approach for the detection and classification of continuous movement, referred to as gesture, from EEG signals by assigning orthogonal basis functions for representing different sub-gestures. This is mainly to translate each 3-D gesture movement profile to a 1-D signal from an orthogonal set, as shown in Fig. 5.2. This method makes use of a cooperative learning technique for the development of a cooperative regressor which, together with a DNN, can identify each gesture. Various order Bessel functions [162] are used

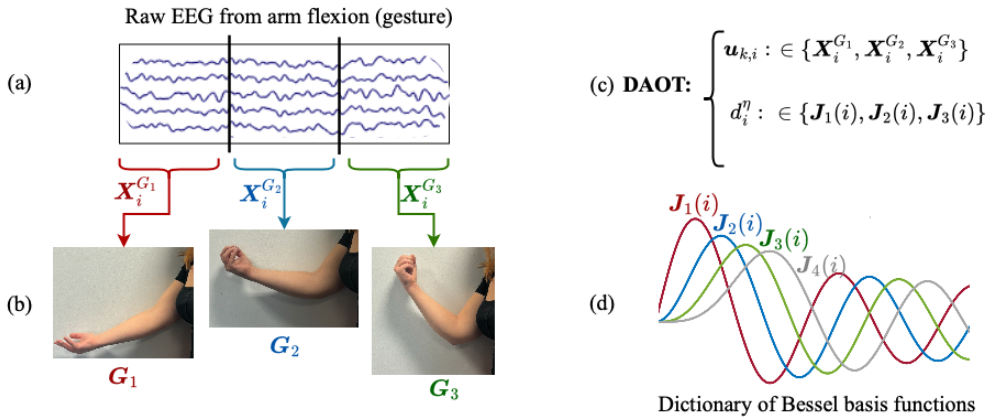

Figure 5.2: Illustration of the translation of a 3D prolonged movement (gesture), such as flexion of an arm as shown in Fig. 5.1, to a 1D signal profile composed of a sequence of Bessel functions representing various sub-gestures. (a) shows the EEG signals that are recorded for the full gesture, and how they are divided so different segments of the signal represent various sub-gestures that compose the full gesture. (b) shows a visual representation of the sequence of sub-gestures that compose the gesture, where each sub-gesture corresponds to an EEG segment of the initial EEG prolonged movement. The Bessel function assignment to the sub-gestures is depicted in (c) and a dictionary of Bessel functions in (d).

for this purpose and assigned to each sub-gesture. However, the size and number of orthogonal Bessel functions is limited, so similar sub-gestures are allocated similar functions. The number of sub-gestures is decided based on the duration and characteristics of the performed motor task.

In the proposed approach, the single task DA strategy, defined in Eq. (2.12), is employed for the reconstruction of each sub-gesture that forms the sequence of sub-gestures (or full gesture). In the approach proposed in this chapter, for each sub-gesture, $d_{k,i}$ represents the assigned order Bessel function.

5.4.1 Bessel Functions as Orthogonal Targets

Bessel functions [162] are the solutions to the Bessel's differential equation, defined as:

$$i^2 \frac{\mathrm{d}^2 y}{\mathrm{d}i^2} + i \frac{\mathrm{d}y}{\mathrm{d}i} + (i^2 - \nu^2) y = 0 \tag{5.1}$$

They have been used in multiple applications, such as for solving physics or engineering problems [163]. Some techniques based on Bessel functions, such as the Fourier-Bessel expansion [164], have also been extensively used as feature extraction techniques for the modelling and classification of brain responses [165, 166]. However, they have never been used for translating 3-D profiles to 1-D orthogonal vectors which can be used as distinct targets for adaptation.

Although there are different types of Bessel functions, the most common types are the Bessel functions of the first and second kinds. In this application, we make use of the Bessel functions of the first kind, defined by:

$$J_\nu(i) = \left(\frac{i}{2}\right)^\nu \sum_{n=0}^{\infty} \frac{\left(\frac{-i^2}{4}\right)^n}{n!\Gamma(\nu+n+1)} \tag{5.2}$$

where Γ represents the Gamma function, n is a non-negative integer that indexes the terms in the series representation, and ν is a constant parameter that determines the order of the Bessel function.

Other orthogonal time series, such as Fourier functions [167], could also be used. However, the Bessel functions of first kind are selected due to their analytical simplicity. These functions manifest more variations in shape and amplitude compared to sinusoids, and have the ability to accommodate variable boundary conditions.

5.4.2 Diffusion Adaptation with Orthogonal Target for EEG-based BCI

For the proposed method, as in Chapter 4, consider a set of EEG electrodes that form a cooperative network. Each electrode k forms a neighbourhood N_k with all the electrodes. Each electrode k has access to time realisations $\{d_{k,i}, \boldsymbol{u}_{k,i}\}$ of the electrodes from its neighbourhood, as shown in Fig. 5.3. For each electrode or channel k, we obtain a $1 \times M$ regression vector in the form:

$$\boldsymbol{u}_{k,i} = [u_k(i) \quad u_k(i-1) \quad \cdots \quad u_k(i+M-1)] \tag{5.3}$$

where M is a non-negative scalar equal to the number of taps in the cooperative

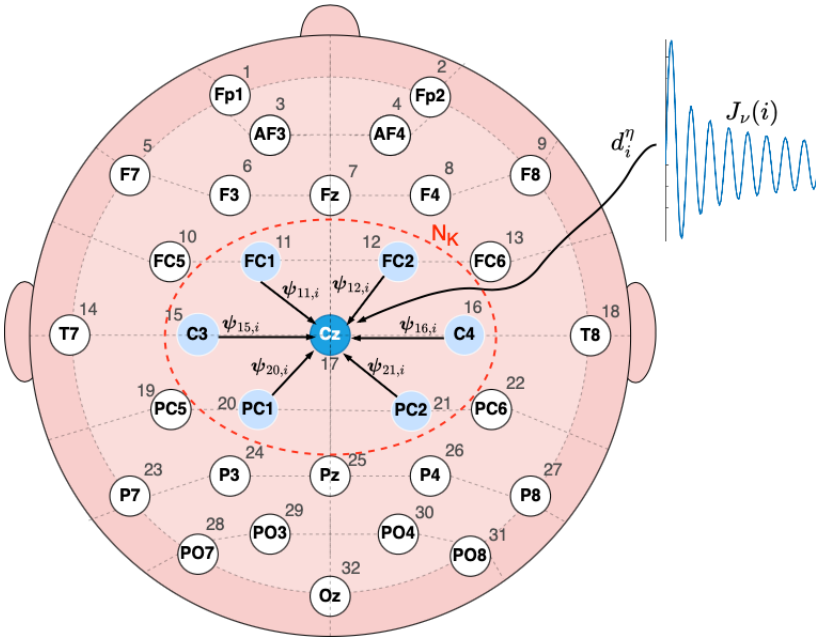

Figure 5.3: An illustration of the cooperation in a DA setup. The electrodes displayed in this figure follow the 10-20 international electrode placement system for 32 electrodes. The nodes from the neighbourhood cooperate between each other while following the target signal d_i^η. N_k represents the neighbourhood of node k when $k = C_z$. $J_\nu(i)$ is the Bessel function of the first kind of order ν associated with the target signal d_i^η.

adaptive filter or the corresponding signal segment duration.

Prior to the application of DA, we need to allocate a limited set of Bessel functions to the existing sub-gestures. Each sub-gesture is represented by a different order Bessel basis function. This basis function is then used as the common target $d_{k,i}$ of the DA strategy for all the nodes, which results in that sub-gesture G_η, such as:

$$d_i^\eta = J_\nu(i), \qquad \nu \in D \tag{5.4}$$

where D is the dictionary of selected Bessel functions, and η represents the number of sub-gestures. In the presence of multiple gestures, each containing a number of sub-gestures, each sub-gesture is assigned a Bessel basis function, with similar Bessel functions allocated to similar sub-gestures. Therefore, there is a

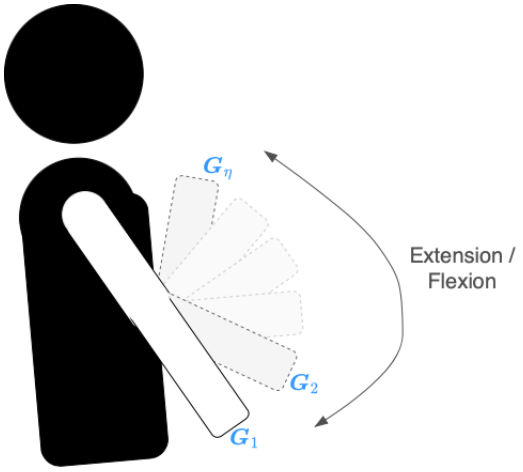

Figure 5.4: Representation of a continuous arm motor execution. $Q = \{G_1, G_2, \cdots, G_\eta\}$, where Q represents the arm extension or arm flexion gesture, which is composed of η sub-gestures.

ratio between η and ν, and generally, $\eta \geq \nu$. This implies that the larger the dictionary of Bessel functions, the better performance of the system is expected and vice versa. Given all this, Eq. (2.12) can be re-defined as:

$$\boldsymbol{\psi}_{k,i} = \boldsymbol{w}_{k,i-1} + \mu_k \boldsymbol{u}_{k,i}^T [d_i^\eta - \boldsymbol{u}_{k,i} \boldsymbol{w}_{k,i-1}] \tag{5.5}$$

$$\boldsymbol{w}_{k,i} = \sum_{l \in N_k} a_{l,k} \boldsymbol{\psi}_{l,i} \tag{5.6}$$

Figs. 5.2 and 5.4 demonstrate a clear view of what the considered sub-gestures of a motor executed movement are.

Finally, for a more realistic approach, the combination weights $a_{l,k}$, as in Chapter 4, are estimated through estimation of the brain connectivity. For this approach, it was decided to employ the time-domain MVAR GC [168] between the EEG channels as the brain connectivity estimation due to the higher dependency of the proposed method on small changes on the connectivity measures over time. The time-domain MVAR GC is defined as [133]:

$$GC_{x_l \to x_k} = \ln \frac{|\boldsymbol{\Sigma}'_{x_k x_l}|}{|\boldsymbol{\Sigma}_{x_k x_k}|} \tag{5.7}$$

where $\mathbf{\Sigma}_{x_k x_k} = \text{cov}(\boldsymbol{\varepsilon}_{k,i})$ and $\mathbf{\Sigma}'_{x_k x_l} = \text{cov}(\boldsymbol{\varepsilon}'_{k,i})$ are the residuals covariance matrices of the MVAR model as an approximation to GC. $\boldsymbol{\varepsilon}_{k,i}$ and $\boldsymbol{\varepsilon}'_{k,i}$ are defined in Chapter 4. As in Chapter 4, the MVAR model is estimated from the EEG signals.

5.5 Application of Diffusion Adaptation with Orthogonal Targets for EEG prolonged movements

The proposed DA with orthogonal target method models the interface between the brain and each sub-gesture. The estimated output signals are then given to a DNN classifier.

Before applying the proposed method, the data are preprocessed to mitigate the effects of EEG artifacts. Although nowadays several artifact detection and rejection techniques have been successfully developed [104, 134], EEG artifact removal is still considered a difficult task as it depends on the BCI tasks, spurious brain functions, and the nature of the recorded data. For this reason, and to reduce the preprocessing time and complexity of the system, only minimal artifact removal and preprocessing steps are applied to the data.

For validation of the method, two EEG motor movement datasets are used. For the first dataset (Dataset 1) we use the data from experiments 1 and 2 described in Section 3.3.2. In these experiments, the subjects were asked to perform an extension and flexion of the left or right arm and the simultaneous opening and closing of the respective hand. For simplicity, only the left arm movement was considered in this application. Although the experiments contain hyperscanning data, we separated the EEG data from each subject out of each pair and analysed each individual's EEG independently.

The second dataset (Dataset 2) was obtained from a public dataset available at [169]. It contains EEG data from 52 subjects performing imagery and physical motor movements. For simplicity and consistency with the previous dataset, only the physical left hand movements from the first sixteen subjects were considered in this application. This allows us to evaluate the proposed method on a consid-

erably small dataset, and compare the method's performance to those of other methods that have previously used the full size dataset. The movement recorded in this dataset involves continuous movement of all the fingers, starting from the index and touching each finger to their thumb within 3 seconds. The subjects were given time to practice and get used to the task before the recording sessions. The data from this dataset were recorded at 512 Hz sampling frequency using 64 channels per subject.

We selected these two datasets since, compared to other publicly available motor-related EEG datasets, they contain recordings of complex fine movements. At the same time, we validate the proposed method on two datasets containing different motor tasks and recorded under completely different settings. This is to prove the ability of the proposed method in providing good performance under different settings.

5.5.1 Data Preprocessing

For both datasets, the motor movement segments $\boldsymbol{x}_{k,i}$ are obtained for each subject in the form of a $N_k \times L_s \times S$ matrix, where N_k represents the number of channels, L_s represents the segment length, and S represents the number of segments. A bandpass filter of 8-30 Hz is applied to the data to reduce possible power line noise and retain the frequency bands of interest associated with motor movements (alpha and beta).

Then, the two datasets are analysed independently and the bad channels are removed. In Dataset 1, channel 27 of odd subjects and 32 of even subjects present a low electrode-skin impedance value across most of the recording sessions, so they are considered bad channels (both channels are distant from the motor area so they do not affect the analysis). For consistency across all the subjects of Dataset 1 and simplicity of the system, both channels are removed from the EEGs of all the subjects. Dataset 2, on the other hand, does not contain any bad channels.

After the signals are filtered, the baseline is also removed from the signals by subtracting the mean from the filtered signal. The MVAR model is obtained from the EEG data and the GC matrix is estimated for each subject and motor movement. If the signal length of each sub-gesture is enough to obtain a stable

MVAR model, for each sub-gesture, we potentially have different weights. The MVGC MATLAB toolbox [133] is used to calculate the MVAR model and GC values. Finally, the preprocessed signals are normalised.

5.5.2 Feature Augmentation

Once all the segments are preprocessed, we decide on the desired number of sub-gestures and the dictionary of Bessel functions. The window length for each sub-gesture is obtained as:

$$L_w = \frac{L_s}{\eta} \tag{5.8}$$

For each sliding window $\boldsymbol{x}_{k,i}^{\eta}$ with no overlap over each motor movement segment, we apply Eq. (5.5)-(5.6) to estimate the optimum $\boldsymbol{w}_{k,i}$. Then, we obtain the estimated sub-gesture signal segments as:

$$\boldsymbol{y}_{k,i}^{\eta} = \boldsymbol{x}_{k,i}^{\eta} \boldsymbol{w}_{k,i} \tag{5.9}$$

These estimated segments obtained using DA, which contain differentiable signals for each sub-gesture, are used to estimate the DA filtered output matrix ($\boldsymbol{Y}^{Q}$) that represents each gesture. The output signals are given as the input to the chosen classifier. Algorithm 2 shows the pseudocode to obtain the estimated signal segments.

5.5.3 Classification

Once the estimated signals are obtained, a classifier is selected and trained to recognise the sequences of sub-gestures.

Given the discussion presented in Section 5.3, it was decided to implement a CNN as the classifier for the proposed system. The implemented CNN architecture follows the recommendations reviewed in [30] for implementation of a shallow CNN for EEG pattern recognition. This CNN consists of a 2D input layer, 2 hidden layers (two 2D convolutional layers), and a fully connected output layer. A rectified linear unit (ReLU) is used as the activation function for the two 2D convolutional layers. Each 2D convolutional layer has 8 filters of size 4×4. Finally,

Algorithm 2 DA with Orthogonal Targets

Input: EEG Segment (**EEG**: $N_k \times L_s$)

Prerequisites: Window Length L_w, Initial segment signal point i, d_i^η, $a_{l,k}$

while $L_w + i - 1 \leq$ Segment Length (L_s) **do**
  $\boldsymbol{x}_{k,i}^\eta = EEG(i : i + L_w - 1)$
  while $i + M \leq L_w$ **do**
    for $k = 1$ to N_k **do**
      $\boldsymbol{\psi}_{k,i} = \boldsymbol{w}_{k,i-1} + \mu_k \boldsymbol{u}_{k,i}^T [d_i^\eta - \boldsymbol{u}_{k,i} \boldsymbol{w}_{k,i-1}]$
      $\boldsymbol{w}_{k,i} = \sum_{l \in N_k} a_{l,k} \boldsymbol{\psi}_{l,i}$
    end for
  end while
  $\boldsymbol{y}_{k,i}^\eta = \boldsymbol{x}_{k,i}^\eta \boldsymbol{w}_{k,i}$
  $i = i + L_w$
end while

Output: estimated sub-gesture signal $(\boldsymbol{y}_{k,i}^\eta)$

a Softmax activation function is used in the output layer for decision making. The CNN is trained using a maximum of 30 epochs and a minimum batch size of 8, with cross-entropy as the loss function and an Adam optimiser. The initial learning rate is set to 0.001. The hyperparameters were selected following hyper-tuning and based on the recommendations given in [30], which reviews the most common CNNs and hyperparameters used for EEG classification that leads to the best results.

The same CNN was used for both datasets, with 90% of the full dataset used for training and validation, and the remaining for blind testing of the trained model. During training, we followed a leave-one-subject-out k-fold cross-validation for the split of the training and validation data, and used the model of the fold with the best validation accuracy as the final trained model. To verify the robustness of the model against subject-dependency, we followed an inter-subject evaluation approach for the models, where the split between the training, validation, and testing data sets were done per subject, and not per observation. Therefore, the data from the same subject is always given to only one of the sets, either training, validation, or testing. We used this approach to evaluate the subject-independency of the system. This way, the system is tested, during validation and blind testing, against the data from subjects not used for training.

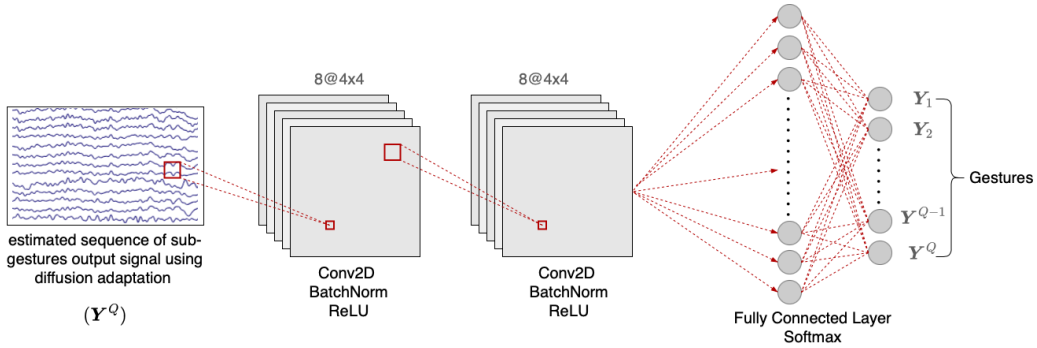

Figure 5.5: Schematic diagram of the proposed CNN. The proposed CNN is a shallow CNN which consists of an input layer, 2 hidden layers (two 2D convolutional layers) and a fully connected output layer. Multichannel EEG is applied to the input as 2D data. Both 2D convolutional layers are composed of 8 filters of 4×4 size. ReLU and Softmax activation functions are used respectively for the convolutional and output layers. The proposed CNN is trained using a maximum of 30 epochs and a minimum batch size of 8, with an initial learning rate of 0.001. It is trained using an Adam optimiser together with cross-entropy as the loss function.

The detailed pipeline of the implemented shallow CNN is shown in Fig. 5.5.

Fig. 5.6 summarises the steps for classification of the DA estimated sequence of sub-gestures. The EEG segments for each sub-gesture are obtained as depicted in Fig. 5.2. The proposed DA with orthogonal targets is abbreviated as DAOT. The classifier is trained using the DA filtered signals thanks to the dictionary of Bessel functions. In the testing stage, the filter output is classified as one of the Bessel functions in the dictionary, which corresponds to one of the sub-gestures. It is assumed that the size of Bessel functions dictionary is always less or equal to the number of sub-gestures. In the next section, the effect of dictionary size is evaluated.

5.6 Results

To validate the performance improvement of the proposed method against state-of-the-art methods, it is compared with EEGNet [170], a well established CNN model that has been successfully used as a benchmark model for motor-related EEG classification. We also compare the proposed method against another highly

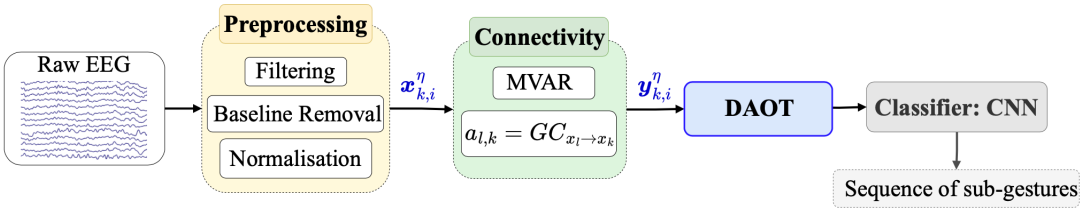

Figure 5.6: A step-by-step illustration of the proposed method. The number of tasks to classify depends on the number of sequence of sub-gestures. The proposed DA with orthogonal targets is abbreviated as DAOT.

successful and more recently developed CNN model, called SCCNet [171], which has shown promising inter-subject results.

Both competing methods, EEGNet and SCCNet, use the EEG signals directly after applying the preprocessing steps defined in Section 5.5.1 and described in [151] as the input. For consistency, the raw signals are segmented for the proposed method, and the same filtering and normalisation techniques, which enhance the quality of the data, are applied to the raw signals before applying each method. For further consistency, the data are split in the same way for all the methods, and the data from the same subjects are used for the same training, validation, and testing sets of each system. For easier comparison, we calculated the accuracy, which measures the proportion of correct predictions of the model, as:

$$\text{Acc} = \frac{TP + TN}{TP + TN + FP + FN} \tag{5.10}$$

where TP represents the true positives, TN the true negatives, FP the false positives, and FN the false negatives, and the sum of all of them represents the total number of observations. We also obtained the F-score, which is a metric that combines the precision and the recall. It reflects the model effectiveness in identifying positive cases and avoiding false positives. F-score is estimated as:

$$\text{F-score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{5.11}$$

where the precision metric is obtained as:

$$\text{precision} = \frac{TP}{TP + FP} \tag{5.12}$$

and the recall as:

$$\text{recall} = \frac{TP}{TP + FN} \tag{5.13}$$

We also obtained the average area under the curve (AUC) score for each fold to estimate the model's generalisability and robustness. This is the area under the Receiver Operating Characteristic (ROC) curve, across all the classes for each fold. Finally, we estimated the pairwise p-value between the proposed method (under the ideal, $\eta \leq \nu$, and non-ideal, $\eta > \nu$, situations) and each comparison method for both datasets. Although multiple statistical tests could be used, we chose to perform a Quade test [172], followed by the Nemenyi test [173], which are non-parametric statistical tests suitable for the comparison of multiple models across repeated samples. Furthermore, the Quade test incorporates relative rankings within each fold, making it an ideal test for the leave-one-subject-out cross-validation approach. Compared to other non-parametric statistical tests, this test also places weights on different folds, to consider the complexity of different folds. We make use of the validation accuracy during each fold of the leave-one-subject-out cross-validation to estimate the Quade test statistics. Then, if the statistical test is significant, we perform the Nemenyi test, which is a commonly used post-hoc statistical test, across pairwise models to obtain the p-value for each method, with the significance level as 0.05.

As previously mentioned, the size of Bessel functions dictionary affects the performance. It is expected that by decreasing the number of orthogonal targets the overall performance drops and vice versa. This has been validated by progressively increasing the number of Bessel functions from one to the total number of sub-gestures from all considered gestures for Dataset 1 and the results are shown in Fig. 5.7. When the size of Bessel functions dictionary is smaller than the total number of sub-gestures, similar Bessel basis functions are allocated to similar sub-gestures. Two symbolic hypothetical movement (gesture) trajectories, each with four sub-gestures, are illustrated in Fig. 5.8. From this figure, it can be appreciated that the correct allocation of the same Bessel basis functions to similar sub-gestures can influence the performance of the system and allow for the correct recognition of the two sequences of sub-gestures. On the other hand,

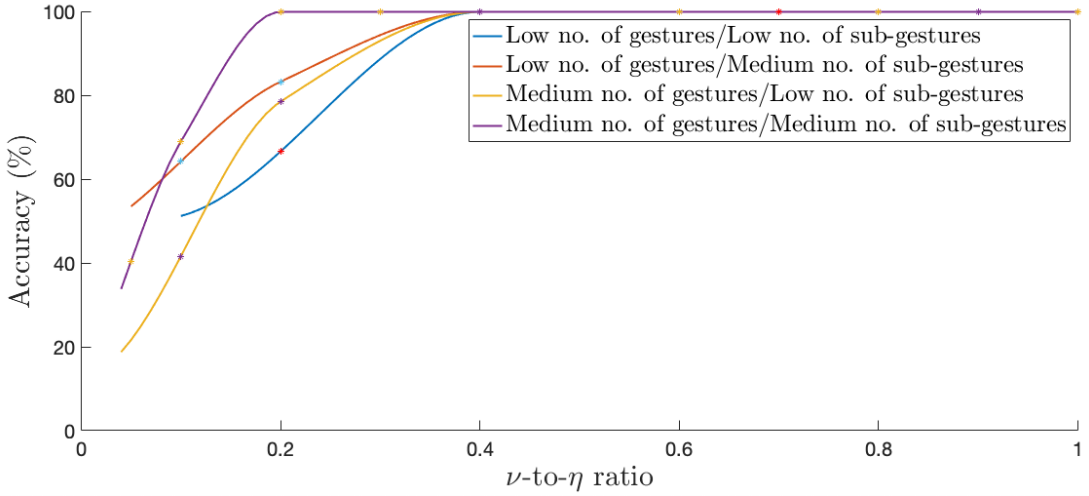

Figure 5.7: Average validation accuracy when classifying a number of sequences of sub-gestures of Dataset 1 for the proposed method using different dictionary sizes of Bessel functions. A low number of gestures represents two gestures while a medium number of gestures represents four gestures. Two sub-gestures per gesture are considered a low number of sub-gestures, while four sub-gestures per gesture are considered a medium number of sub-gestures. What is considered a low or high number of gestures and sub-gestures per gesture depends on the characteristics of the recorded gestures to be classified as well as their nature, and the dictionary size that is available.

it can be appreciated in Fig. 5.7 that the higher the number of sub-gestures, the more orthogonal functions are necessary to obtain a good performance. The overall number of gestures (movements) that can be classified is also radically affected by the number of selected basis functions (dictionary size).

In comparison to Fig. 5.7, for Dataset 1, EEGNet obtained an average validation accuracy of 37.50% for the recognition of four gestures with a low number of sub-gestures, and 50.00% for the recognition of two gestures with a medium number of sub-gestures, while SCCNet obtained 26.79% and 64.29%, respectively.

The proposed method can also be used to accurately recognise highly complex sub-gestures, such as finger movements. This enables the proposed CNN model to accurately classify the sub-gestures, obtaining an inter-subject average validation accuracy of 69.59% for Dataset 1 and 87.68% for Dataset 2, under a non-ideal situation ($\nu = 2/3\eta$) for the classification of ten sub-gestures for Dataset 1 (five

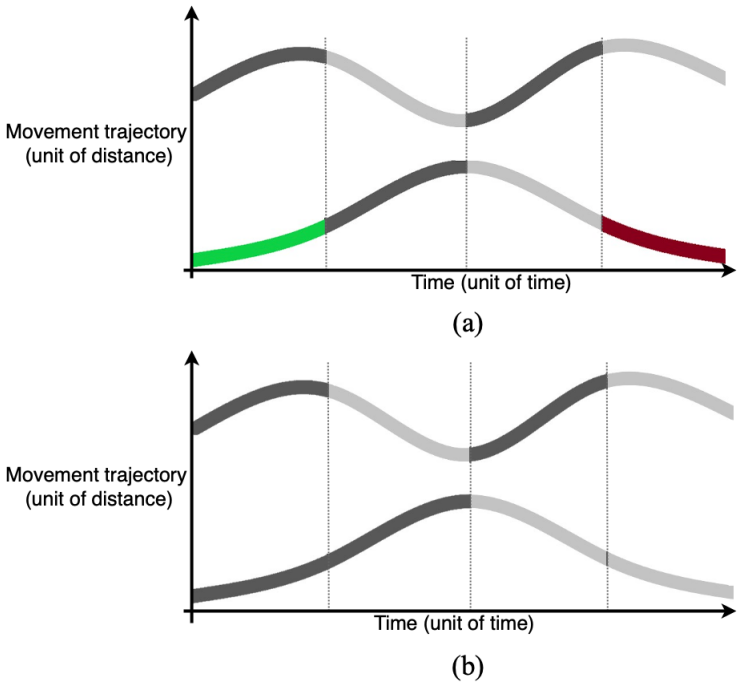

Figure 5.8: Hypothetical 2D signals trajectories of two movements (gestures), each divided into four sub-gestures, with some sub-gestures in common. This represents a non-ideal situation where the number of available Bessel functions is smaller than the total number of sub-gestures ($\eta > \nu$), so similar sub-gestures are assigned the same Bessel functions. In this example, two different gestures, each composed of four sub-gestures, which yields a total of 8 sub-gestures, is to be represented with a Bessel functions Dictionary (D) of size 4 ($\nu = 4$) for (a), and size 2 ($\nu = 2$) for (b).

sub-gestures obtained from the extension of an arm and the other five sub-gestures from the flexion of the same arm) and 4 sub-gestures for Dataset 2 (each sub-gesture representing the movement of one finger: index, middle, ring and little). The recognition of each finger movement as a sub-gesture of sequential finger movements allows comparing the proposed method's performance against the performance of other methods for the recognition of the same fine movements.

Under an ideal situation ($\nu = \eta$), we obtained 99.99% average accuracy for Dataset 1 and 100% for Dataset 2. As expected, when each sub-gesture is filtered using a different orthogonal function as the DA target, the classification accuracy of these sub-gestures is close to 100%. This is due to denoising by adaptive fil-

Table 5.1: Model performance of the proposed and competing methods for sub-gestures classification. The high accuracy of the proposed method is as expected due to denoising by adaptive filtering and the Bessel basis functions' orthogonality. The low performance of the competing methods is also expected due to the strong EEG noise, variability in movement, and small size datasets. The reported metrics are the inter-subject metrics of the models, obtained using data from different subjects for the training, validation, and testing sets. For validation, we performed leave-one-subject-out cross-validation. The best F-score is the metric from the fold with the best validation accuracy. The reported p-value is obtained for each pairwise (proposed method vs. compared method) for each dataset, where we consider the significance level p-value< 0.05.

	Dataset 1				Dataset 2			
	DAOT$_{\nu=\eta}$	DAOT$_{\nu=2/3\eta}$	EEGNet	SCCNet	DAOT$_{\nu=\eta}$	DAOT$_{\nu=2/3\eta}$	EEGNet	SCCNet
Best Validation Acc	1.00	0.83	0.50	0.40	1.00	1.00	0.75	0.75
Best F-score	1.00	0.78	0.41	0.30	1.00	1.00	0.67	0.67
Avg. Validation Acc	0.99 ± 0.004	0.70 ± 0.094	0.17 ± 0.144	0.11 ± 0.122	1.00 ± 0.00	0.88 ± 0.150	0.48 ± 0.199	0.68 ± 0.113
Avg. Validation F-score	0.99 ± 0.004	0.65 ± 0.094	0.12 ± 0.12	0.06 ± 0.090	1.00 ± 0.00	0.85 ± 0.178	0.40 ± 0.200	0.58 ± 0.130
Avg. Validation AUC score	1.00 ± 0.00	0.94 ± 0.041	0.57 ± 0.100	0.52 ± 0.080	1.00 ± 0.00	0.97 ± 0.044	0.74 ± 0.180	0.79 ± 0.090
Testing Acc	0.99	0.72	0.15	0.10	1.00	0.80	0.25	0.50
Testing F-score	0.99	0.72	0.15	0.10	1.00	0.76	0.12	0.53
p-value (DAOT$_{\nu=\eta}$ vs. Competing Method)	–	0.170052	0.000008	$3.64e-07$	–	0.459525	0.000003	0.001086
p-value (DAOT$_{\nu=2/3\eta}$ vs. Competing Method)	0.170052	–	0.027748	0.004232	0.459525	–	0.001903	0.105335

tering, making the outputs more orthogonal and separable. The lower accuracy is expected for Dataset 1. This is because the sub-gestures are more similar and there are higher number of classes to be classified compared to Dataset 2. On the other hand, with EEGNet we obtained an average validation accuracy of 17.14% for Dataset 1 and 48.21% for Dataset 2, and 10.71% and 67.86% respectively for SCCNet. Although, as expected, SCCNet shows a better inter-subject classification performance than EEGNet, both competing methods still show a worse performance than the proposed method. The low accuracy obtained for both competing methods can be associated with the strong EEG noise, the variability in movement, and the small dataset size, compared to the successful performance of EEGNet in [170] and [174], and SCCNet in [171], where both models were trained with considerably larger datasets.

Furthermore, for Dataset 2, if we decrease the number of subjects from 16

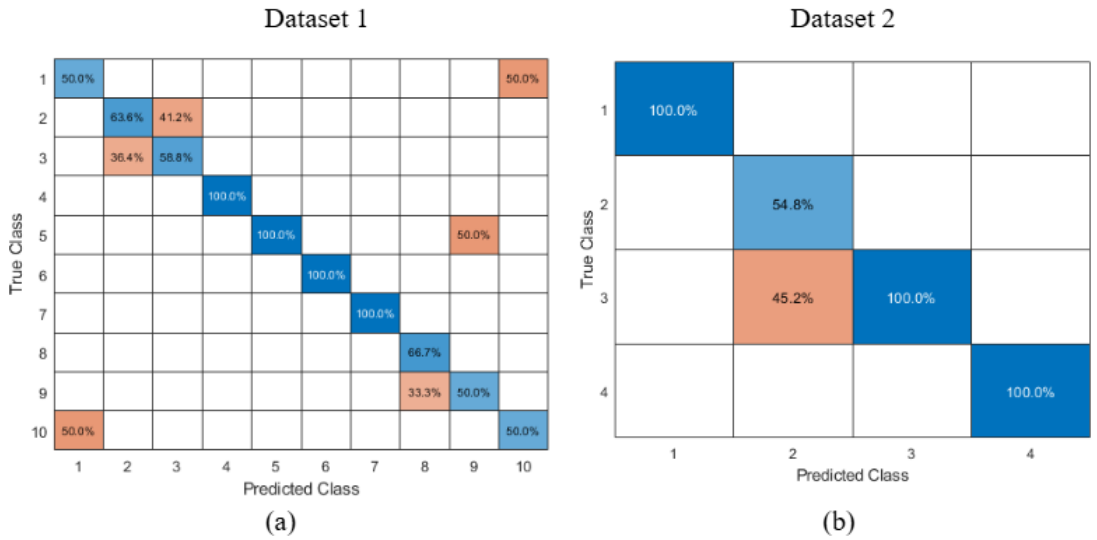

Figure 5.9: Confusion matrix obtained for the testing set using the DAOT method under non-ideal conditions ($\nu = 2/3\eta$) for both datasets, where each class of the confusion matrix represents a sub-gesture. The trained model with the best validation accuracy and corresponding F-score was selected as the final model. (a) shows the confusion matrix of the 10 sub-gestures obtained for Dataset 1, while (b) shows the confusion matrix of the 4 sub-gestures obtained for Dataset 2.

to 8 for training the competing models, we obtained 32.14% average validation accuracy for EEGNet, and 50.00% for SCCNet, which confirms that the methods obtain a better performance for a bigger dataset, and also obtain a worse performance if the dataset size is reduced.

The confusion matrix in Fig. 5.9, which shows the accuracy of the proposed method for the recognition of sub-gestures during testing, further verifies the good performance of the proposed method. Evidently, the proposed method, even under non-ideal conditions (e.g., $\nu = 2/3\eta$), can classify accurately the majority of sub-gestures. As expected, the proposed method can classify with no error all the sub-gestures that were assigned a unique Bessel basis function. The classification error increases when similar Bessel functions are allocated to relatively (not exactly) similar sub-gestures.

The proposed and the competing methods were trained on an Intel i9-13900F CPU. The MATLAB Parallel Computing Toolbox was used to improve the per-

formance of all the trained models through optimisation of the available resources, using 8 cores. To compare the computational cost and complexity between the proposed and competing methods, we obtained the training run time of all these methods for the two datasets. Since the proposed method has a heavier preprocessing stage but a smaller CNN, the training run time includes the preprocessing time. Given all that, the proposed method demonstrated a 32.74% improvement in the training speed over EEGNet for Dataset 1, reducing the training time from 3.74 to 2 minutes. For Dataset 2, it obtained a 25.54% improvement. When compared against SCCNet, the proposed method obtained a 49.43% and 77.99% improvement for Dataset 1 and 2 respectively. Although SCCNet obtains better performance results than EEGNet, SCCNet is more computationally intensive. Dataset 2 contains more data than Dataset 1, which increases the training time for the models, and also explains the higher performance of all the models for this dataset. Overall, these results prove the lower computational cost and training time of the proposed method compared to latest state-of-the-art methods.

5.6.1 Ablation Study: Proposed Model vs Variations

To better understand the relevance of the different components involved in the proposed method and how each component contributes to the overall performance, we conducted an ablation study where we evaluated four variations of the proposed method. Each variation was obtained by removing one of the main features. For a better evaluation, we tested the ablated methods on Dataset 1, the more complex and smaller dataset, following the leave-one-subject-out cross-validation described in Section 5.5.3, and obtained the same metrics described in Section 5.6. To evaluate the contribution of the proposed filtering method, we also estimated the mean squared error (MSE) and SNR between the original signal that represents a sub-gesture and its filtered version. MSE and SNR are estimated only for the ablated models where the adaptive filtering is performed. Therefore, these values are not estimated for the ablated model that does not perform DA. This is because in this case, both pre- and post-filtered signals are the same. In Table 5.2 we reported the average MSE and SNR over the first sub-gesture for all the subjects. MSE is estimated as:

Table 5.2: Ablation study results: the p-value is obtained between the full method and each ablated method through the Quade test followed up by the post-hoc Nemenyi test.

	Experiment	No Diffusion Adaptation	No GC ($a_{l,k} = 1/N_k$)	Reduced CNN (only 1 Conv2D layer)	Reduced CNN (num. Filters=2)
Dataset 1	Best Validation Acc	0.30	0.90	0.91	0.94
	Best F-score	0.27	0.87	0.90	0.94
	Avg. Validation Acc	0.14 ± 0.10	0.71 ± 0.12	0.71 ± 0.10	0.71 ± 0.16
	Avg. F-score	0.10 ± 0.08	0.64 ± 0.12	0.67 ± 0.11	0.68 ± 0.16
	Avg. AUC Score	0.55 ± 0.08	0.93 ± 0.05	0.93 ± 0.05	0.92 ± 0.07
	MSE	–	0.21 ± 0.02	0.20 ± 0.03	0.20 ± 0.03
	SNR	–	-0.23 ± 0.06	-0.30 ± 0.06	-0.30 ± 0.06
	p-value	0.002	0.994	0.994	0.919

$$\text{MSE} = \frac{1}{L_s} \sum_{i=0}^{L_s-1} (\boldsymbol{X}_i - \boldsymbol{X}'_i)^2 \tag{5.14}$$

where $\boldsymbol{X}_i$ represents the signal segment of length L_s at time instant i, and $\boldsymbol{X}'_i$ represents the corresponding filtered signal. SNR is estimated as:

$$\text{SNR} = 10 * \log_{10}\left(\frac{\sum_{i=0}^{L_s-1} \boldsymbol{X}_i^2}{\sum_{i=0}^{L_s-1} (\boldsymbol{X}_i - \boldsymbol{X}'_i)^2}\right) \tag{5.15}$$

Through the ablation study, we are particularly interested in evaluating the impact of the proposed adaptive filtering technique, and how the use of GC for estimating the combination weights ($a_{l,k} = GC_{x_l \to x_k}$) affects the final system as compared to applying the same combination weights to all the nodes ($a_{l,k} = 1/N_k$). We are also interested in confirming if a smaller CNN could be used.

As shown in Table 5.2, the full method outperformed the majority of ablated models for every performance metric. The MSE and SNR values show the efficiency of the proposed filtering method and that the use of GC as the combination weights helps retain more information from the original signal. This increases the robustness of the system against individual subjects' EEG variations, as shown by the higher standard deviation of the average accuracy, F-score and AUC. The

results obtained from the ablated model without DA testifies the importance and relevance of the proposed filtering technique. The performance of the ablated model with no DA also shows similar performance to the two competing methods, EEGNet and SCCNet, and therefore worse performance than the proposed full system. This shows that a shallow CNN with no special preprocessing steps is not enough for the highly complex task of EEG fine movement recognition.

Furthermore, based on the results obtained for the reduced CNN ablated methods, we can conclude that the originally proposed CNN, depicted in Fig. 5.5, provides the best performance with the smallest possible architecture. Although the ablated models with only one 2D convolutional layer, or reduced number of filters, obtain very similar results, the standard deviations of the average metrics for this model are higher than for the full model, showcasing how the use of two convolutional layers provides a more consistent performance across all subjects. This leads to a robust system against inter-subject variations.

5.6.2 Discussions

As mentioned in previous sections, during the recognition of a prolonged movement, denoted by a sequence of sub-gestures, the number of gestures to be recognised and the number of sub-gestures that compose each gesture, can have an impact on the method's performance considering the non-ideal scenario where we only have access to a limited-size Bessel functions Dictionary. In the specific scenario of Fig. 5.7, due to the limited recording time for each available gesture from Dataset 1, empirically, four sub-gestures per gesture have been selected. Gestures with longer duration usually require more sub-gestures.

The two state-of-the-art methods, EEGNet and SCCNet, were selected for comparison due to their successful performance in the recognition of EEG motor-related tasks. EEGNet has been used successfully for the classification of fine movements similar to Dataset 2 as in [175] and [174], making it an ideal method for comparison. On the other hand, although SCCNet has not been implemented for classification of fine hand movements, this method has been used for classification of more common motor task scenarios (i.e., left hand, right hand, both feet, and tongue motions), and shown good results in inter-subject classification

settings, showcasing the model's ability to generalise and its robustness to EEG subject variations.

Despite the fact that the proposed method outperforms the two competing methods, similar to other methods, it performs worse for Dataset 1 compared to Dataset 2. This is expected due to the more complex data of this dataset. Still, the confusion matrix in Fig. 5.9 shows that the proposed method can recognise highly similar sub-gestures with an acceptable accuracy even after filtering. This shows that the filtered segments still retain their underlying unique information.

Table 5.1 verifies the above claims. The estimated p-values for the competing methods further validate the statistical superiority of the proposed method. On the other hand, the p-value of the proposed method under ideal and non-ideal situations evidences that, by increasing the dictionary size above a certain level for each dataset, the methods' performance metrics no longer show any statistical significant improvement.

It is also worth mentioning that EEGNet as well as other DNNs have already been used for the classification of fine hand movements, such as in [174], where EEGNet obtained an average intra-subject accuracy of 62.30% with a dataset size bigger than the datasets used in this chapter. In [152], the authors proposed an autonomous deep learning model instead, obtaining a subject-independent, commonly referred to as inter-subject, average accuracy of 78%, showcasing a similar performance to the method proposed in this chapter. Even so, their proposed deep learning model is larger and more complex, which increases the computational complexity of the method, and usually requires larger datasets.

These results further validate the advantages of the proposed method, which can obtain a higher performance even with small datasets. This reduces the necessity for big datasets and the training time for a good performance. However, although increasing the size of the datasets enhances the performance for the other competing DNN methods, obtaining a more extensive dataset can be expensive, difficult, and may not always be possible. Furthermore, increasing the size of the training dataset also increases the computational cost and training time of the models, reducing their feasibility for real-world implementation. This might require adjustments to the changes to the motor tasks that the model should recognise. Therefore, the low computational cost of the proposed method in com-

parison to the two competing methods, as stated in the previous section of this chapter, highlights the method's suitability for rapid and resource-constrained retraining. This shows that the proposed method can be implemented and trained with standard hardware without the need for extensive and expensive computational resources, making the proposed method more suitable for cost-effective applications, including those for neurorehabilitation.

Although the inter-subject variability still remains a major challenge in MI classification, most current studies still focus on subject-dependent MI. This makes such systems difficult to implement in real-world scenarios where the same system may need to be used by different users without having time for recalibrating the system [151]. The proposed method has high performance even when trained and tested with small inter-subject data, showing the high robustness to inter-subject variability of the proposed system.

Even so, regardless of its considerably good performance, certain aspects of the proposed system could be further improved. In the current system, under non-ideal conditions, similar Bessel basis functions are allocated to similar sub-gestures. The number of Bessel functions (dictionary size) is directly related to the number of gestures and sub-gestures, and complexity of the movement trajectories, ranging from less than the number of gestures to as high as the total number of sub-gestures, depending on the similarity between the movements and the desired accuracy. To enhance the system, and improve its feasibility for real real-world settings, a dictionary learning algorithm can be developed and used to optimise the Bessel basis functions allocation process. In addition, the evaluation of the proposed system with real data from subjects with neurological disorders or under neurorehabilitation can verify its practical benefits.

5.7 Conclusions

In this chapter, we have established a new research direction in BCI where prolongation of physical movement is linked to the communications between different brain zones within the motor cortex captured by scalp EEG. This interface has been modelled as a cooperative network optimised by DA. In addition, incorporating the combination weights into the cooperative adaptive filtering algorithm

boosts its accuracy.

Each EEG segment is decoded to a segment of hand movement called a sub-gesture. The combination weights of the cooperative model are estimated through GC estimation. The sub-gestures are then sequenced for recognition and identification of the prolonged movement. This, together with using a shallow CNN, forms a regressor that can be used to classify the entire movement. The results verify the high accuracy in recognition of fine movements, especially when the movements are similar, meaning that their EEG signatures are highly similar. Finger movement is an example of such fine physical activities. Allocation of orthogonal Bessel functions to the set of gestures further enhances the classification accuracy.

The proposed method outperforms other state-of-the-art methods in the case of inter-subject classification providing a high inter-subject performance with a small training dataset and low computational cost. This allows the system to adapt and re-train for different scenarios that might require different motor tasks, on highly accessible low-cost devices.

Furthermore, through the use of leave-one-subject-out cross-validation, we can evaluate the robustness of the proposed method to subject's EEG variability. Consequently, we can conclude that the proposed method is robust to subject variability, alleviating the need for re-training for each subject, improving the method's feasibility to be used in a real-world setting, such as in a neurorehabilitation clinic.

Moreover, the robustness of the method was further validated under a range of noise conditions and motor task complexities to assess its practical applicability in real-world BCI scenarios. Two distinct datasets were employed for this purpose: a publicly available dataset containing finger movements, and our own dataset containing hand gestures performed in an uncontrolled environment, containing spontaneous and unpredictable motor disturbances. These conditions inherently resulted in increased signal variability, including artifacts induced by muscle activity, variations in attention, and inconsistencies in gesture execution.

Despite the differences in recording environments and the greater susceptibility to spontaneous artifacts in the second dataset, the proposed method consistently maintained high classification performance. This outcome demonstrates

its strong resilience to real-world EEG noise sources and confirms the model's capability to extract meaningful inter-regional neural information, even when data quality is compromised. Furthermore, the method was evaluated under only minimal preprocessing, indicating its robustness to non-ideal input conditions and its suitability for application in naturalistic or clinical settings where extensive preprocessing pipelines may not be feasible, and the system might be subject to higher signal noise.

The comparative evaluation between the two datasets also highlighted the method's adaptability to varying levels of motor task complexity. While both datasets involved fine motor movements, the continuous hand opening and closing task presents greater complexity than easier to isolate finger movements. This increased complexity arises from the need to coordinate multiple muscle groups simultaneously, the temporal continuity of the gesture, the lower task synchrony across trials, and the higher degree of intra-task similarity, which makes discriminating between movement segments more challenging. Such gestures, characterised by overlapping or gradually evolving EEG patterns, introduce subtle noise that can interfere with accurate classification.

Even so, the proposed method demonstrated consistent performance even under these challenging conditions. This contrasts with traditional approaches that typically perform better on more discrete or well-separated motor actions, such as binary hand open and close decisions. These results confirm the method's robustness not only to inter-subject variability, but also to significant signal degradation and increased motor complexity. Its performance across datasets recorded under controlled and uncontrolled conditions, and across tasks with varying levels of temporal and neuromuscular complexity, even under only minimal signal preprocessing, illustrates its potential for generalisation.

The small size of datasets used to validate the proposed method further proves its suitability to be implemented in a real-world scenario. In such cases, access to extensive human data is not easy and the recording time is highly limited. These are still the main challenges in BCI [111]. The proposed method paves the path for a new BCI generation where the system can learn a movement without requiring any audio, visual, or haptic feedback.

To further facilitate the potential implementation of the BCIs presented in

Chapter 4 as well as this chapter in uncontrolled environments, the following chapter introduces a novel filtering technique specifically designed for hyperscanning data. This proposed filtering technique is based on the reformulation of the CSP method. Its effectiveness for uncontrolled environments is evaluated through the use of the dataset described in Chapter 3.

Chapter 6

Formulation of Common Spatial Patterns for Multi-task Hyperscanning BCI

This chapter is an expansion of our published work in [176].

Exploiting the data from multiple subjects performing the same task has great potential for more effective training of motor-related BCI systems in an uncontrolled environment. This is mainly due to the brain engagement in performing irrelevant mental or physical activities while performing a BCI task. Therefore, an approach that extracts the common, or relevant, brain motor activity from multiple brains and avoids the effect of undesired or irrelevant ones for training the BCI system can be a significant outcome of multi-subject, or hyperscanning, BCI.

Over the years, many algorithms have been developed and improved for analysis of EEG data, such as the feature extraction techniques discussed in [177]. However, yet no feature extraction technique tailored to hyperscanning data has been developed.

In hyperscanning settings, multiple subjects perform at least one task together at the same time. This requires a level of cognitive involvement between the subjects. As discussed in previous studies [54, 109, 178], this interaction also results in an enhanced synchronisation between the subjects' neural mechanisms,

and some additional connections and relationships between different brain regions for all the interacting subjects. This information therefore, is not present in traditional, non-hyperscanning settings.

As discussed in Chapters 1 and 3, the majority of EEG motor-related studies have been performed under ideal scenarios, where the data is recorded in a controlled environment and the subjects are asked to constantly concentrate on a single motor-related task. This poses some challenges in the implementation of these systems in an uncontrolled environment where the brain is inherently engaged in multiple tasks. In most of these cases, the available dataset size is small due to various practical limitations.

Through the use of additional brain connections between the subjects generated in hyperscanning settings, we can develop a system for non-ideal data to establish a BCI that can perform favourably even in uncontrolled environments. Specifically, we propose the development of a system that enables classification of EEG hyperscanning data depending on whether the subjects are performing common tasks or not. Furthermore, the proposed system can classify motor-related multi-task data where the tasks are all in common, even in the presence of strong undesired tasks, inherently or intentionally performed by the subjects.

To achieve these goals, a new filtering technique based on the reformulation of CSP [179] catered for hyperscanning data is introduced. The proposed filtering technique can obtain satisfactory performance even when using a smaller dataset than what is used in related studies [180]. The proposed system allows for recording under non-ideal conditions, levering the potential of this technique to be implemented in a real-world BCI. This new technique is proposed for a scenario where the brain signals of at least two subjects are recorded simultaneously during the performance of common and uncommon tasks in a multi-task scenario.

6.1 Common Spatial Patterns; Concept and Extensions

CSP separates a multivariate signal into additive subcomponents which have maximum differences in variance. It was first introduced in 1990 as a novel technique to discriminate between the EEGs of two populations [179] through the optimisation of the variance separation process for the two classes. This is obtained through the projection of the EEG signals to a set of spatial filters. This way, it maximises the variance of one state while minimising the variance of the other state, enabling the extraction of discriminative features.

Therefore, given $\boldsymbol{X} \in \mathbb{R}^{N_k \times L}$ as a segment of EEG signal, where N_k represents the number of channels and L the number of samples, the traditional CSP cost function is formulated as:

$$J(\boldsymbol{W}) = \frac{\boldsymbol{W}^T \boldsymbol{X}_1^T \boldsymbol{X}_1 \boldsymbol{W}}{\boldsymbol{W}^T \boldsymbol{X}_2^T \boldsymbol{X}_2 \boldsymbol{W}} = \frac{\boldsymbol{W}^T \boldsymbol{C}_1 \boldsymbol{W}}{\boldsymbol{W}^T \boldsymbol{C}_2 \boldsymbol{W}} \tag{6.1}$$

where $\boldsymbol{W}$ represents the filter coefficients that separate two classes of a brain activity, or the two populations, and $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ represent the multichannel signals from classes 1 and 2 respectively. In the classical CSP, $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$ refer respectively to the covariance matrices for the subspaces of the signals from tasks, or states, 1 and 2 or those of the desired or undesired signals.

6.1.1 CSP Applications to EEG and BCI

Since its development, CSP has been widely used for the extraction of feature matrices and the analysis of EEGs particularly for efficient BCI development. It has been successfully used for the separation of different mental states, including cognitive state estimation, for Parkinson's [181] or epilepsy [182, 183], as well as for emotion recognition [184,185] or MI classification, introduced for the first time in 2000 for imagery hand movement [186].

Even so, the original formulation presents some limitations, including sensitivity to noise, non-stationarity, and overfitting in high-dimensional data. To overcome these limitations, several extensions have been proposed to accommo-

date the original formulation to multiple problem domains [187]. As reviewed in [188], multiple regularisation techniques have been developed to mitigate the overfitting problem of CSP and improve its generalisation. In problems where relevant features are distributed in multiple frequency bands, variations to CSP have been introduced to capture the frequency-specific information, such as the widely known and used filter bank CSP (FB-CSP) [189], its regularised variant [190], and other proposed frequency based extensions [191]. Furthermore, some CSP extensions, such as [192,193], incorporated sparsity constraints to select the most relevant spatial filters and reduce the dimensionality of the data.

Some CSP extensions have also been proposed to overcome inter-subject variability problems, such as composite CSP (CCSP) [194], or composite local temporal correlation CSP (CLTCCSP) [195], where CSP is modified to include information from different subjects. This is particularly important in BCI, which has become the main application of CSP and its variants.

The use of CSP for BCI incorporates further challenges, such as the necessity for a shorter training time and a higher inter-subject robustness compared to the use of CSP in other applications. For this reason, some CSP extensions have been developed focusing on these challenges specifically for BCI settings and MI classification scenarios. For example, in [196], the authors propose an extension of the CSP formulation based on an Euclidean alignment together with transfer learning to overcome these challenges.

Finally, another main limitation of the original CSP formulation is that it is inherently binary, not allowing for the simultaneous discrimination of multiple classes. This is especially relevant in its application to BCI, where multi-task scenarios are highly common. Nevertheless, extensions like [197–199] have been developed to allow CSP to successfully handle multi-task classification problems.

6.2 Problem Formulation for Hyperscanning CSP

In this section, we propose a novel CSP-based spatial filtering approach as a feature extraction technique to support the classification of a common task performed by at least two subjects, independently of what the specific task is. The proposed hyperscanning CSP (hyperCSP) technique is a reformulation of the

well-known CSP method.

6.2.1 Hyperscanning Common Spatial Patterns

The proposed hyperCSP aims to learn the spatial filters that maximise the variance of the EEG signals when a common task is performed by multiple subjects, while minimising the variance for the uncommon tasks. In this way, the cost function for hyperCSP can be approximated as:

$$J(\boldsymbol{W}) = \frac{\boldsymbol{W}^T\left(\sum_{z=1}^{K}\boldsymbol{C}_z\right)\boldsymbol{W}}{\boldsymbol{W}^T\left(\sum_{b=1}^{K}\sum_{a=1,a\neq b}^{K}\boldsymbol{C}^g_{a,b}\right)\boldsymbol{W}} \tag{6.2}$$

In this equation, the global correlation $\boldsymbol{C}^g_{1,2,\ldots K}$ has been approximated by the sum of mutual correlations, i.e. $\boldsymbol{C}_{a,b}$.

Since $\boldsymbol{C}_{a,b} = \boldsymbol{C}_{b,a}$, eq. (6.2) can be simplified to:

$$J(\boldsymbol{W}) = \frac{\boldsymbol{W}^T\left(\sum_{z=1}^{K}\boldsymbol{C}_z\right)\boldsymbol{W}}{\boldsymbol{W}^T\left(\sum_{b=1}^{K-1}\sum_{a=b+1}^{K}\boldsymbol{C}_{a,b}\right)\boldsymbol{W}} \tag{6.3}$$

where K in this scenario is the number of subjects, $\boldsymbol{W}$ is the spatial filter, $\boldsymbol{C}_z$ is the spatial covariance matrix for subject z, and $\boldsymbol{C}_{a,b}$ is the mutual correlation matrix of subjects a and b.

The use of multiple heads in EEG hyperscanning could enhance the classification accuracy, especially in uncontrolled environments where the subjects are free to perform undesired or irrelevant tasks. Here, due to resource limitations, we only work with two subjects. Therefore, the cost function for hyperCSP in (6.2) becomes:

$$J(\boldsymbol{W}) = \frac{\boldsymbol{W}^T(\boldsymbol{C}_1 + \boldsymbol{C}_2)\boldsymbol{W}}{\boldsymbol{W}^T\boldsymbol{C}_{1,2}\boldsymbol{W}} \tag{6.4}$$

where $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$ represent the spatial covariance matrices for Subjects 1 and 2 respectively, estimated as:

$$\boldsymbol{C}_z = \frac{\boldsymbol{X}_z\boldsymbol{X}_z^T}{\text{trace}(\boldsymbol{X}_z\boldsymbol{X}_z^T)} \tag{6.5}$$

$\boldsymbol{X}_z = [\boldsymbol{x}_{z,1}, \boldsymbol{x}_{z,2}, ...\boldsymbol{x}_{z,L}]$ represents a trial of EEG signal for subject z, and L denotes the trial length. $\boldsymbol{X}_z^T$ represents the transpose of $\boldsymbol{X}_z$ and trace($\boldsymbol{C}$) is the sum of the diagonal elements of matrix $\boldsymbol{C}$. On the other hand, in (6.4), $\boldsymbol{C}_{1,2}$ represents the linear mutual correlation matrix between the two subjects, defined as:

$$\boldsymbol{C}_{a,b} = \frac{\boldsymbol{X}_a \boldsymbol{X}_b^T}{\text{trace}(\boldsymbol{X}_a \boldsymbol{X}_b^T)} \tag{6.6}$$

where, in this scenario, $\boldsymbol{X}_a$ and $\boldsymbol{X}_b$ refer respectively to the EEG signals of subjects a and b. Here, it is assumed that the signals have zero mean.

6.2.2 Solution to the hyperCSP Problem

To solve the maximisation problem in (6.4), we use the Lagrange multiplier method, as shown in [198]. Following this method, the constrained problem in (6.4) is converted to an unconstrained problem:

$$\hat{J}_{\lambda,\boldsymbol{w}} = \boldsymbol{w}^T(\boldsymbol{C}_1 + \boldsymbol{C}_2)\boldsymbol{w} - \lambda(\boldsymbol{w}^T\boldsymbol{C}_{1,2}\boldsymbol{w} - 1) \tag{6.7}$$

where λ is the Lagrange multiplier. Then, to calculate the spatial filter $\boldsymbol{w}$, the derivative of $\hat{J}_{\lambda,\boldsymbol{w}}$ is taken and set equal to zero with respect to $\boldsymbol{w}$:

$$\begin{aligned} \frac{\partial \hat{J}_{\lambda,\boldsymbol{w}}}{\partial \boldsymbol{w}} = 2\boldsymbol{w}^T(\boldsymbol{C}_1 + \boldsymbol{C}_2) - 2\lambda\boldsymbol{w}^T\boldsymbol{C}_{1,2} = 0; \\ (\boldsymbol{C}_1 + \boldsymbol{C}_2)\boldsymbol{w} = \lambda\boldsymbol{C}_{1,2}\boldsymbol{w}; \\ \boldsymbol{C}_{1,2}^{-1}(\boldsymbol{C}_1 + \boldsymbol{C}_2)\boldsymbol{w} = \lambda\boldsymbol{w} \end{aligned} \tag{6.8}$$

This leads to the eigenvalue problem. Therefore, we obtain the eigenvectors of $\boldsymbol{\Upsilon} = \boldsymbol{C}_{1,2}^{-1}(\boldsymbol{C}_1 + \boldsymbol{C}_2)$, and the spatial filter $\boldsymbol{w}$ is estimated as the row of $\boldsymbol{\Upsilon}$ eigenvector matrix corresponding to its largest eigenvalue. Such a filter also alleviates the effects of other undesired brain sources and artifacts.

6.3 Application of hyperCSP

In this section, we introduce the steps performed in the experiments to process and analyse the raw EEG hyperscanning data using hyperCSP. Before applying the method, the data are preprocessed to mitigate the effects of some EEG artifacts. The proposed hyperCSP is used to obtain the feature vectors that are given to the classifier. To classify the spatial feature matrix containing all the feature vectors, we employ a number of well-known classifiers.

The EEG hyperscanning data used in this experiment are the data from experiments 1, 2, 3, 4 and 5, described in Chapter 3, Section 3.3.2.

6.3.1 Data Preprocessing

The raw EEG hyperscanning data is preprocessed to remove any bad channels and main artifacts.

The data from both subjects are combined into a single data block $\boldsymbol{X}_{a+b}$ in the form of a $2N_k \times L$ matrix, where L represents the signal length. The first N_k channels are the EEG signals from Subject 1 (a) while the other N_k channels are the EEG signals from Subject 2 (b). Then, $\boldsymbol{X}_{a+b}$ is segmented into equal-length blocks based on the labels associated to the EEG hyperscanning signals. In this way, we obtain EEG hyperscanning blocks in the form of a $2N_k \times L_s$ matrix, where L_s represents the segment length with a single label, either Task 1 or Task 2 activity, both potentially including uncommon (undesired) tasks.

As in previous chapters, channel 27 from the first subject and channel 32 from the second subject (both distant from the motor area) present a low electrode-skin impedance value across most of the recording sessions, so they are considered bad channels. Therefore, these two channels are removed from both subjects to maintain the symmetry between the two subjects' data. The locations of these channels are coloured grey in Fig. 6.1. A FIR bandpass filter of 1-50 Hz is applied to the data from both subjects to remove the baseline and reduce any possible power line noise. Then, average re-referencing is applied to each subject's data separately. These preprocessing methods are applied with the help of EEGLAB Toolbox [200]. No other artifact removal procedure is applied to the data since the proposed hyperCSP method is also able to filter these undesired artifacts.

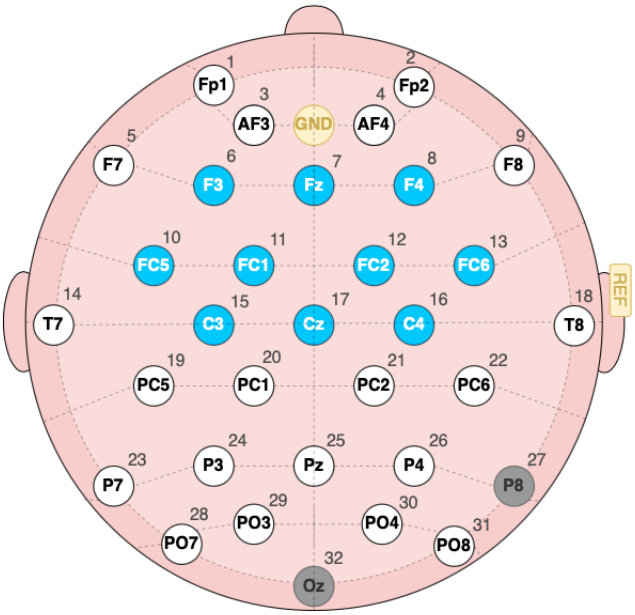

Figure 6.1: The electrode setup used for the experiment. The locations of the electrodes follow the standard 10-20 international EEG electrode placement system for 32 electrodes. The highlighted electrodes in blue form a subsystem that represents the motor area. The highlighted electrodes in grey represent channels 27 and 32 that have been removed during the preprocessing stage.

6.3.2 hyperCSP as Feature Extraction Technique

Unlike many CSP-based BCI applications, here the intention is to classify the entire body movement rather than the movement onset. For that reason, once the signals from both subjects are preprocessed, we select an appropriate window size over which the signal is considered stationary and slide the window over all the signal segments. In this case, we selected a window size between 2.7s and 3.3s, depending on the length of each segment, with no overlap. For each sliding window, we apply Eq. (6.4)-(6.6) using the first N_k channels from subject a as $\boldsymbol{X}_a$ and the second N_k channels from subject b as $\boldsymbol{X}_b$. We obtain the right eigenvectors that leads to the best $\boldsymbol{W}$, as the desired spatial filters. Then, we sort $\boldsymbol{W}$ in descending order of the eigenvalues and apply the filters $\boldsymbol{W}$ corresponding to the largest eigenvalues to the signal window segment following the equations:

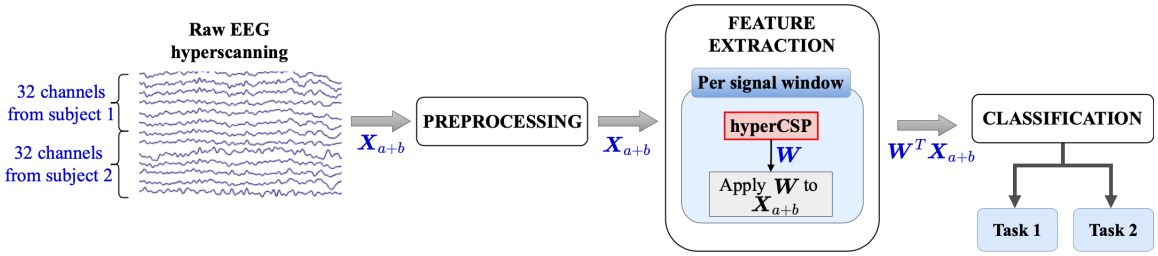

Figure 6.2: A step-by-step illustration of the proposed method. The classification of the tasks is done in the presence of strong uncommon tasks.

$$\boldsymbol{X}'_a = \boldsymbol{W}^T \boldsymbol{X}_a \tag{6.9}$$

$$\boldsymbol{X}'_b = \boldsymbol{W}^T \boldsymbol{X}_b \tag{6.10}$$

Given the projected signals, we apply a log-variance method similar to those in [198], [201], and [202] to extract the feature vector from the projected signals. Hence, we calculate the logarithm of the variances for the projected EEG window segment of each subject and normalise them to have zero mean and unit variance:

$$\boldsymbol{f}_a = \log\left(\frac{\text{var}(\boldsymbol{X}'_a)}{\sum \text{var}(\boldsymbol{X}'_a)}\right) \tag{6.11}$$

$$\boldsymbol{f}_b = \log\left(\frac{\text{var}(\boldsymbol{X}'_b)}{\sum \text{var}(\boldsymbol{X}'_b)}\right) \tag{6.12}$$

Once the feature vectors for all the signal segments are obtained, they are combined into a feature matrix, obtaining a $N_k \times S$ matrix, where S represents the number of segments for both subjects. This feature matrix is the input to the chosen classifiers. Fig. 6.2 summarises the steps defined in this section, while Algorithm 3 represents the calculation of the feature matrix for each segment in pseudo code.

6.3.3 Classification

Once the feature matrix is obtained, a classifier is selected and trained to classify the tasks performed by individuals.

Algorithm 3 Algorithm for hyperCSP

Input: EEG Segment (**EEG**: $2N_k \times L_s$)

Prerequisites: Window Length L_w, Initial signal point i

while $L_w + i - 1 \leq$Segment Length (L_s) **do**
  $\boldsymbol{X}_a = \boldsymbol{EEG}[1 : N_k, i : i + L_w - 1]$
  $\boldsymbol{X}_b = \boldsymbol{EEG}[N_k + 1 : 2N_k, i : i + L_w - 1]$
  $\boldsymbol{C}_1 = \frac{\boldsymbol{X}_a \cdot \boldsymbol{X}_a^T}{\text{trace}(\boldsymbol{X}_a \cdot \boldsymbol{X}_a^T)} + \frac{\boldsymbol{X}_b \cdot \boldsymbol{X}_b^T}{\text{trace}(\boldsymbol{X}_b \cdot \boldsymbol{X}_b^T)}$
  $\boldsymbol{C}_2 = \frac{\boldsymbol{X}_a \cdot \boldsymbol{X}_b^T}{\text{trace}(\boldsymbol{X}_a \cdot \boldsymbol{X}_b^T)}$
  $\boldsymbol{W} = \text{eigenvalues}(\boldsymbol{C}_1, \boldsymbol{C}_2)$
  Sort $\boldsymbol{W}$ in descending order
  $\boldsymbol{X}'_a = \boldsymbol{W}^T \cdot \boldsymbol{X}_a$
  $\boldsymbol{X}'_b = \boldsymbol{W}^T \cdot \boldsymbol{X}_b$
  $\boldsymbol{f}_a = \text{zero mean } \log(\text{var}(\boldsymbol{X}'_a)/\sum \text{var}(\boldsymbol{X}'_a))$
  $\boldsymbol{f}_b = \text{zero mean } \log(\text{var}(\boldsymbol{X}'_b)/\sum \text{var}(\boldsymbol{X}'_b))$
  $i = i + L_w$
end while

Output: Normalised Feature Vectors: $\boldsymbol{f}_a, \boldsymbol{f}_b$

To better validate the performance of the proposed hyperCSP method, we applied three well-known classifiers that have shown promising results in previous CSP applications [197]. The chosen classifiers are linear kernel SVM, LDA, and KNN with 3 neighbours. The number of neighbours was established through the hypertunning of the KNN model.

The obtained feature matrix is divided into three subsets, one containing 75% for training, another containing 15% for validation, and one for testing, containing the remaining 10% of the feature matrix. In this scenario, due to the small size of the feature matrix, we used a 10-fold cross-validation for the evaluation of the models.

6.3.4 Online EEG hyperscanning-based BCI system

The use of the proposed hyperCSP as a filtering technique becomes especially important in EEG Hyperscanning-based BCIs, where a long-duration common task can be isolated. In this scenario, hyperCSP provides cleaner data, which helps improve the performance of other feature extraction and classification methods,

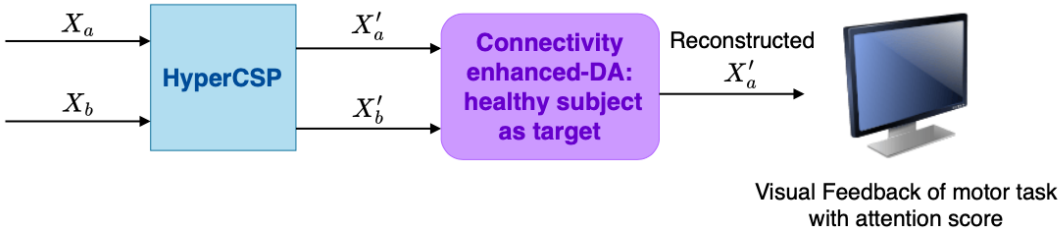

Figure 6.3: A schematic illustration of the proposed online BCI system. The system incorporates the proposed hyperCSP method to filter the data and provide cleaner data to the connectivity-informed DA method, described in Chapter 4. This online BCI system can detect if a subject is following the motor activity of another subject based on their neural activity. X_a represents the EEG signals from the following subject (individual under rehabilitation), while X_b represents the EEG signals from the leading subject (healthy individual). X'_a and X'_b represent the filtered signals of X_a and X_b respectively, which have been filtered with hyperCSP to remove undesired mental activity.

and therefore, improves the performance of the final BCI system.

Given this, it can be presumed that the other analysis and recognition techniques for EEG hyperscanning and complex continuous movements, which were introduced in Chapters 4 and 5 respectively, could further be improved through the application of hyperCSP as part of the preprocessing steps of the systems.

Therefore, the proposed hyperCSP and the connectivity-informed DA method, defined in Chapter 4, can be combined for the development of an online attention evaluation hyperscanning BCI system, as shown in Fig. 6.3. Such a system can detect if two subjects are performing the same task or not. This system can further be extended to include multiple subjects instead of only two. In that case, only one subject leads while the remaining subjects, one or more, follow. This online system can be especially beneficial during neurorehabilitation sessions to control if the follower subjects, i.e., those under rehabilitation, perform the tasks correctly or get distracted. Providing feedback on the concentration of the subjects during the neurorehabilitation sessions is essential to detect when specific tasks may need modification.

For this system, the attention score was obtained based on obtained aggregated error per electrode between the follower and the leader subjects. Therefore, given Eq. (4.10), which provides the error per node, we obtain the aggregated

error as:

$$E_{aggregate} = \frac{1}{N_k} \sum_{k=1}^{N_k} E_k \tag{6.13}$$

which provides a measure on the similarity between the EEGs from both subjects. Given this, we can estimate the attention score as:

$$\text{Attention Score} = 1 - E_{norm} \tag{6.14}$$

where E_{norm} is the normalised $E_{aggregate}$ between [0 1]. For a more accurate system, E_{norm} can be calibrated to be normalised between the $E_{aggregate}$ obtained during full synchronisation and no synchronisation between the subjects.

Furthermore, given the BCI system proposed in Fig. 6.3, visual feedback in the form of simple instructions, such as different colours representing the different attention scores could be provided. This way the system is able to provide real-time feedback that can guide not only the clinicians, but also the individual under rehabilitation to inform them if the mental activity is not being followed correctly. This same feedback can be used as control output signals when necessary depending on the final purpose of the system.

The proposed BCI system was evaluated through the implementation of the system in Simulink as an online real-time BCI system, as shown in Fig. 6.4. For the evaluation of the system, the same data described in this chapter in Section 6.3 was used. The pre-recorded data was passed to the system as sequential time series to emulate real-time online data for a more realistic evaluation of the system. The same minimal preprocessing steps described in Section 6.3.1 were also applied together with hyperCSP as preprocessing steps.

hyperCSP as well as the connectivity-informed DA method require an EEG window segment to be passed instead of each signal time instant. For this reason, the implemented system contains a buffer that enables the collection of the data over each window before passing the data to hyperCSP. The window size is adjustable and can be modified as needed depending on the nature of the data. This reliance of the proposed methods on window segments generates a small delay on the system, preventing it from been a fully real-time BCI system. Still,

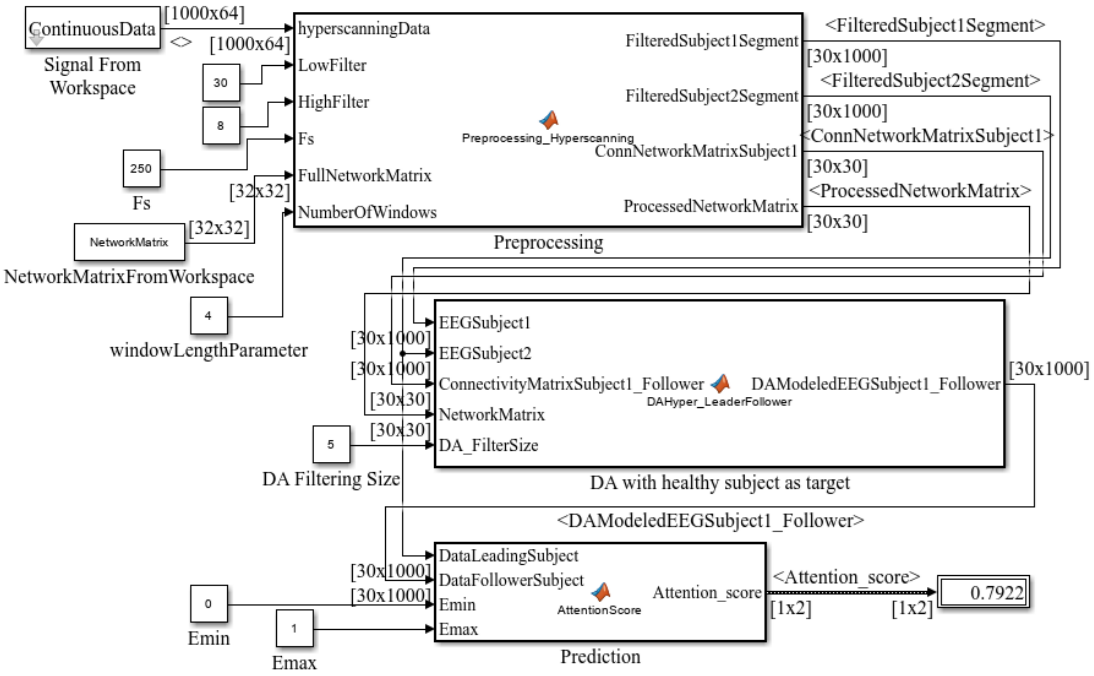

Figure 6.4: An illustration of the implemented online BCI system, implemented in Simulink. Here, *Fs* represents the EEG sampling rate, and *ConnNetworkMatrixSubject1*, represents the estimated connectivity measures from Subject 1.

the use of window segments that incur a small delay on the real-time system is a common and acceptable practice, usually necessary in these BCI systems. Even so, through optimisation processes and reuse of the same window segment for both the preprocessing and feature extraction methods, we are able to reduce the delay of the system to a minimum acceptable delay that does not undermine the final purpose of the system. The final delay of the system is mostly determined by the window size.

The proposed system can also be implemented together with the DAOT system, described in Chapter 5, after the reconstruction of X'_a for further improvement.

6.4 Results

Given the dataset described in Section 3.3.2, we present and analyse the results obtained after applying the proposed hyperCSP and the traditional CSP under two scenarios: a multitask scenario through the separation and classification of two common tasks, left and right hand movement; and a filtering scenario, through the separation of desired tasks (tasks in common) from undesired tasks (tasks not in common) and their classification.

6.4.1 Multi-task Separation and Classification

For the classification of common multi-tasks performed by two subjects in the presence of undesired tasks, a subset of the dataset described in Section 3.3.2 is used. Only experiments 1 and 2 were considered where the two subjects performed two different common tasks (left- and right-hand movement) and each performed a different (uncommon) task. Therefore, this subset of the dataset contains 8 trials. In here, we show the results obtained when applying hyperCSP and classifying the two common tasks: move right hand and move left hand, all in the presence of strong undesired tasks.

To better validate the results, using the same dataset, we compare the results of the proposed method against results obtained using the traditional CSP, as defined in Eq. (6.1), using the data from one subject from each pair of subjects. This can show how a movement classification using EEG can be affected by undesired brain activities.

Fig. 6.5 shows the topoplots from Subject 1 for the first two trials of left-hand movement before and after applying hyperCSP and the comparison system. Since we are focusing on motor movements, the topoplots are represented for alpha and beta band frequencies (also known as alpha-mu and beta-mu). In Fig. 6.5a, the topoplots before applying any CSP filtering are presented, and it can be appreciated that there is a mix of sources due to the undesired tasks performed by the subject while moving the left hand. In Fig. 6.5b, c, the topoplots after applying the filtering are presented, and it can be appreciated that hyperCSP is able to isolate the common task and obtain a clearer and more isolated alpha-mu and beta-mu. On the other hand, due to the presence of different sources, CSP

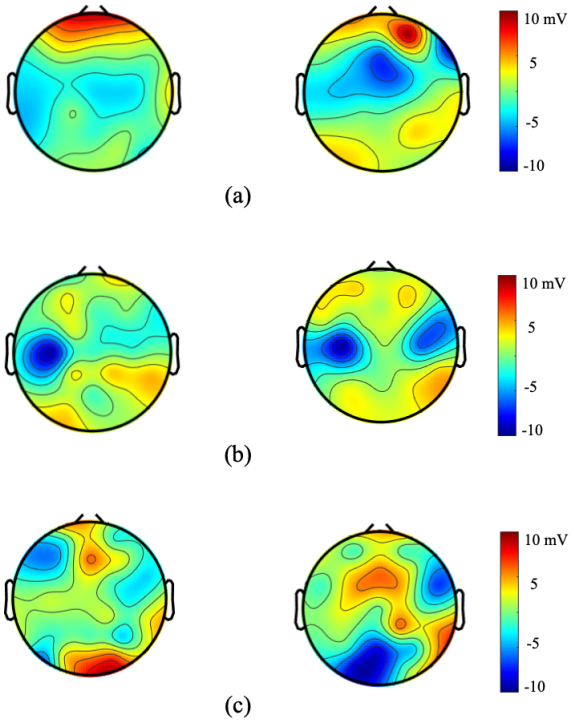

Figure 6.5: The topoplots from Subject 1 for the left-hand movement after applying CSP and hyperCSP to isolate the desired common task; (a) before any CSP filtering, (b) after applying CSP, and (c) after applying hyperCSP. In the left the activity in alpha band, and in the right the activity in beta band, have been demonstrated. The main conclusion is that the hyperCSP better isolates the brain's common and desired prolonged activity.

is not able to clearly isolate the common task.

Table 6.1 shows the accuracy obtained for each of the mentioned methods used to validate the proposed system during the testing and training phases for each classifier. The results are the average validation accuracy and the standard deviation of each system obtained after 10-fold cross-validation. A maximum validation accuracy for each system, as well as the testing accuracy and the F-score for the two classes, are obtained using the classifiers with the lowest validation error.

As explained in Section 6.3.3, three classifiers are used. For consistency in the results, the parameters for the classifiers remained the same for the proposed and traditional methods.

Table 6.1: Classification accuracy during validation and testing of hyperCSP and the comparison systems during the classification of common multi-tasks.

	hyperCSP			CSP		
	SVM	LDA	KNN	SVM	LDA	KNN
Avg. Validation Acc	0.63±0.08	0.60 ± 0.08	0.50 ± 0.08	0.44 ± 0.07	0.44 ± 0.07	0.45 ± 0.07
Max. Validation Acc	0.76	0.74	0.71	0.59	0.59	0.59
Testing Acc	0.82	0.73	0.45	0.64	0.64	0.36
Testing F-score	0.82	0.73	0.45	0.61	0.61	0.34

As shown in Table 6.1, for the classification of right and left hand movements, SVM results in the best performance for hyperCSP. We obtained an average validation accuracy of 62.77%, with a maximum validation accuracy of 76.47%, a testing accuracy of 81.82%, and a F-score of 0.82 for this model. On the other hand, for the traditional formulation of CSP, the three classifiers yield a highly similar performance. Still, SVM and LDA provides the best performance for the traditional CSP, obtaining the highest F-score. Even so, the higher performance of the proposed method can be verified based on these values, especially on the F-score, which clearly shows the merit of hyperCSP for robust BCI compared to the traditional CSP formulation.

6.4.2 Separation of Target tasks from Irrelevant Brain Activities and their Classification

Now, we show that hyperCSP, similar to the traditional CSP, can separate a relevant (desired task), which is in common between the participants, from irrelevant (undesired) tasks.

For the classification of desired (common) and undesired (uncommon) tasks for the two subjects, experiments 1, 2, 3, 4, and 5 from the dataset described in Section 3.3.2, is used. This subset of the dataset comprises 14 trials in total. For this experiment, we followed the same preprocessing steps and applied the same parameters as in Section 6.4.1.

In Table 6.2, for both methods, the KNN classifier results in the best perfor-

mance, even so the three models' performance is highly similar. The proposed method with the KNN classifier obtained an average validation accuracy of 64%, with a maximum validation accuracy of 70.97%, and a testing accuracy of 68.29%. This shows that the proposed system is able to distinguish between tasks performed simultaneously in the presence of undesired (or random) individual tasks without any prior information about the random tasks, even with a small dataset. In comparison, the classical CSP obtained an average validation accuracy of 47%, with a maximum validation accuracy of 57%, and a testing accuracy of 43%. This shows that the competing method, CSP, is not able to deal with separating the tasks in the presence of strong random or undesired tasks, while hyperCSP can achieve a reasonable accuracy for such scenarios.

Table 6.2: Classification accuracy during validation and testing of hyperCSP and the comparison systems during the classification of desired and undesired tasks.

	hyperCSP			CSP		
	SVM	LDA	KNN	SVM	LDA	KNN
Avg. Validation Acc	0.53 ± 0.03	0.52 ± 0.03	0.64 ± 0.03	0.45 ± 0.05	0.45 ± 0.05	0.47 ± 0.05
Max. Validation Acc	0.58	0.59	0.71	0.57	0.53	0.57
Testing Acc	0.71	0.68	0.68	0.48	0.57	0.43
Testing F-score	0.65	0.63	0.67	0.45	0.56	0.43

The results obtained in both experiments show that, given the same dataset, CSP is not able to distinguish between two tasks performed simultaneously in the presence of an undesired (or random) individual task without any prior information about the uncommon tasks performed individually.

Therefore, we can conclude that hyperCSP outperforms the popular and well-established feature detection method, the traditional CSP formulation, mainly because it benefits from more than one subject's EEG to marginalise the uncommon, unrelated, or spurious tasks. Due to practical limitations, in this study we only worked with two subjects simultaneously. We expect that, by increasing the number of subjects in the hyperscanning setting, the training error will decrease. In addition, hyperCSP may also be regularised to better incorporate the task properties and provide a better performance for the proposed system.

6.5 Conclusions

In this chapter, we proposed a novel feature extraction technique based on CSP, namely hyperCSP, that can be used to analyse EEG hyperscanning data and classify different brain motor tasks. This is applicable to a synchronised multi-brain study, where the subjects perform a common task in the presence of strong undesired tasks, similar to what can be obtained in a real-world, uncontrolled environment. These uncontrolled environments are commonly characterised by the presence of varying lighting conditions, unpredicted stimuli, involuntarily movements, and spontaneous undesired neural activity, which introduce substantial noise and artifact contamination in the EEG signals.

The results from this study show that hyperCSP can achieve a satisfactory classification accuracy of 81.82% using an SVM classifier mainly because it benefits from the synchronised data from two subjects during a common movement task performance. The hyperCSP filter, $\boldsymbol{W}$, can be applied to an individual's EEG to highlight the desired task and suppress all other undesired tasks. Furthermore, this approach does not require prior heavy preprocessing of the signal, avoiding the need for manual artifact removal, making the method highly suitable for its implementation in more dynamic, real-world settings, where such noise is unavoidable and difficult to remove.

To validate the robustness of the proposed method under various noise conditions and task complexities, a dataset containing multiple distinct motor tasks performed under uncontrolled conditions was employed. This dataset contained diverse desired, and in common between the subjects, motor tasks as well as spontaneous and sporadic motor activity, different for each subject. In real-world, uncontrolled scenarios, EEG recordings are often contaminated by involuntary neural activity resulting from involuntary movements, fluctuations in attention, and variations in task execution timing. The consistent performance of hyperCSP on such diverse and noisy data provides strong empirical support for its resilience to both signal noise and task variability.

Moreover, through its evaluation with this dataset, the proposed method demonstrated its adaptability to diverse tasks with varying levels of motor complexity. Tasks involving fine motor control and temporal continuity typically

generate overlapping EEG patterns that are difficult to differentiate using traditional CSP. However, hyperCSP showed an enhanced ability to isolate the neural activity from the desired tasks, even when the desired movements were embedded within or followed by spontaneous motor noise. This robustness indicates that the method is capable of generalising across a range of motor behaviours, from simple hand movements to continuous and more complex gestures. The results show that hyperCSP can therefore improve the performance of a BCI.

While this study focused on hyperscanning involving two participants, the proposed method can be extended to scenarios involving a larger number of subjects. In such scenarios, the common task across all participants would define the desired task, while all other neural activity would be considered as the undesired tasks and attenuated by hyperCSP. However, this extended setting introduces some challenges, including increased computational complexity, task synchronisation difficulties, and higher inter-subject variability. These factors can affect the reliable extraction of common neural patterns, which could impact the robustness of the proposed method.

Various regularisation techniques proposed for the conventional CSP by different researchers, such as in [180] and [188], can also be extended to hyperCSP to further improve its performance. Finally, with a larger dataset, the application of deep learning classification models, such as DNNs could be considered in future works, which may enhance the system performance, as shown in other studies [203, 204].

The next chapter presents the thesis conclusions, containing a summary of the key findings presented in this thesis, their potential limitations, and a discussion on potential directions for future work.

Chapter 7

Conclusions and Future Research

In this thesis, a new direction in EEG-based BCI has been established. In this development, multi-subject data through hyperscanning have been analysed.

Exploiting the data from multiple subjects performing the same task has great potential for more effective training of motor-related BCI systems in an uncontrolled environment. This is mainly due to the brain engagement in performing irrelevant mental or physical activities while performing the desired BCI task. Although over the years several signal processing and machine learning techniques have been developed and improved for the analysis of EEG data and the implementation of BCIs, none has been tailored to hyperscanning data.

In this thesis, we explored the possibility of implementing novel EEG hyperscanning BCI systems highly applicable to neurorehabilitation by exploiting the cooperation parameters. More specifically, we developed three new analysis methods, two based on the single task DA approach, and one as a reformulation of the classical CSP especially tailored to hyperscanning data. Furthermore, complex motor multi-task EEG hyperscanning data have been released to help promote research in this field. This work opens up a new research direction for the development and implementation of BCIs in uncontrolled environments. In this chapter, we conclude the thesis with an overview of the main contributions, discuss possible limitations of the proposed systems, and propose new possible directions for future investigations.

The contributions of this work can be summarised as:

- Generating the first publicly available multi-task motor EEG hyperscanning dataset, which contains data from multiple subjects recorded in an uncontrolled environment [108]. Main challenges during the design paradigm and data recording of hyperscanning scenarios are also reviewed and discussed, and the guidelines on how to overcome them are provided.

- A new EEG hyperscanning analysis method has been proposed which can be used to evaluate the state of a patient based on how well the patient can follow the motor tasks performed by another individual. This technique is based on improving the single task DA approach by brain connectivity analysis as the estimates for combination weights of the DA. In this approach, the brain network of one subject and a target channel of the second subject are used to model the pathway between the two subjects.

- A new classification technique is developed for the recognition of prolonged and fine brain motor potentials. This method allows for a realistic and accurate recognition of motor movements with no audio, visual, or haptic feedback. This provides a new platform for the development of more advanced prosthetic systems and BCIs. This new method makes use of the previously developed connectivity-informed DA approach, where in this case, the objective of the distributed network is represented by a member of an orthogonal functions dictionary, each one associated with a specific sub-gesture.

- Finally, a new formulation for CSP, namely hyperCSP, especially tailored to hyperscanning data is developed. This method can be used to filter undesired tasks from EEGs, providing cleaner data, and improving the final performance of the BCI. The implementation of this technique leverages the BCI implementation in a more realistic and uncontrolled environment, where the subjects can be distracted.

Moreover, a pipeline for real-time multi-subject BCI has been proposed with the possibility of combining hyperCSP and other proposed techniques. This shows the feasibility of the proposed methods to be implemented as an online system, showcasing their potential deployment in real-world scenarios.

7.1 Limitations

While the proposed methods presented in this thesis demonstrate promising performance for the development of novel EEG hyperscanning-based BCIs, certain limitations remain.

A valuable contribution of this work is the introduction of a motor-related EEG hyperscanning dataset comprising multiple subjects from diverse geographical, age, gender and backgrounds. However, as discussed in earlier chapters, the dataset size is relatively small, which constrains the through evaluation of the proposed methods, including their generalisability, and limits direct comparison with some of the latest state-of-the-art distributed signal processing and BCI approaches, such as graph neural networks (GNNs) [50] and transformer-based models [33].

Even so, the dataset size is comparable to other publicly available small motor-related EEG datasets, providing sufficient data to validate the effectiveness of the proposed techniques. Moreover, whenever possible, the proposed methods were additionally evaluated on larger publicly available datasets, further demonstrating their adaptability, robust performance, and confirming that the proposed dataset offers a representative evaluation framework comparable to established datasets.

Furthermore, this dataset was designed to approximate more realistic conditions by recording it in a standard office environment, simulating a more realistic neurorehabilitation clinic setting. This introduced common uncontrolled variables found in real-world scenarios, including varying lighting conditions, auditory and visual distractions, ambient environmental noise, and spontaneous motor movements. These factors introduce artifacts, noise variability, and increase task complexity, introducing additional challenges for the accurate recognition of neural activity. Some of these conditions were however minimised to preserve data quality. This provides an opportunity for the proposed methods to demonstrate their robustness against diverse noise conditions and task challenges as well as their performance under uncontrolled environments, demonstrating their feasibility to be implemented in real-world scenarios.

However, it should be emphasised that the dataset presented in this thesis does not represent a fully uncontrolled environment, and the methods have not

yet been tested in actual clinical settings. Consequently, the evaluation of the system robustness and its feasible implementation outside of controlled laboratory environments remains limited. Similarly, the assessment of the proposed methods as an online BCI system is constrained by the fact that online performance was simulated offline, rather than validated through real-time implementation.

7.2 Future Works

In this research, we laid the ground for hyperscanning BCI, which is in its infancy stage. While the proposed methods show promising performance, several avenues remain open for further exploration and enhancement.

Given the encouraging results obtained in this thesis, it would be interesting to further analyse the proposed BCI and newly developed methods with a broader dataset containing a more extensive collection of diverse motor and cognitive tasks, as well as data from real patients undergoing neurorehabilitation with different cognitive conditions. This would allow a more thorough evaluation of the proposed methods.

The implementation of the proposed EEG hyperscanning-based BCI in a neurorehabilitation clinic, with real patients, for a controlled long-term group therapy study would further highlight the benefits and feasibility of this new neurorehabilitation method, while providing further evaluation of the feasible implementation of the proposed methods under a more realistic environment. This study would require careful consideration of the number of patients that would form the group therapy, as well as the type of patients and type of neurorehabilitation exercises that would be considered for the study. Furthermore, a control group with patients with similar conditions that follows a more traditional neurorehabilitation therapy may be necessary for a more realistic statistical analysis of the possible benefits of hyperscanning therapy compared to traditional therapy.

Furthermore, to help alleviate some of the previously discussed system limitations, some improvements could also be explored to enhance the performance of the newly developed methods.

The developed hyperCSP may be regularised by exploiting the statistical properties of the task signals to be extracted, similar to what has been done for the

traditional CSP [188]. More robust methods, such as xDAWN [205], can be extended to the hyperscanning applications too. Exploring non-linear decompositions and source-space extensions could also enhance the interpretability of the method.

On the other hand, to optimise the dictionary of orthogonal functions for the DAOT method, there is a need for the design of an optimal dictionary learning algorithm, which is beyond the scope of this thesis. Such dictionary learning algorithm could facilitate the more dynamic allocation of the set of orthogonal Bessel functions to the sub-gestures, increasing the automatisation of the proposed method and its ability to adjust to a larger amount of prolonged motor movements.

Moreover, with access to a larger dataset and the potentially future development of more cost-efficient computational resources, latest signal processing and feature recognition techniques could be analysed and combined with the proposed methods.

Recent advances in graph signal processing and GNNs [206–208] have shown promising results in the modelling of inter-brain connectivity, which could provide more accurate estimations of functional connectivity measures. Its combination with DA could enhance the proposed methods.

With no doubt, latest deep learning strategies could also be applied for this research.

In particular, techniques such as generative adversarial networks (GANs) [209] have shown promising results in syntetically augmenting EEG datasets, particularly in clinical applications where data availability is often limited. By generating diverse synthetic samples, GANs can improve model robustness and potentially reduce overfitting. Furthermore, recent studies suggest that GANs could assist in domain adaptation by mapping patient EEG signals to distributions resembling those of healthy participants, which may help mitigate misclassification due to inter-population variability. However, careful consideration must be taken when using synthetic data for the proposed methods to avoid possible overfitting.

Techniques such as transfer learning [36, 210] have also become increasingly popular in EEG-based BCIs to improve subject-to-subject generalisation and reduce training data requirements. Pre-trained models can often be adapted to

new users with minimal data, making them appealing for clinical settings. Nevertheless, depending on the complexity of the model and the adaptation strategy, transfer learning may require significant computational resources and training time, which can hinder its feasible implementation in dynamic environments or time and resource-constrained clinical scenarios.

While both GANs and transfer learning offer valuable solutions under constrained and limited data conditions such as those presented in this thesis, their practical deployment remains limited outside laboratory-controlled environments due to computational demands, implementation complexity, and the need for thorough validation.

Recent advances in attention mechanism, such as transformer-based architectures [211, 212], and representation learning techniques, such as autoencoders [35, 213], could also be developed and adopted in this emerging research field to pave the path for highly desired semantic BCIs [214, 215]. While transformers could capture long-range temporal dependencies in EEG signals, autoencoders and unsupervised methods might enhance feature extraction and representation learning.

Despite their potential, these models require careful validation due to their high computational cost, training complexity, and limited adaptability to clinical environments. To help alleviate some of these challenges, the combination of these methods with knowledge distillation [216, 217] for EEG hyperscanning, which allows for the compression of large models into more efficient architectures more suitable for real-time applications, could be investigated.

Finally, the exploration of hybrid deep learning models [218] and multimodal approaches, such as through the combination of EEG and fNIRs [219, 220] could provide a suitable solution for the development of more reliable neurorehabilitation BCIs.

Given all this, we can conclude that this thesis has laid a strong foundation for EEG hyperscanning-based BCIs, opening a new research field, and the proposed future directions aim to transform this emerging field into a practical tool for next-generation neurorehabilitation BCIs.

Bibliography

[1] M. Zhuang, Q. Wu, F. Wan, and Y. Hu, "State-of-the-art non-invasive brain-computer interface for neural rehabilitation: A review", *Journal of Neurorestoratology*, vol. 8, no. 1, pp. 12–25, 2020. 1

[2] E. R. Kandel, J. H. Schwartz, T. M. Jessell, S. Siegelbaum, A. J. Hudspeth, and S. Mack, *Principles of neural science*, vol. 4, McGraw-hill New York, 2000. 3

[3] A. Rouzitalab, C. B. Boulay, J. Park, and A. J. Sachs, "Intracortical brain-computer interfaces in primates: a review and outlook", *Biomedical Engineering Letters*, vol. 13, no. 3, pp. 375–390, 2023. 3

[4] M. Vilela and L. R. Hochberg, "Applications of brain-computer interfaces to the control of robotic and prosthetic arms", *Handbook of Clinical Neurology*, vol. 168, pp. 87–99, 2020. 4

[5] S. Tirunagari, S. Kouchaki, D. Abasolo, and N. Poh, "One dimensional local binary patterns of electroencephalogram signals for detecting Alzheimer's disease", in *2017 22nd International Conference on Digital Signal Processing (DSP)*, 2017, pp. 1–5. 5

[6] R. Mane, T. Chouhan, and C. Guan, "BCI for stroke rehabilitation: motor and beyond", *Journal of Neural Engineering*, vol. 17, no. 4, pp. 041001, 2020. 5

[7] X. Chai, T. Cao, Q. He, N. Wang, X. Zhang, X. Shan, Z. Lv, W. Tu, Y. Yang, and J. Zhao, "Brain-computer interface digital prescription for

neurological disorders", *CNS Neuroscience & Therapeutics*, vol. 30, no. 2, pp. e14615, 2024. 5

[8] M. Sawan, M. T. Salam, J. Le Lan, A. Kassab, S. Gélinas, P. Vannasing, F. Lesage, M. Lassonde, and D. K. Nguyen, "Wireless recording systems: From noninvasive EEG-NIRS to invasive EEG devices", *IEEE Transactions on Biomedical Circuits and Systems*, vol. 7, no. 2, pp. 186–195, 2013. 5

[9] S. Sanei and J. A. Chambers, *EEG signal processing and machine learning*, John Wiley & Sons, 2021. 5

[10] M. E. Vanutelli, M. Salvadore, and C. Lucchiari, "BCI applications to creativity: Review and future directions, from little-c to C2", *Brain Sciences*, vol. 13, no. 4, 2023. 5

[11] G. A. M. Vasiljevic and L. C. De Miranda, "Brain–computer interface games based on consumer-grade EEG devices: A systematic literature review", *International Journal of Human–Computer Interaction*, vol. 36, no. 2, pp. 105–142, 2020. 5

[12] A. Al-Nafjan, M. Hosny, Y. Al-Ohali, and A. Al-Wabil, "Review and classification of emotion recognition based on EEG brain-computer interface system research: a systematic review", *Applied Sciences*, vol. 7, no. 12, pp. 1239, 2017. 5

[13] E. P. Torres, E. A. Torres, M. Hernández-Álvarez, and S. G. Yoo, "EEG-based BCI emotion recognition: A survey", *Sensors*, vol. 20, no. 18, pp. 5083, 2020. 5

[14] M. Bagheri and S. D. Power, "Simultaneous classification of both mental workload and stress level suitable for an online passive brain–computer interface", *Sensors*, vol. 22, no. 2, pp. 535, 2022. 5

[15] P. Aricò, G. Borghini, G. Di Flumeri, A. Colosimo, S. Pozzi, and F. Babiloni, "A passive brain–computer interface application for the mental workload assessment on professional air traffic controllers during realistic

air traffic control tasks", *Progress in Brain Research*, vol. 228, pp. 295–328, 2016. 5

[16] Y. Wang, X. Gao, B. Hong, C. Jia, and S. Gao, "Brain-computer interfaces based on visual evoked potentials", *IEEE Engineering in Medicine and Biology Magazine*, vol. 27, no. 5, pp. 64–71, 2008. 5

[17] M. Bamdad, H. Zarshenas, and M. A. Auais, "Application of BCI systems in neurorehabilitation: a scoping review", *Disability and Rehabilitation: Assistive Technology*, vol. 10, no. 5, pp. 355–364, 2015. 5

[18] O. Maslova, Y. Komarova, N. Shusharina, A. Kolsanov, A. Zakharov, E. Garina, and V. Pyatin, "Non-invasive EEG-based BCI spellers from the beginning to today: a mini-review", *Frontiers in Human Neuroscience*, vol. 17, pp. 1216648, 2023. 5

[19] A. Kawala-Sterniuk, N. Browarska, A. Al-Bakri, M. Pelc, J. Zygarlicki, M. Sidikova, R. Martinek, and E. J. Gorzelanczyk, "Summary of over fifty years with brain-computer interfaces—a review", *Brain Sciences*, vol. 11, no. 1, 2021. 6

[20] D. Huang, K. Qian, D.-Y. Fei, W. Jia, X. Chen, and O. Bai, "Electroencephalography (EEG)-based brain–computer interface (BCI): A 2-D virtual wheelchair control based on event-related desynchronization/synchronization and state control", *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 20, no. 3, pp. 379–388, 2012. 8

[21] P. Ahmadian, S. Sanei, L. Ascari, L. González-Villanueva, and M. A. Umiltà, "Constrained blind source extraction of readiness potentials from EEG", *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 21, no. 4, pp. 567–575, 2013. 8

[22] L. M. McCane, S. M. Heckman, D. J. McFarland, G. Townsend, J. N. Mak, E. W. Sellers, D. Zeitlin, L. M. Tenteromano, J. R. Wolpaw, and T. M. Vaughan, "P300-based brain-computer interface (BCI) event-related

potentials (ERPs): People with amyotrophic lateral sclerosis (ALS) vs. age-matched controls", *Clinical Neurophysiology*, vol. 126, no. 11, pp. 2124–2131, 2015. 8

[23] Y. I. Fishman, "The mechanisms and meaning of the mismatch negativity", *Brain Topography*, vol. 27, pp. 500–526, 2014. 9

[24] 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. 9

[25] T. Hinterberger, S. Schmidt, N. Neumann, J. Mellinger, B. Blankertz, G. Curio, and N. Birbaumer, "Brain-computer communication and slow cortical potentials", *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 1011–1018, 2004. 9

[26] P. W. Ferrez and J. del R. Millán, "Error-related EEG potentials generated during simulated brain–computer interaction", *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 3, pp. 923–929, 2008. 9

[27] J. J. Vidal, "Toward direct brain-computer communication", *Annual review of Biophysics and Bioengineering*, vol. 2, no. 1, pp. 157–180, 1973. 10

[28] L. F. Nicolas-Alonso and J. Gomez-Gil, "Brain computer interfaces, a review", *Sensors*, vol. 12, no. 2, pp. 1211–1279, 2012. 10

[29] S. Saha, K. A Mamun, K. Ahmed, R. Mostafa, G. R. Naik, S. Darvishi, A. H. Khandoker, and M. Baumert, "Progress in brain computer interface: Challenges and opportunities", *Frontiers in Systems Neuroscience*, vol. 15, pp. 578875, 2021. 11, 47

[30] A. Craik, Y. He, and J. L. Contreras-Vidal, "Deep learning for electroencephalogram (EEG) classification tasks: a review", *Journal of Neural Engineering*, vol. 16, no. 3, pp. 031001, 2019. 11, 86, 94, 95

[31] A. Al-Saegh, S. A. Dawwd, and J. M. Abdul-Jabbar, "Deep learning for motor imagery EEG-based classification: A review", *Biomedical Signal Processing and Control*, vol. 63, pp. 102172, 2021. 11, 86

[32] P. Deny, S. Cheon, H. Son, and K. W. Choi, "Hierarchical transformer for motor imagery-based brain computer interface", *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 11, pp. 5459–5470, 2023. 11

[33] M. A. Pfeffer, S. S. H. Ling, and J. K. W. Wong, "Exploring the frontier: Transformer-based models in EEG signal analysis for brain-computer interfaces", *Computers in Biology and Medicine*, p. 108705, 2024. 11, 132

[34] J. F. Hwaidi and T. M. Chen, "Classification of motor imagery EEG signals based on deep autoencoder and convolutional neural network approach", *IEEE Access*, vol. 10, pp. 48071–48081, 2022. 11

[35] S. Phadikar, N. Sinha, and R. Ghosh, "Unsupervised feature extraction with autoencoders for EEG based multiclass motor imagery BCI", *Expert Systems with Applications*, vol. 213, pp. 118901, 2023. 11, 135

[36] Z. Wan, R. Yang, M. Huang, N. Zeng, and X. Liu, "A review on transfer learning in EEG signal analysis", *Neurocomputing*, vol. 421, pp. 1–14, 2021. 11, 134

[37] S. Aggarwal and N. Chugh, "Review of machine learning techniques for EEG based brain computer interface", *Archives of Computational Methods in Engineering*, vol. 29, no. 5, pp. 3001–3020, 2022. 11

[38] K. Liao, R. Xiao, J. Gonzalez, and L. Ding, "Decoding individual finger movements from one hand using human EEG signals", *PloS one*, vol. 9, no. 1, pp. e85192, 2014. 11, 85

[39] R. Alazrai, H. Alwanni, and M. I. Daoud, "EEG-based BCI system for decoding finger movements within the same hand", *Neuroscience Letters*, vol. 698, pp. 113–120, 2019. 11, 82, 85

[40] G. Varone, W. Boulila, M. Driss, S. Kumari, M. K. Khan, T. R. Gadekallu, and A. Hussain, "Finger pinching and imagination classification: A fusion of CNN architectures for IoMT-enabled BCI applications", *Information Fusion*, vol. 101, pp. 102006, 2024. 12, 85, 86

[41] K. Eftaxias, S. Sanei, and A. H. Sayed, "Modelling brain cortical connectivity using diffusion adaptation", in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2013, pp. 959–962. 12, 44, 66, 68

[42] K. Eftaxias and S. Sanei, "Diffusion adaptive filtering for modelling brain responses to motor tasks", in *2013 18th International Conference on Digital Signal Processing (DSP)*. IEEE, 2013, pp. 1–5. 12, 64, 66, 68

[43] K. Eftaxias and S. Sanei, "Discrimination of task-related eeg signals using diffusion adaptation and s-transform coherency", in *2014 IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*. IEEE, 2014, pp. 1–6. 12, 44

[44] S. Monajemi, K. Eftaxias, S. Sanei, and S. H. Ong, "An informed multi-task diffusion adaptation approach to study tremor in Parkinson's disease", *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 7, pp. 1306–1314, 2016. 12, 27, 43, 44, 64, 66

[45] S. Sanei, O. Geman, S. Monajemi, K. Eftaxias, and D. Jarchi, "Cooperative learning for biomedical signal processing and recognition", in *2017 E-Health and Bioengineering Conference (EHB)*. IEEE, 2017, pp. 687–692. 12, 43, 64, 66

[46] S. Sanei, C. C. Took, D. Jarchi, and A. Prochazka, "EEG connectivity-informed cooperative adaptive line enhancer for recognition of brain state", in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 1195–1199. 12, 27, 44

[47] A. H. Sayed, "Chapter 9 - diffusion adaptation over networks", in *Academic Press Library in Signal Processing: Volume 3*, vol. 3 of *Academic Press*

Library in Signal Processing, pp. 323–453. Elsevier, 2014. 12, 14, 18, 24, 25, 26, 28, 39

[48] E. Bullmore and O. Sporns, "Complex brain networks: graph theoretical analysis of structural and functional systems", *Nature Reviews Neuroscience*, vol. 10, no. 3, pp. 186–198, 2009. 12

[49] C. Seguin, O. Sporns, and A. Zalesky, "Brain network communication: concepts, models and applications", *Nature Reviews Neuroscience*, vol. 24, no. 9, pp. 557–574, 2023. 12

[50] R. Li, X. Yuan, M. Radfar, P. Marendy, W. Ni, T. J. O'Brien, and P. M. Casillas-Espinosa, "Graph signal processing, graph neural network and graph learning on biological data: A systematic review", *IEEE Reviews in Biomedical Engineering*, vol. 16, pp. 109–135, 2023. 12, 132

[51] P. R. Montague, G. S. Berns, J. D. Cohen, S. M. McClure, G. Pagnoni, M. Dhamala, M. C. Wiest, I. Karpov, R. D. King, N. Apple, and R. E. Fisher, "Hyperscanning: simultaneous fMRI during linked social interactions", *Neuroimage*, vol. 16, no. 4, pp. 1159–1164, 2002. 13

[52] D. Liu, S. Liu, X. Liu, C. Zhang, A. Li, C. Jin, Y. Chen, H. Wang, and X. Zhang, "Interactive brain activity: review and progress on EEG-based hyperscanning in social interactions", *Frontiers in Psychology*, vol. 9, pp. 1862, 2018. 14, 48, 53, 62

[53] A. Czeszumski, S. Eustergerling, A. Lang, D. Menrath, M. Gerstenberger, S. Schuberth, F. Schreiber, Z. Z. Rendon, and P. König, "Hyperscanning: a valid method to study neural inter-brain underpinnings of social interaction", *Frontiers in Human Neuroscience*, vol. 14, pp. 39, 2020. 14, 48, 62

[54] M. Balconi and M. E. Vanutelli, "Cooperation and competition with hyperscanning methods: Review and future application to emotion domain", *Frontiers in Computational Neuroscience*, vol. 11, pp. 86, 2017. 14, 48, 53, 62, 111

[55] M. R. Short, J. C. Hernandez-Pavon, A. Jones, and J. L. Pons, "EEG hyperscanning in motor rehabilitation: a position paper", *Journal of Neuroengineering and Rehabilitation*, vol. 18, no. 1, pp. 1–15, 2021. 14, 48, 51, 63

[56] S. Sanei, *Adaptive processing of brain signals*, John Wiley & Sons, 2013. 18, 27, 43

[57] A. H. Sayed, "Adaptation, learning, and optimization over networks", *Foundations and Trends® in Machine Learning*, vol. 7, no. 4-5, pp. 311–801, 2014. 18, 20, 21, 22, 23, 24, 26, 39

[58] L. eon Bottou, "Online learning and stochastic approximations", *Online learning in neural networks*, vol. 17, no. 9, pp. 142, 1998. 20

[59] C. G. Lopes and A. H. Sayed, "Incremental adaptive strategies over distributed networks", *IEEE Transactions on Signal Processing*, vol. 55, no. 8, pp. 4064–4077, 2007. 20, 21, 22, 28

[60] M. H. DeGroot, "Reaching a consensus", *Journal of the American Statistical Association*, vol. 69, no. 345, pp. 118–121, 1974. 20, 23

[61] R. Olfati-Saber, J. A. Fax, and R. M. Murray, "Consensus and cooperation in networked multi-agent systems", *Proceedings of the IEEE*, vol. 95, no. 1, pp. 215–233, 2007. 20

[62] C. G. Lopes and A. H. Sayed, "Diffusion least-mean squares over adaptive networks: Formulation and performance analysis", *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3122–3136, 2008. 20, 26, 27

[63] J. Chen and A. H. Sayed, "Diffusion adaptation strategies for distributed optimization and learning over networks", *IEEE Transactions on Signal Processing*, vol. 60, no. 8, pp. 4289–4305, 2012. 20, 27

[64] S. Y. Tu and A. H Sayed, "Diffusion strategies outperform consensus strategies for distributed estimation over adaptive networks", *IEEE Transactions on Signal Processing*, vol. 60, no. 12, pp. 6217–6234, 2012. 20, 24, 44

[65] R. Kukde, M. S. Manikandan, and G. Panda, "Incremental learning based adaptive filter for nonlinear distributed active noise control system", *IEEE Open Journal of Signal Processing*, vol. 1, pp. 1–13, 2020. 22

[66] L. Georgopoulos and M. Hasler, "Distributed machine learning in networks by consensus", *Neurocomputing*, vol. 124, pp. 2–12, 2014. 24

[67] Y. Shi, C. T. Lin, Y. C. Chang, W. Ding, Y. Shi, and X. Yao, "Consensus learning for distributed fuzzy neural network in big data environment", *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 5, no. 1, pp. 29–41, 2020. 24

[68] W. Chen, S. Hua, and S. S. Ge, "Consensus-based distributed cooperative learning control for a group of discrete-time nonlinear multi-agent systems using neural networks", *Automatica*, vol. 50, no. 9, pp. 2254–2268, 2014. 24

[69] A. Bertrand, "Distributed signal processing for wireless EEG sensor networks", *IEEE Transactions on Neural Systems and Rehabilitation engineering*, vol. 23, no. 6, pp. 923–935, 2015. 24

[70] S. Monajemi, S. Sanei, and S. H. Ong, "Advances in bacteria motility modelling via diffusion adaptation", in *2014 22nd European Signal Processing Conference (EUSIPCO)*. IEEE, 2014, pp. 2335–2339. 26

[71] X. Zhao and A. H. Sayed, "Learning over social networks via diffusion adaptation", in *2012 Conference Record of the Forty Sixth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*. IEEE, 2012, pp. 709–713. 26

[72] F. S. Cattivelli and A. H. Sayed, "Modeling bird flight formations using diffusion adaptation", *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2038–2051, 2011. 27

[73] S. Y. Tu and A. H. Sayed, "Cooperative prey herding based on diffusion adaptation", in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2011, pp. 3752–3755. 27, 29, 34

[74] S. Monajemi, S. Ensafi, S. Lu, A. A. Kassim, C. L. Tan, S. Sanei, and S. H. Ong, "Classification of HEp-2 cells using distributed dictionary learning", in *2016 24th European Signal Processing Conference (EUSIPCO)*. IEEE, 2016, pp. 1163–1167. 27

[75] S. Monajemi, S. Sanei, S. H. Ong, and A. H. Sayed, "Adaptive regularized diffusion adaptation over multitask networks", in *2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP)*. IEEE, 2015, pp. 1–5. 27, 40

[76] R. Nassif, C. Richard, A. Ferrari, and A. H. Sayed, "Multitask diffusion adaptation over asynchronous networks", *IEEE Transactions on Signal Processing*, vol. 64, no. 11, pp. 2835–2850, 2016. 27, 40

[77] M. Latifi, A. Khalili, A. Rastegarnia, and S. Sanei, "Fully distributed demand response using the adaptive diffusion–stackelberg algorithm", *IEEE Transactions on Industrial Informatics*, vol. 13, no. 5, pp. 2291–2301, 2017. 27

[78] A. Falcon-Caro and S. Sanei, "Diffusion adaptation for crowd analysis", in *2021 International Conference on e-Health and Bioengineering (EHB)*. IEEE, 2021, pp. 1–4. 27, 29, 38

[79] A. Falcon-Caro, E. Peytchev, and S. Sanei, "Adaptive network model for assisting people with disabilities through crowd monitoring and control", *Bioengineering*, vol. 11, no. 3, pp. 283, 2024. 27

[80] A. Falcon-Caro and S. Sanei, "Cooperative networking approach to assisting blinds in a crowd using air trackers", in *2022 4th International Conference on Emerging Trends in Electrical, Electronic and Communications Engineering (ELECOM)*. IEEE, 2022, pp. 1–5. 29, 38

[81] S. Y. Tu and A. H. Sayed, "Tracking behavior of mobile adaptive networks", in *2010 Conference Record of the Forty Fourth Asilomar Conference on Signals, Systems and Computers*, 2010, pp. 698–702. 29, 30

[82] J. Chen, C. Richard, and A. H. Sayed, "Multitask diffusion adaptation over networks", *IEEE Transactions on Signal Processing*, vol. 62, no. 16, pp. 4129–4144, 2014. 40, 41, 42

[83] J. Chen, C. Richard, and A. H. Sayed, "Diffusion lms over multitask networks", *IEEE Transactions on Signal Processing*, vol. 63, no. 11, pp. 2733–2748, 2015. 40, 42

[84] T. Evgeniou and M. Pontil, "Regularized multi–task learning", in *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2004, pp. 109–117. 42

[85] S. Ji and J. Ye, "An accelerated gradient method for trace norm minimization", in *Proceedings of the 26th Annual International Conference on Machine Learning*, 2009, pp. 457–464. 42

[86] J. Zhou, J. Chen, and J. Ye, "Clustered multi-task learning via alternating structure optimization", *Advances in Neural Information Processing Systems*, vol. 24, 2011. 42

[87] J. Chen, L. Tang, J. Liu, and J. Ye, "A convex formulation for learning shared structures from multiple tasks", in *Proceedings of the 26th Annual International Conference on Machine Learning*, 2009, pp. 137–144. 43

[88] O. Chapelle, P. Shivaswamy, S. Vadrevu, K. Weinberger, Y. Zhang, and B. Tseng, "Multi-task learning for boosting with application to web search ranking", in *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2010, pp. 1189–1198. 43

[89] J. Zhou, L. Yuan, J. Liu, and J. Ye, "A multi-task learning formulation for predicting disease progression", in *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2011, pp. 814–822. 43

[90] M. Tangermann, K.-R. Müller, A. Aertsen, N. Birbaumer, C. Braun, C. Brunner, R. Leeb, C. Mehring, K. J. Miller, G. R. Müller-Putz, et al.,

"Review of the BCI competition IV", *Frontiers in neuroscience*, vol. 6, pp. 55, 2012. 46, 53

[91] C. He, Y.-Y. Chen, C.-R. Phang, C. Stevenson, I-P. Chen, T.-P. Jung, and L.-W. Ko, "Diversity and suitability of the state-of-the-art wearable and wireless EEG systems review", *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 8, pp. 3830–3843, 2023. 47

[92] A. Bizzego, G. Gabrieli, A. Azhari, M. Lim, and G. Esposito, "Dataset of parent-child hyperscanning functional near-infrared spectroscopy recordings", *Scientific Data*, vol. 9, no. 1, pp. 625, 2022. 47

[93] M. A. Ramírez-Moreno, J. G. Cruz-Garza, and J. L. Contreras-Vidal, "Mobile EEG recordings of musical (jazz) improvisation", 2022. 47

[94] R. Hari and M. V. Kujala, "Brain basis of human social interaction: from concepts to brain imaging", *Physiological Reviews*, vol. 89, no. 2, pp. 453–479, 2009. 47

[95] E. Tognoli, J. Lagarde, G. C. DeGuzman, and J. S. Kelso, "The phi complex as a neuromarker of human social coordination", *Proceedings of the National Academy of Sciences*, vol. 104, no. 19, pp. 8190–8195, 2007. 48

[96] G. Dumas, J. Nadel, R. Soussignan, J. Martinerie, and L. Garnero, "Inter-brain synchronization during social interaction", *PloS one*, vol. 5, no. 8, pp. e12166, 2010. 48

[97] K. Yun, K. Watanabe, and S. Shimojo, "Interpersonal body and neural synchronization as a marker of implicit social interaction", *Scientific Reports*, vol. 2, no. 1, pp. 959, 2012. 48

[98] A. Nijholt and H. Gürkök, "Multi-brain games: cooperation and competition", in *Universal Access in Human-Computer Interaction. Design Methods, Tools, and Interaction Techniques for eInclusion: 7th International Conference, UAHCI 2013, Held as Part of HCI International 2013, Las Vegas, NV, USA, July 21-26, 2013, Proceedings, Part I 7*. Springer, 2013, pp. 652–661. 48

[99] Hans Berger, "Über das elektroenkephalogramm des menschen", *Archiv für Psychiatrie und Nervenkrankheiten*, vol. 87, no. 1, pp. 527–570, 1929. 48

[100] H. H. Jasper, "Ten-twenty electrode system of the international federation", *Electroencephalogr Clin Neurophysiol*, vol. 10, pp. 371–375, 1958. 49

[101] 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. 49

[102] M. A. Lopez-Gordo, D. Sanchez-Morillo, and F. P. Valle, "Dry EEG electrodes", *Sensors*, vol. 14, no. 7, pp. 12847–12870, 2014. 50

[103] M. Rashid, N. Sulaiman, A. PP Abdul Majeed, R. M. Musa, A. F. Ab. Nasir, B. S. Bari, and S. Khatun, "Current status, challenges, and possible solutions of EEG-based brain-computer interface: a comprehensive review", *Frontiers in Neurorobotics*, vol. 14, pp. 25, 2020. 50

[104] I. Kaya, "A brief summary of EEG artifact handling", *Brain-Computer Interface*, , no. 9, 2019. 50, 92

[105] R. Guarnieri, M. Marino, F. Barban, M. Ganzetti, and D. Mantini, "Online EEG artifact removal for BCI applications by adaptive spatial filtering", *Journal of Neural Engineering*, vol. 15, no. 5, pp. 056009, 2018. 51

[106] X. Jiang, G.-B. Bian, and Z. Tian, "Removal of artifacts from EEG signals: a review", *Sensors*, vol. 19, no. 5, pp. 987, 2019. 51

[107] P. Arpaia, A. Esposito, A. Natalizio, and M. Parvis, "How to successfully classify EEG in motor imagery BCI: a metrological analysis of the state of the art", *Journal of Neural Engineering*, vol. 19, no. 3, pp. 031002, 2022. 53, 85

[108] A. Falcon-Caro, S. Shirani, J. F. Ferreira, J. J. Bird, and S. Sanei, "Formulation of common spatial patterns for multi-task hyperscanning BCI dataset", 2024. 56, 131

[109] J. Toppi, G. Borghini, M. Petti, E. J. He, V. De Giusti, B. He, L. Astolfi, and F. Babiloni, "Investigating cooperative behavior in ecological settings: an EEG hyperscanning study", *PloS one*, vol. 11, no. 4, pp. e0154236, 2016. 60, 111

[110] F. Babiloni, F. Cincotti, D. Mattia, F. D. V. Fallani, A. Tocci, L. Bianchi, S. Salinari, M.G. Marciani, A. Colosimo, and L. Astolfi, "High resolution EEG hyperscanning during a card game", in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2007, pp. 4957–4960. 60, 63

[111] Z. Khademi, F. Ebrahimi, and H. M. Kordy, "A review of critical challenges in MI-BCI: From conventional to deep learning methods", *Journal of Neuroscience Methods*, vol. 383, pp. 109736, 2023. 60, 109

[112] P. Comon, "Independent component analysis, a new concept?", *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994. 61

[113] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications", *Neural Networks*, vol. 13, no. 4-5, pp. 411–430, 2000. 61

[114] A. Falcon-Caro, M. Frincu, and S. Sanei, "A diffusion adaptation approach to model brain responses in an EEG-based hyperscanning study", in *2023 IEEE Statistical Signal Processing Workshop (SSP)*. IEEE, 2023, pp. 393–397. 62

[115] A. F. de C. Hamilton, "Hyperscanning: beyond the hype", *Neuron*, vol. 109, no. 3, pp. 404–407, 2021. 62

[116] C. O'Reilly, J. D. Lewis, and M. Elsabbagh, "Is functional brain connectivity atypical in autism? a systematic review of EEG and MEG studies", *PloS One*, vol. 12, no. 5, pp. e0175870, 2017. 63

[117] Y. Du, Z. Fu, and V. D. Calhoun, "Classification and prediction of brain disorders using functional connectivity: promising but challenging", *Frontiers in Neuroscience*, vol. 12, pp. 525, 2018. 63

[118] K. Helm, K. Viol, T. M. Weiger, P. A. Tass, C. Grefkes, D. Del Monte, and G. Schiepek, "Neuronal connectivity in major depressive disorder: a systematic review", *Neuropsychiatric Disease and Treatment*, pp. 2715–2737, 2018. 63

[119] Y. Kong, J. Gao, Y. Xu, Y. Pan, J. Wang, and J. Liu, "Classification of autism spectrum disorder by combining brain connectivity and deep neural network classifier", *Neurocomputing*, vol. 324, pp. 63–68, 2019. 63

[120] W. Zhang, X. Han, S. Qiu, T. Li, C. Chu, L. Wang, J. Wang, Z. Zhang, R. Wang, and M. et al. Yang, "Analysis of brain functional network based on EEG signals for early-stage Parkinson's disease detection", *IEEE Access*, vol. 10, pp. 21347–21358, 2022. 63

[121] V. Sakkalis, "Review of advanced techniques for the estimation of brain connectivity measured with EEG/MEG", *Computers in Biology and Medicine*, vol. 41, no. 12, pp. 1110–1117, 2011. 64

[122] A. M. Bastos and J.-M. Schoffelen, "A tutorial review of functional connectivity analysis methods and their interpretational pitfalls", *Frontiers in Systems Neuroscience*, vol. 9, pp. 175, 2016. 64

[123] G. Chiarion, L. Sparacino, Y. Antonacci, L. Faes, and L. Mesin, "Connectivity analysis in EEG data: a tutorial review of the state of the art and emerging trends", *Bioengineering*, vol. 10, no. 3, pp. 372, 2023. 64

[124] 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. 64

[125] C. J. Stam, G. Nolte, and A. Daffertshofer, "Phase lag index: assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources", *Human Brain Mapping*, vol. 28, no. 11, pp. 1178–1193, 2007. 64

[126] L. A. Baccalá and K. Sameshima, "Partial directed coherence: a new concept in neural structure determination", *Biological Cybernetics*, vol. 84, no. 6, pp. 463–474, 2001. 64

[127] M. J. Kaminski and K. J. Blinowska, "A new method of the description of the information flow in the brain structures", *Biological Cybernetics*, vol. 65, no. 3, pp. 203–210, 1991. 64

[128] J. Geweke, "Measurement of linear dependence and feedback between multiple time series", *Journal of the American Statistical Association*, vol. 77, no. 378, pp. 304–313, 1982. 65

[129] C. W.J. Granger, "Investigating causal relations by econometric models and cross-spectral methods", *Econometrica: Journal of the Econometric Society*, pp. 424–438, 1969. 65

[130] S. Monajemi, D. Jarchi, S.-H. Ong, and S. Sanei, "Cooperative particle filtering for tracking ERP subcomponents from multichannel EEG", *Entropy*, vol. 19, no. 5, pp. 199, 2017. 66

[131] W. Penny and L. Harrison, "Multivariate autoregressive models", *Statistical Parametric Mapping: The Analysis of Functional Brain Images. Academic Press, Amsterdam*, pp. 534–540, 2007. 68

[132] G. Deshpande, S. LaConte, G. A. James, S. Peltier, and X. Hu, "Multivariate Granger causality analysis of fMRI data", *Human Brain Mapping*, vol. 30, no. 4, pp. 1361–1373, 2009. 69

[133] L. Barnett and A. K. Seth, "The MVGC multivariate Granger causality toolbox: a new approach to Granger-causal inference", *Journal of Neuroscience Methods*, vol. 223, pp. 50–68, 2014. 69, 91, 94

[134] Luca Pion-Tonachini, Ken Kreutz-Delgado, and Scott Makeig, "ICLabel: An automated electroencephalographic independent component classifier, dataset, and website", *NeuroImage*, vol. 198, pp. 181–197, 2019. 74, 92

[135] H. Akaike, "A new look at the statistical model identification", *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, 1974. 74

[136] R. Oostenveld, P. Fries, E. Maris, and J-M. Schoffelen, "FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data", *Computational Intelligence and Neuroscience*, vol. 2011, no. 1, pp. 156869, 2011. 74

[137] A. Falcon-Caro, F. J. Ferreira, and S. Sanei, "Cooperative identification of prolonged motor movement from EEG for BCI without feedback", *IEEE Access*, vol. 13, pp. 11765–11777, 2025. 82

[138] R. Xiao and L. Ding, "Evaluation of EEG features in decoding individual finger movements from one hand", *Computational and Mathematical Methods in Medicine*, vol. 2013, no. 1, pp. 243257, 2013. 82

[139] P. D. E. Baniqued, E. C. Stanyer, M. Awais, A. Alazmani, A. E. Jackson, M. A. Mon-Williams, F. Mushtaq, and R. J. Holt, "Brain–computer interface robotics for hand rehabilitation after stroke: a systematic review", *Journal of Neuroengineering and Rehabilitation*, vol. 18, pp. 1–25, 2021. 82

[140] O.-Y. Kwon, M.-H. Lee, C. Guan, and S.-W. Lee, "Subject-independent brain–computer interfaces based on deep convolutional neural networks", *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 10, pp. 3839–3852, 2020. 83

[141] R. J. Kobler, A. I. Sburlea, V. Mondini, M. Hirata, and G. R. Müller-Putz, "Distance-and speed-informed kinematics decoding improves M/EEG based upper-limb movement decoder accuracy", *Journal of Neural Engineering*, vol. 17, no. 5, pp. 056027, 2020. 84

[142] P. Wang, Z. Li, P. Gong, Y. Zhou, F. Chen, and D. Zhang, "Mtrt: Motion trajectory reconstruction transformer for EEG-Based BCI decoding", *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 2349–2358, 2023. 84

[143] K. Anam, M. Nuh, and A. Al-Jumaily, "Comparison of EEG pattern recognition of motor imagery for finger movement classification", in *2019 6th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*. IEEE, 2019, pp. 24–27. 85

[144] M. Kato, S. Kanoga, T. Hoshino, and T. Fukami, "Motor imagery classification of finger motions using multiclass CSP", in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2020, pp. 2991–2994. 85

[145] N. Robinson, A. P. Vinod, C. Guan, K. K. Ang, and T. K. Peng, "A wavelet-CSP method to classify hand movement directions in EEG based BCI system", in *2011 8th International Conference on Information, Communications Signal Processing*, 2011, pp. 1–5. 85

[146] A. Subasi and M. I. Gursoy, "EEG signal classification using PCA, ICA, LDA and support vector machines", *Expert Systems with Applications*, vol. 37, no. 12, pp. 8659–8666, 2010. 85

[147] G. Bressan, G. Cisotto, G. R. Müller-Putz, and S. C. Wriessnegger, "Deep learning-based classification of fine hand movements from low frequency EEG", *Future Internet*, vol. 13, no. 5, 2021. 85, 86

[148] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep learning for computer vision: A brief review", *Computational Intelligence and Neuroscience*, vol. 2018, 2018. 85

[149] A. Mehrish, N. Majumder, R. Bharadwaj, R. Mihalcea, and S. Poria, "A review of deep learning techniques for speech processing", *Information Fusion*, p. 101869, 2023. 85

[150] S. Srivastava, A. V. Divekar, C. Anilkumar, I. Naik, V. Kulkarni, and V. Pattabiraman, "Comparative analysis of deep learning image detection algorithms", *Journal of Big Data*, vol. 8, no. 1, pp. 66, 2021. 85

[151] 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. 14681–14722, 2023. 85, 86, 87, 97, 107

[152] K. Anam, S. Bukhori, F. S. Hanggara, and M. Pratama, "Subject-independent classification on brain-computer interface using autonomous deep learning for finger movement recognition", in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2020, pp. 447–450. 86, 106

[153] H. N. Zahra, H. Zakaria, and B. R. Hermanto, "Exploration of pattern recognition methods for motor imagery EEG signal with convolutional neural network approach", in *Journal of Physics: Conference Series*. IOP Publishing, 2022, vol. 2312, p. 012064. 86

[154] M. Degirmenci, Y. K. Yuce, M. Perc, and Y. Isler, "EEG-based finger movement classification with intrinsic time-scale decomposition", *Frontiers in Human Neuroscience*, vol. 18, pp. 1362135, 2024. 86

[155] S. Tortora, S. Ghidoni, C. Chisari, S. Micera, and F. Artoni, "Deep learning-based BCI for gait decoding from EEG with LSTM recurrent neural network", *Journal of Neural Engineering*, vol. 17, no. 4, pp. 046011, 2020. 86

[156] J. Ruiz-Pinales, J. G. Avina-Cervantes, J. J. Gonzalez-Barbosa, and J. L. et al. Contreras-Hernandez, "Empirical mode decomposition and a bidirectional LSTM architecture used to decode individual finger MI-EEG signals", *Journal of Advances in Applied & Computational Mathematics*, vol. 9, pp. 32–48, 2022. 86

[157] L. Ji, L. Yi, C. Huang, H. Li, W. Han, and N. Zhang, "Classification of hand movements from EEG using a FusionNet based LSTM network", *Journal of Neural Engineering*, vol. 21, no. 6, pp. 066013, 2024. 86

[158] M. Zhou, C. Tian, R. Cao, B. Wang, Y. Niu, T. Hu, H. Guo, and J. Xiang, "Epileptic seizure detection based on EEG signals and CNN", *Frontiers in Neuroinformatics*, vol. 12, pp. 95, 2018. 86

[159] B. TaghiBeyglou, A. Shahbazi, F. Bagheri, S. Akbarian, and M. Jahed, "Detection of ADHD cases using CNN and classical classifiers of raw EEG",

Computer Methods and Programs in Biomedicine Update, vol. 2, pp. 100080, 2022. 86

[160] Z. Wang, L. Cao, Z. Zhang, X. Gong, Y. Sun, and H. Wang, "Short time fourier transformation and deep neural networks for motor imagery brain computer interface recognition", *Concurrency and Computation: Practice and Experience*, vol. 30, no. 23, pp. e4413, 2018. 87

[161] Z. Tayeb, J. Fedjaev, N. Ghaboosi, C. Richter, L. Everding, X. Qu, Y. Wu, G. Cheng, and J. Conradt, "Validating deep neural networks for online decoding of motor imagery movements from EEG signals", *Sensors*, vol. 19, no. 1, pp. 210, 2019. 87

[162] F. Bowman, *Introduction to Bessel functions*, Courier Corporation, 2012. 87, 88

[163] B. G. Korenev, *Bessel Functions and their Applications*, CRC Press, 2002. 89

[164] J. Schroeder, "Signal processing via Fourier-Bessel series expansion", *Digital Signal Processing-New York*, vol. 3, pp. 112–112, 1993. 89

[165] S. Bhalerao, S. Ainwad, and R. Pachori, "FBSE based automated classification of motor imagery EEG signals in brain–computer interface", *Handbook of Neural Engineering (Handbook of Neural Engineering)*, vol. 2, 2022. 89

[166] A. Nalwaya and R. B. Pachori, "Fourier–Bessel domain adaptive wavelet transform-based method for emotion identification from EEG signals", *IEEE Sensors Letters*, vol. 8, no. 2, pp. 1–4, 2024. 89

[167] M. J. Lighthill, *An Introduction to Fourier Analysis and generalised Functions*, Cambridge University Press, 1958. 89

[168] K. J. Blinowska, "Review of the methods of determination of directed connectivity from multichannel data", *Medical & Biological Engineering & Computing*, vol. 49, pp. 521–529, 2011. 91

[169] H. Cho, M. Ahn, S. Ahn, M. Kwon, and S. C. Jun, "Supporting data for EEG datasets for motor imagery brain computer interface", *GigaScience Database*, 2017. 92

[170] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces", *Journal of Neural Engineering*, vol. 15, no. 5, pp. 056013, 2018. 96, 101

[171] C.-S. Wei, T. Koike-Akino, and Y. Wang, "Spatial component-wise convolutional network (SCCNet) for motor-imagery EEG classification", in *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*, 2019, pp. 328–331. 97, 101

[172] D. Quade, "Using weighted rankings in the analysis of complete blocks with additive block effects", *Journal of the American Statistical Association*, vol. 74, no. 367, pp. 680–683, 1979. 98

[173] P. B. Nemenyi, *Distribution-free multiple comparisons.*, Princeton University, 1963. 98

[174] Y. Rao, L. Zhang, R. Jing, J. Huo, K. Yan, J. He, X. Hou, J. Mu, W. Geng, and H. et al. Cui, "An optimized EEGNet decoder for decoding motor image of four class fingers flexion", *Brain Research*, p. 149085, 2024. 101, 105, 106

[175] N. J. Limbaga, K. L. Mallari, N. R. Yeung, and J. C. Monje, "Development of an EEG-based brain-controlled system for a virtual prosthetic hand", in *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2022, pp. 1714–1717. 105

[176] A. Falcon-Caro, S. Shirani, J. F. Ferreira, J. J. Bird, and S. Sanei, "Formulation of common spatial patterns for multi-task hyperscanning BCI", *IEEE Transactions on Biomedical Engineering*, 2024. 111

[177] D. Jaipriya and K.C. Sriharipriya, "Brain computer interface-based signal processing techniques for feature extraction and classification of motor im-

agery using EEG: A literature review", *Biomedical Materials & Devices*, vol. 2, no. 2, pp. 601–613, 2024. 111

[178] N. Sinha, T. Maszczyk, Zhang W., J. Tan, and J. Dauwels, "EEG hyperscanning study of inter-brain synchrony during cooperative and competitive interaction", in *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2016, pp. 004813–004818. 111

[179] Z. J. Koles, M. S. Lazar, and S. Z. Zhou, "Spatial patterns underlying population differences in the background EEG", *Brain Topography*, vol. 2, pp. 275–284, 1990. 112, 113

[180] H. Lu, H.-L. Eng, C. Guan, K. N. Plataniotis, and A. N. Venetsanopoulos, "Regularized common spatial pattern with aggregation for EEG classification in small-sample setting", *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 12, pp. 2936–2946, 2010. 112, 129

[181] M. Aljalal, S. A. Aldosari, K. AlSharabi, A. M. Abdurraqeeb, and F. A. Alturki, "Parkinson's disease detection from resting-state EEG signals using common spatial pattern, entropy, and machine learning techniques", *Diagnostics*, vol. 12, no. 5, pp. 1033, 2022. 113

[182] Y. Zhang, Y. Guo, P. Yang, W. Chen, and B. Lo, "Epilepsy seizure prediction on EEG using common spatial pattern and convolutional neural network", *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 2, pp. 465–474, 2019. 113

[183] C. Li, W. Zhou, G. Liu, Y. Zhang, M. Geng, Z. Liu, S. Wang, and W. Shang, "Seizure onset detection using empirical mode decomposition and common spatial pattern", *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 458–467, 2021. 113

[184] M. Li and B.-L. Lu, "Emotion classification based on gamma-band EEG", in *2009 Annual International Conference of the IEEE Engineering in medicine and biology society*. IEEE, 2009, pp. 1223–1226. 113

[185] M. D. Basar, A. D. Duru, and A. Akan, "Emotional state detection based on common spatial patterns of EEG", *Signal, Image and Video Processing*, vol. 14, no. 3, pp. 473–481, 2020. 113

[186] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement", *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 4, pp. 441–446, 2000. 113

[187] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Muller, "Optimizing spatial filters for robust EEG single-trial analysis", *IEEE Signal Processing Magazine*, vol. 25, no. 1, pp. 41–56, 2007. 114

[188] F. Lotte and C. Guan, "Regularizing common spatial patterns to improve BCI designs: Unified theory and new algorithms", *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 2, pp. 355–362, 2011. 114, 129, 134

[189] K. K. Ang, Z. Y. Chin, C. Wang, C. Guan, and H. Zhang, "Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b", *Frontiers in Neuroscience*, vol. 6, pp. 39, 2012. 114

[190] S.-H. Park, D. Lee, and S.-G. Lee, "Filter bank regularized common spatial pattern ensemble for small sample motor imagery classification", *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 2, pp. 498–505, 2017. 114

[191] S. Sethi, R. Upadhyay, and H. S. Singh, "Stockwell-common spatial pattern technique for motor imagery-based brain computer interface design", *Computers & Electrical Engineering*, vol. 71, pp. 492–504, 2018. 114

[192] F. Goksu, N. F. Ince, and A. H. Tewfik, "Sparse common spatial patterns in brain computer interface applications", in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011, pp. 533–536. 114

[193] M. Amiri, H. Aghaeinia, and H. R. Amindavar, "Automatic epileptic seizure detection in EEG signals using sparse common spatial pattern and adaptive

short-time Fourier transform-based synchrosqueezing transform", *Biomedical Signal Processing and Control*, vol. 79, pp. 104022, 2023. 114

[194] H. Kang, Y. Nam, and S. Choi, "Composite common spatial pattern for subject-to-subject transfer", *IEEE Signal Processing Letters*, vol. 16, no. 8, pp. 683–686, 2009. 114

[195] S. Hatamikia and A. M. Nasrabadi, "Subject transfer BCI based on composite local temporal correlation common spatial pattern", *Computers in Biology and Medicine*, vol. 64, pp. 1–11, 2015. 114

[196] Q. Wei and X. Ding, "Intra- and inter-subject common spatial pattern for reducing calibration effort in MI-based BCI", *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 904–916, 2023. 114

[197] T. Nguyen, I. Hettiarachchi, A. Khatami, L. Gordon-Brown, C. P. Lim, and S. Nahavandi, "Classification of multi-class BCI data by common spatial pattern and fuzzy system", *IEEE Access*, vol. 6, pp. 27873–27884, 2018. 114, 120

[198] K. Darvish Ghanbar, T. Yousefi Rezaii, A. Farzamnia, and I. Saad, "Correlation-based common spatial pattern (CCSP): A novel extension of CSP for classification of motor imagery signal", *Plos One*, vol. 16, no. 3, pp. e0248511, 2021. 114, 116, 119

[199] T. Long, M. Wan, W. Jian, H. Dai, W. Nie, and J. Xu, "Application of multi-task transfer learning: The combination of EA and optimized subband regularized CSP to classification of 8-channel EEG signals with small dataset", *Frontiers in Human Neuroscience*, vol. 17, pp. 1143027, 2023. 114

[200] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis", *Journal of Neuroscience Methods*, vol. 134, no. 1, pp. 9–21, 2004. 117

[201] Z.-c. Tang, C. Li, J.-f. Wu, P.-c. Liu, and S.-w. Cheng, "Classification of EEG-based single-trial motor imagery tasks using a B-CSP method for

BCI", *Frontiers of Information Technology & Electronic Engineering*, vol. 20, no. 8, pp. 1087–1098, 2019. 119

[202] M. Aljalal, R. Djemal, K. AlSharabi, and S. Ibrahim, "Feature extraction of EEG based motor imagery using CSP based on logarithmic band power, entropy and energy", in *2018 1st International Conference on Computer Applications & Information Security (ICCAIS)*. IEEE, 2018, pp. 1–6. 119

[203] I. Sturm, S. Lapuschkin, W. Samek, and K.-R. Müller, "Interpretable deep neural networks for single-trial EEG classification", *Journal of Neuroscience Methods*, vol. 274, pp. 141–145, 2016. 129

[204] M. Dehghani, A. Mobaien, and R. Boostani, "A deep neural network-based transfer learning to enhance the performance and learning speed of BCI systems", *Brain-Computer Interfaces*, vol. 8, no. 1-2, pp. 14–25, 2021. 129

[205] B. Rivet, A. Souloumiac, V. Attina, and G. Gibert, "xDAWN algorithm to enhance evoked potentials: application to brain–computer interface", *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 8, pp. 2035–2043, 2009. 134

[206] C. Liu, X. Zhou, Y. Wu, R. Yang, Z. Wang, L. Zhai, Z. Jia, and Y. Liu, "Graph neural networks in EEG-based emotion recognition: a survey", *arXiv preprint arXiv:2402.01138*, 2024. 134

[207] A. Hajisafi, H. Lin, Y.-Y. Chiang, and C. Shahabi, "Dynamic GNNs for precise seizure detection and classification from EEG data", in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, 2024, pp. 207–220. 134

[208] S. Abadal, P. Galván, A. Mármol, N. Mammone, C. Ieracitano, M. L. Giudice, A. Salvini, and F. C. Morabito, "Graph neural networks for electroencephalogram analysis: Alzheimer's disease and epilepsy use cases", *Neural Networks*, vol. 181, pp. 106792, 2025. 134

[209] C. Fan, B. Yang, X. Li, S. Gao, and P. Zan, "EEG-based feature classification combining 3D-convolutional neural networks with generative adver-

sarial networks for motor imagery", *Journal of Integrative Neuroscience*, vol. 23, no. 8, pp. 153, 2024. 134

[210] A. M. Azab, L. Mihaylova, K. K. Ang, and M. Arvaneh, "Weighted transfer learning for improving motor imagery-based brain–computer interface", *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 7, pp. 1352–1359, 2019. 134

[211] A. Hameed, R. Fourati, B Ammar, A. Ksibi, A. S. Alluhaidan, M. B. Ayed, and H. K. Khleaf, "Temporal–spatial transformer based motor imagery classification for BCI using independent component analysis", *Biomedical Signal Processing and Control*, vol. 87, pp. 105359, 2024. 135

[212] J. Liu, R. Wang, Y. Yang, Y. Zong, Y. Leng, W. Zheng, and S. Ge, "Convolutional transformer-based cross subject model for SSVEP-based BCI classification", *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 11, pp. 6581–6593, 2024. 135

[213] A. Shoeibi, N. Ghassemi, R. Alizadehsani, M. Rouhani, H. Hosseini-Nejad, A. Khosravi, M. Panahiazar, and S. Nahavandi, "A comprehensive comparison of handcrafted features and convolutional autoencoders for epileptic seizures detection in EEG signals", *Expert Systems with Applications*, vol. 163, pp. 113788, 2021. 135

[214] M. K. Ardali, A. Rana, M. Purmohammad, N. Birbaumer, and U. Chaudhary, "Semantic and BCI-performance in completely paralyzed patients: Possibility of language attrition in completely locked in syndrome", *Brain and Language*, vol. 194, pp. 93–97, 2019. 135

[215] D.-H. Kim, D.-H. Shin, and T.-E. Kam, "Bridging the BCI illiteracy gap: a subject-to-subject semantic style transfer for EEG-based motor imagery classification", *Frontiers in Human Neuroscience*, vol. 17, pp. 1194751, 2023. 135

[216] X.-Y. Huang, S.-Y. Chen, and C.-S. Wei, "Enhancing low-density EEG-based brain-computer interfacing with similarity-keeping knowledge distil-

lation", *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2023. 135

[217] M. Ferrante, T. Boccato, S. Bargione, and N. Toschi, "Decoding visual brain representations from electroencephalography through knowledge distillation and latent diffusion models", *Computers in Biology and Medicine*, vol. 178, pp. 108701, 2024. 135

[218] N. A. Alzahab, L. Apollonio, A. Di Iorio, M. Alshalak, S. Iarlori, F. Ferracuti, A. Monteriu, and C. Porcaro, "Hybrid deep learning (hdl)-based brain-computer interface (bci) systems: a systematic review", *Brain sciences*, vol. 11, no. 1, pp. 75, 2021. 135

[219] Z. Wang, L. Yang, Y. Zhou, L. Chen, B. Gu, S. Liu, M. Xu, F. He, and D. Ming, "Incorporating EEG and fNIRS patterns to evaluate cortical excitability and MI-BCI performance during motor training", *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 2872–2882, 2023. 135

[220] J. Chen, Y. Xia, X. Zhou, E. Vidal Rosas, A. Thomas, R. Loureiro, R. J. Cooper, T. Carlson, and H. Zhao, "fNIRS-EEG BCIs for motor rehabilitation: a review", *Bioengineering*, vol. 10, no. 12, pp. 1393, 2023. 135

Appendix A. Statement of IEEE Copyright

In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of Nottingham Trent University's products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink. If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation